MAPS: A Comprehensive Feature Model for Prostate Cancer Diagnosis with Multiparametric MRI

by

Andrew Cameron

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Systems Design Engineering

Waterloo, Ontario, Canada, 2014

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Prostate cancer killed over 33 000 North American men in 2013 [3, 9]. However, the survival outlook for prostate cancer is very good if it is caught early. Prostate cancer screening is therefore very important. Although many methods are currently used to screen for prostate cancer, the use of multiparametric magnetic resonance imaging (mpMRI) is increasing in clinical practice and has been shown to have some power in differentiating between healthy and cancerous tissue. This thesis presents a comprehensive feature model for performing prostate cancer diagnosis using mpMRI. It incorporates a novel tumour candidate identification algorithm to efficiently and thoroughly identify regions of concern and a feature model to grade these regions for severity. Unlike conventional automated classification schemes, this feature model aims to ground its decisions in a way that can be interpreted and understood by the diagnostician. It does this by grouping features into high-level feature categories which are already used by radiologists to diagnose prostate cancer: Morphology, Asymmetry, Physiology, and Size (MAPS), using biomarkers inspired by the PI-RADS guidelines for performing structured reporting on prostate MRI [5]. To the author's best knowledge, the proposed feature model is the first using morphology and asymmetry features for prostate cancer detection.

Clinical mpMRI data were collected from thirteen men with biopsy-confirmed prostate cancer and labeled by an expert radiologist with thirteen years of experience diagnosing prostate MRI. These annotated data were used to train classifiers using the proposed feature model in order to evaluate classification performance. Training was performed using cross-validation in order to avoid overlearning the training set. Experimental results indicated that the proposed model outperformed each of its constituent feature groups as well as a comparable state of the art feature model.

Further work on the MAPS feature model is still warranted. Although the initial results are promising, more data are needed to refine the feature model and discard those features with no predictive power. Additional features should be investigated for inclusion in the model, so that the existing features may be conditioned on the prostate region to reflect the different characteristics between, for instance, the peripheral and the transition zones. Finally, user experience and user acceptance studies would help investigate the degree of cognitive support to diagnosticians that the MAPS model provides.

Acknowledgements

I couldn't have written this thesis without the support of my supervisor, Alex Wong. Thank you for pushing me, for believing in me, and for joining me on countless treks to the plaza.

Many thanks also to Dr. Haider at the Sunnybrook Health Sciences Centre. Your generosity, both with your time, and in providing thoroughly-annotated ground truth data, made this research possible.

To Mom and Dad: words fail me at describing everything you've done for me. Thank you for teaching me to think for myself, for buying me VCRs to disassemble, for wearing the very latest in carrot-based fashions, and for everything else.

Finally, my gratitude to all of the Vision and Image Processing group members, past and present; those with whom I have toiled and those upon whose shoulders I have stood. I hope we have the good fortune of collaborating again in future.

Dedication

À la mémoire de Louis Sylvestre.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Prostate cancer is the most-diagnosed form of cancer in Canadian men, with an estimated 23,600 new cases projected to be diagnosed in 2013 with 3,900 of those cases expected to result in death [9]. Although it is the third leading cause of cancer death in men, prostate cancer has good prognosis if detected early [10]. Effective and reliable screening methods for prostate cancer are therefore very important.

The main contribution of this work is the introduction of a comprehensive feature model for performing prostate cancer detection and localization using multiparametric magnetic resonance imaging (mpMRI). Multiparametric MRI involves repeated imaging of a patient's tissues while varying certain parameters of the MRI acquisition protocol, resulting in several image series (or "modalities"), each emphasizing different tissue characteristics. The proposed model gathers features into various groups in order to present classification decisions to diagnosticians in a way that supports their existing diagnostic reasoning process. The features presented in this model are evaluated on clinical MRI data to show their suitability for prostate cancer diagnosis.

1.1 What Is Being Done?

Conventional methods of screening for prostate cancer include the prostate-specific antigen (PSA) test, trans-rectal ultrasound (TRUS), and magnetic resonance imaging (MRI). Some of these methods, such as the PSA test, have come under fire for oversensitivity, causing expensive and often painful follow-up procedures. Imaging-based methods produce image series which require expert review from an experienced diagnostician, who looks for

subtle features to inform their diagnosis. When multiple types of imaging are performed, they must be registered to a common coordinate system. Even when registered, it can be challenging to search for features across multiple imaging modalities at once, and it's common for a single scene to receive varying diagnoses if evaluated by different diagnosticians (known as "inter-observer variability"), or even if evaluated by a single diagnostician multiple times in multiple sittings ("intra-observer variability"). In addition to diagnosis, imaging can aid in treatment, as in the case of Image Guided Radiation Therapy (IGRT), which uses real-time imagery (such as MRI) to guide the therapeutic beam of ionizing radiation. TRUS similarly uses real-time ultrasound imagery to guide biopsy needles for sample acquisition.

The European Society of Urogenital Radiology recently introduced a set of guidelines for interpreting several MRI image series together known as PI-RADS, or the Prostate Imaging – Reporting and Diagnosis System. This system aims to improve consistency between diagnosticians by providing a common set of criteria to consider when diagnosing a case. However, as the features laid out in PI-RADS are subjective, there is still an opportunity for observer variability to manifest.

Research on automatic methods for detecting and localizing prostate cancer consist mostly of ensemble classifiers trained on huge numbers of features. The classification decisions made by such automatic systems can be difficult for a radiologist to reason about, consisting of a binary output or some kind of probability or confidence score. Without a mechanism to reason about why an automatic classifier computes the result it does, doctors tend to distrust them.

1.2 Thesis Contributions

The aim of this thesis is to introduce a framework to help bridge the gap between the quantitative characterization and clinical interpretation of prostate cancer diagnosis. By grouping classifier features together, we hope to create a diagnostic support system which doctors can leverage to augment their understanding rather than supplant it. Beyond stating that a region is suspected to be cancerous, a system built on this framework can explain that the region has an unusual morphology, for instance. With intuitive features, classification decisions can be more easily justified and explained to medical practitioners, while simultaneously providing more consistently quantifiable measures. In this work, we introduce a novel strategy for computer-aided prostate cancer analysis that combines: i) a tumour candidate identification scheme based on multiparametric MRI and morphology, and ii) a comprehensive feature model comprising groups of features evaluating tumour

candidate morphology, asymmetry, physiology, and size. By presenting results as a collection of factors doctors are already used to using to evaluate prostate cancer, the proposed system should foster user acceptance, and the use of objective, quantitative features can help to reduce inter- and intra-observer variability.

1.3 Thesis Outline

We explore the problem domain in Chapter 2, discussing prostate cancer as well as conventional and automatic methods of prostate cancer detection, before reviewing some important image processing tools. In Chapter 3 we explain the feature framework: motivate the feature groupings and explore each group and its constituent features. Preliminary results using actual clinical data are shown in Chapter 4. Finally, current progress is summarized and directions for future work are presented in Chapter 5.

Chapter 2

Background

This chapter provides fundamental information on what prostate cancer is, whom it affects, and how it's diagnosed. We'll briefly cover prostate anatomy and the prevalence of cancer within the prostate, as well as screening methods, imaging techniques, and their strengths and weaknesses. Finally, we'll discuss previous work in automatic prostate cancer detection.

2.1 Prostate Cancer

Whom does prostate cancer (PCa) affect? Does it occur uniformly throughout the prostate, or is it concentrated in specific areas; and for that matter, how is the prostate structured? How can the aggressiveness of a cancer be determined? The answers to these questions will aid in understanding how PCa is analyzed and diagnosed.

2.1.1 Prevalence

Until recently, prostate cancer was the leading cause of cancer death among North American males [2, 3]. Improved survival rates can be attributed to advances in screening. Despite this, over 30,000 North American men are estimated to have died of prostate cancer in 2013.

2.1.2 Anatomy

The prostate consists of several different volumes of tissue, known as zones, shown in Fig. 2.1. The peripheral zone (PZ) is the largest of these, and tumours in this zone account for 75% of all prostate tumours [32, 46]. The transition zone (TZ) starts out consisting of only 10-20% of the prostate, but starting around age 40 it can grow significantly in a condition known as Benign Prostatic Hyperplasia (BPH), ultimately becoming larger than the PZ. Approximately 20% of prostate cancers arise in the TZ. The remaining 5% or so of tumours occur within the Central Zone (CZ), which surrounds the ejaculatory ducts. These ducts run from the seminal vesicles and vas deferens to the verumontanum, where they join the urethra.

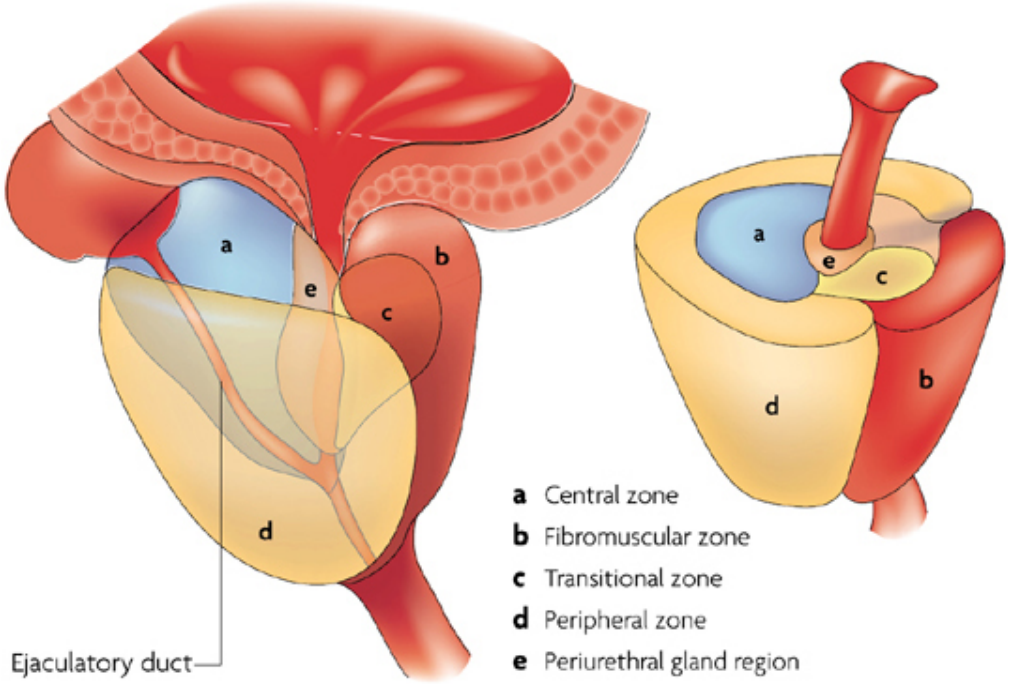


Figure 2.1: Zones of the prostate. Most cancers occur in the peripheral zone, with a sizeable minority occurring in the transitional zone. Very few cases occur elsewhere in the prostate. Adapted by permission from Macmillan Publishers Ltd: Nature Reviews Cancer [14], ©2007

Unlike the structured zones of the prostate gland, prostate cancer can take on many forms, from a focal mass to a patchwork of tissue scattered through a volume. Nevertheless,

prostate cancers are distinguished based on their progress at the cellular level using Gleason grades. Cancerous tissue may have a Gleason grade between 1 and 5, where grade 1 indicates tissue most resembling healthy prostate tissue, and grade 5 indicates the most advanced cancers, with sheets of cancerous cells instead of the gland structures present in healthy prostate tissue.

2.2 Clinical Workflow

Several different methods are currently used to detect prostate cancer, though until recently there has been no widely-accepted standard clinical practice. The prostate-specific antigen (PSA) test measures the concentration of specific markers in the patient's blood of which high levels indicate high risk for prostate cancer. However, use of the PSA test has been criticized for having an unacceptably high occurrence of false positives, causing healthy patients to undergo expensive and sometimes invasive tests to determine whether the PSA test result was simply a false positive. [11, 48]. The PSA test is therefore recommended particularly for patients with a predisposition, such as those with a family history of prostate cancer. Following a positive PSA test, a series of images might be acquired using one or several imaging modalities, including MRI or positron emission tomography (PET). Then a systematic transrectal ultrasound (TRUS) guided biopsy is undertaken where multiple samples are collected. In some cases, the previously-acquired imagery is registered to the TRUS imagery to aid in the biopsy. Biopsy samples are examined under magnification for the presence of cancer and graded using the Gleason scale. Although core biopsies are very accurate, they are intrinsically invasive, causing significant discomfort to the patient and exposing him to possibly unnecessary surgical risk. Furthermore, some tumours are isoechoic and respond to ultrasound similarly to the surrounding tissues, making TRUS-guided biopsies even more difficult as the surgeon must attempt to collect tissue samples from invisible tumours. Depending on the results, further biopsies may be indicated, bearing further surgical risk. Following any biopsies, the patient may be treated with focal therapy, chemotherapy or even a radical prostatectomy, depending on the severity of his cancer.

The European Society of Urogenital Radiology recently introduced, as part of its MR guidelines, the Prostate Imaging – Reporting and Data System (PI-RADS) [5], modeled after the BI-RADS system for breast cancer reporting. PI-RADS breaks the decision process into several component scores, providing guidelines on the scoring of the various recommended modalities. For more details on PI-RADS and the modalities it uses, see Sec. 2.3, below.

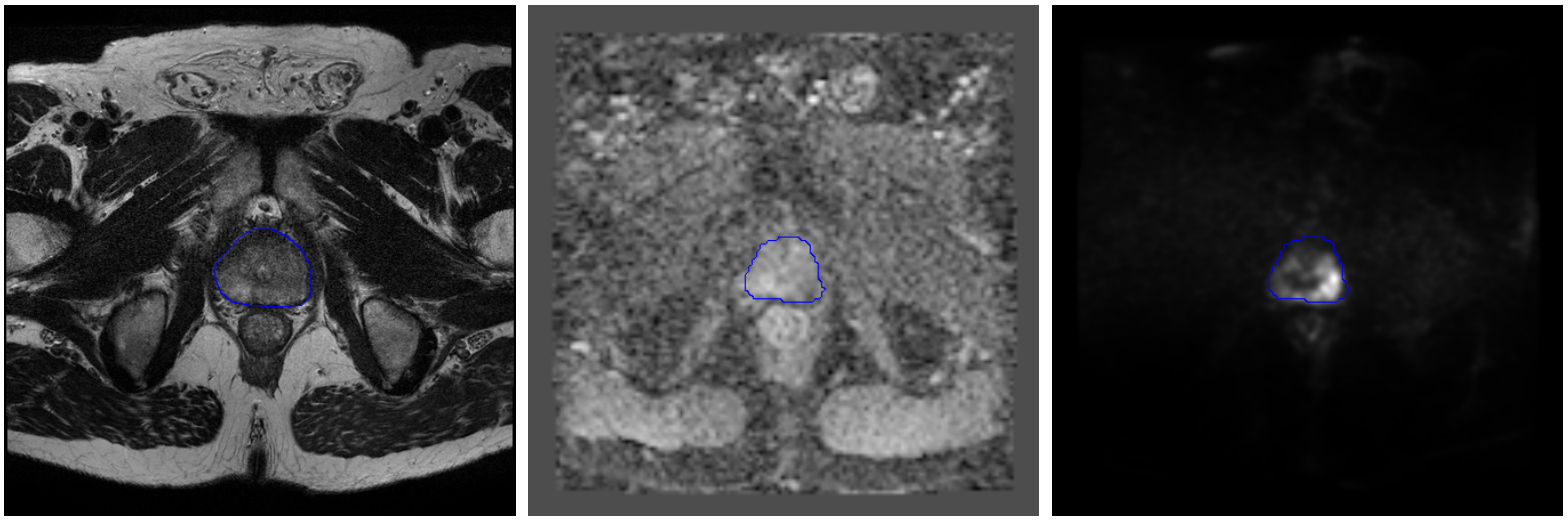
2.3 Multiparametric imaging

Multiparametric magnetic resonance imaging (mpMRI) has shown considerable promise for diagnosis, especially the combination of T2-weighted MRI (T2w) and diffusion-weighted MRI (DWI). T2w affords doctors the ability to see relatively high-resolution imagery of the prostate and surrounding tissue, while allowing detection of subtle structural features. Recent research has focused on DWI and the DWI-derived Apparent Diffusion Coefficient (ADC), which quantifies the diffusion of water molecules through tissue and has been shown to delineate between healthy and cancerous tissue [7, 38].

Other MRI modalities include dynamic contrast-enhanced MRI (DCE-MRI), in which a contrast agent is introduced and the subject is repeatedly imaged while the contrast agent is absorbed into the tissue. The contrast agent uptake behaviour of any given voxel of tissue is modeled and the time-series imagery is used to estimate tissue parameters such as onset time, time to peak, and washout. Magnetic resonance spectroscopy (MRS) is a rarely-used MRI modality, known as magnetic resonance spectroscopic imaging or in vivo nuclear magnetic resonance spectroscopy, which provides information about the chemical composition of the tissue. Correlated diffusion imaging (CDI) is a recently-developed modality which, like ADC, is derived from DWI, and has shown promise in differentiating between healthy and tumorous tissue. One drawback of CDI is that it does not show structural markers, making tumour foci difficult to register to other modalities unless viewed in overlay.

Example multiparametric MRI scans are shown in Figure 2.2. It can be observed that each modality provides different information about the prostate gland, which is useful in aiding radiologists and clinicians in the clinical decision support process. However, interpreting multiple modalities simultaneously can be challenging. Tools designed for the purpose of synchronized viewing of multiple image series use built-in facilities for cross-registering points of interest and for keeping multiple views aligned at the same position in patient-centred coordinates, but even with such tools it can be daunting to simultaneously interpret features across several modalities.

The aforementioned PI-RADS system puts guidelines into place for the interpretation of mpMRI scenes by diagnosticians. In addition, it has subsequently been supplemented with graphic examples to help illustrate the scoring scheme [41]. PI-RADS specifies the use of several different MRI modalities, including T2w, DWI, MRS, and DCE-MRI, each of which is scored, Likert-style, from 1 to 5. Different scoring guidelines are presented for each modality, so while MRS features are relatively straightforward to interpret by reading off the relative heights of two peaks from a spectrogram, other modalities, such as T2w,



(a) T2-weighted

(b) DWI: apparent diffusion coefficient (ADC)

(c) DWI: correlated diffusion imaging (CDI)

Figure 2.2: Example multiparametric MRI slices of the same prostate gland, using T2w and diffusion-weighted imaging (DWI). Each modality provides different information about the prostate gland (outlined in blue) to aid in the clinical decision support process. The lower-right corner of the prostate gland, which appears dark on ADC and bright on CDI, was indicated as tumorous by an expert.

may have very subtle features. Although these features are difficult to interpret without an understanding of prostate physiology, PI-RADS also provides a scale for performing qualitative analysis of MRS which makes the meaning of the scores for all the modalities clear: a score of 1 indicates that "clinically significant disease is highly unlikely to be present", while a score of 5 indicates that "clinically significant cancer is highly likely to to be present".

Unfortunately, despite the wealth of information provided by various MRI modalities, prostate cancer remains difficult to distinguish from other conditions, notably BPH. The additional tissue growth in the TZ occurring due to BPH causes similar MRI signal intensity changes as prostate cancer. Furthermore, attempts to differentiate cancerous tissue from healthy tissue in the central gland have failed to show differences in mean T2w or ADC values [22]. However, since BPH only occurs in the central gland, this issue only arises in the minority of cases which occur elsewhere than the peripheral zone.

2.4 Detecting PCa automatically

Automatic approaches to prostate cancer detection can be grouped into methods which attempt to segment the prostate into discrete regions of concern, and methods which instead assign some kind of numeric score to each voxel in the prostate. A variety of features and classification strategies are used to produce these scores, whether the scores are region-based or voxel-based. Features and classifiers will be discussed later, but for region-based methods an additional step must first be performed, namely the segmentation of the prostate into regions of concern.

2.4.1 Segmenting tumour candidates

The problem of segmenting tumour candidates is not yet well-explored; most classifiers built on MRI intensity features did not perform any tumour candidate segmentation, but instead computed voxelwise feature statistics, using ROC curve analysis, or output from SVMs, RVMs, or MRFs [12, 16, 19, 22–27]. Some high-level classifiers leveraged the added detail apparent in stained histology scans to segment different gland structures in the prostate, namely lumen regions, the cytoplasm surrounding the lumen, and the nuclei of the surrounding cells [35, 36]. Classifiers were trained on features computed from these regions to predict the Gleason grade of prostate tissue, but MRI modalities do not have adequate resolution to observe these structures, so these features cannot be leveraged for MRI tumour candidate identification.

2.4.2 Feature extraction

Detecting prostate cancer automatically has been a popular area of research and to date, these methods have focused primarily on low-level features. Conventional multi-parametric MRI prostate cancer analysis approaches use one or more of the MRI modality values or derived values, including ADC and T2w, with no complementary high-level features [19]. Published values for classification accuracy using these low-level features ranges from 64% to 89% [12, 19, 26, 27, 38, 39].

More sophisticated approaches calculate large numbers of low-level features and then combine them in various ways (see Sec. 2.4.3); low-level features considered in such approaches include statistical features (e.g. local variance) and change-of-basis features such as Gabor filter coefficients [28], discrete cosine transform (DCT) coefficients, and textural features (both first and second-order) such as those derived from co-occurrence matrices (CMs). Chan et al. [12] reported that compared to using MRI intensity features exclusively, DCT and CM features improved accuracy between 6.2% and 11.0% (accuracy of 72.9% without those features, and 79.1% – 83.9% with them). Madabhushi et al. [28] reported a positive predictive value (PPV) of approximately 21% when using a combination of first and second-order statistical features, gradient features, and Gabor features within an ensemble classification framework. However, since the low-level features lack semantic meaning, the classification decisions made by such approaches are difficult to justify to radiologists and clinicians, who typically use subjective heuristics to diagnose patient cases. Although the latter approach is easy to understand and justify, its subjective nature makes it highly inconsistent.

Unlike low-level features, which typically characterize some kind of voxelwise physiology characteristic, high-level features often exploit *a priori* knowledge of the problem to quantify high-level domain concepts. High-level features hand have been used in detecting and localizing prostate cancer; however, their use is still limited, mostly to analysis of histopathology, which necessitates invasive surgery to acquire the tissue samples subsequently scanned at high resolution; for instance, Naik et al. [36] demonstrated that classifiers built using high-level features designed using domain knowledge are effective at diagnosing prostate tissue as cancerous and distinguishing cancerous tissue between two grades on the Gleason scale, with classification accuracy in the range of 86–95%. Notable among these high-level features was the use of region-based features; the added magnification achieved by histopathology scans allows observation and segmentation of structures too small to be apparent via MRI. Naik et al. computed features on the segmentation boundaries of these structures, such as smoothness and area, and used these features to perform classification. The use of such region-based features is not currently well-explored

for MR imagery. These particular high-level features cannot be computed at all for multiparametric MRI, as the structures used are too small to be resolved. Haider et al. [19] described a scoring scheme for diagnosing prostate cancer based on high-level features observed in T2w and ADC imagery. However, these features are not easily quantifiable and are intended for use by trained radiologists and not classification algorithms.

2.4.3 Classification

So far, several approaches to classification have been applied in various prostate cancer imaging contexts. Among approaches investigating multiparametric MRI, support vector machines (SVMs) are widely-used [12, 38, 39, 42], while other methods applied maximum likelihood (ML) classification [12, 19], logistic regression [26], Markov random fields (MRFs) [27, 39]. Approaches attempting classification of histopathology data tended to use more features and thus more sophisticated methods of combining those features; for instance, combining many weak Bayesian predictors using an ensemble method such as AdaBoost [16, 28], or using a manifold learning technique such as Graph Embedding or Local Linear Embedding to reduce the size of the feature space passed into a SVM classifier [15, 29, 36]. Interestingly, the multiparametric MRI approach with the greatest number of features reported greatest classification accuracy with the comparatively simple Fisher's linear discriminant (FLD) classifier [12].

2.5 Summary

Despite the recent introduction of PI-RADS, systematic prostate cancer evaluation schemes are still in their infancy. Although much research has been carried out on the use of various modalities and numerous features for detecting prostate cancer, none of the automatic approaches explored so far attempt to support their decisions with a similar framework as those used by radiologists diagnosing PCa manually. There is no attempt to map automatic classification decisions to the language familiar to diagnosticians in their work. The frameworks which do group features together in this manner, such as those presented by Haider et al. [19] or in PI-RADS [5], are meant for interpretation by radiologists. They are not quantifiable, and due to their subjective nature lack consistency in interpretation from one session to the next and from one doctor to the next. The field is ripe for a comprehensive feature model which enables consistent, quantifiable feature analysis in a manner that is familiar to diagnosticians.

Chapter 3

Methodology

The proposed model consists of an initial tumour candidate identification scheme followed by the MAPS feature model used to score the candidate regions. In the proposed method, initial identification of candidate tumour regions is automatically performed using multi-parametric MRI and region morphology. After candidate regions are identified by the automatic tumour candidate identification algorithm, numerous features are extracted from the various mpMRI modalities. These features are grouped into high-level categories according to the type of characteristics they describe. The Morphology feature group describes region shape; whether the boundary is smoothly-varying or jagged, whether it consists of small protrusions or a focal mass, and so on. The Asymmetry feature group describes the extent to which the two halves of the region appear similar when split along its major and minor axes. Tissue characteristics are captured in the Physiology group, which includes mean signal intensity for all the MRI modalities as well as a series of textural features and filter responses. Finally, the Size feature group captures the physical extent of the region. Collectively, these feature groups afford a diagnostician an improved ability to interpret the system's decisions. The tumour candidate identification algorithm, as well as each of the feature groups and all their constituent features, are described in detail in the following sections.

3.1 Automatic tumour candidate identification

The first step in diagnosing prostate cancer is identifying all regions which might be cancerous. An automated system working from straightforward screening procedures can quickly and thoroughly scan all the imagery for a given case, flagging screened regions for review by

a radiologist, producing screening results with less observer variability. Tumor candidate regions were identified automatically in the proposed system using guidelines for clinical multiparametric MRI prostate cancer screening by a radiologist [19]. Tissues satisfying these criteria were grouped into connected regions and analyzed further with the feature model detailed in Section 3.2. In particular, diffusion characteristics and morphology were used to automatically identify candidate regions in the proposed system.

As low ADC values are associated with tumorous tissue [19], tissue with ADC values below a particular threshold were automatically identified by the proposed system as possible tumour candidates. Considering that the priority of the candidate identification phase is to highlight regions for consideration, rather than to eliminate them, the threshold value used in [19] was relaxed to capture more regions. All voxels with an ADC value less than $700 \times 10^{-6}\,\mathrm{mm}^2/\mathrm{s}$ were flagged as possible tumour candidates. These flagged voxels are then grouped into connected regions, and a second phase is used in the automatic tumour candidate identification algorithm to take morphology into account. In particular, size is taken into consideration and all connected regions larger than $1\,\mathrm{mm}^2$ were considered as the final set of tumour candidate regions.

Example candidate identification results are shown in Figure 3.1. The tendency of the identification scheme to identify many false positives is evident; although the region labelled as tumorous in the ground truth was successfully identified, so were many other regions. This practice is common in clinical screening, which is effectively the purpose of tumour candidate identification: to identify all suspects, so that a more specific test can eliminate the false positives. This helps reduce the number of false negatives, or regions of concern missed by the test that are nevertheless cancerous. For a discussion about screening and test evaluation, see Subsection 4.3.1.

3.2 MAPS feature model

In this section, we discuss the proposed feature model in detail. Each feature group in the MAPS model (Morphology, Asymmetry, Physiology, Size) is expounded along with its component features. Illustrations have been included as much as possible to help visualize and contrast the characteristics captured by each feature.

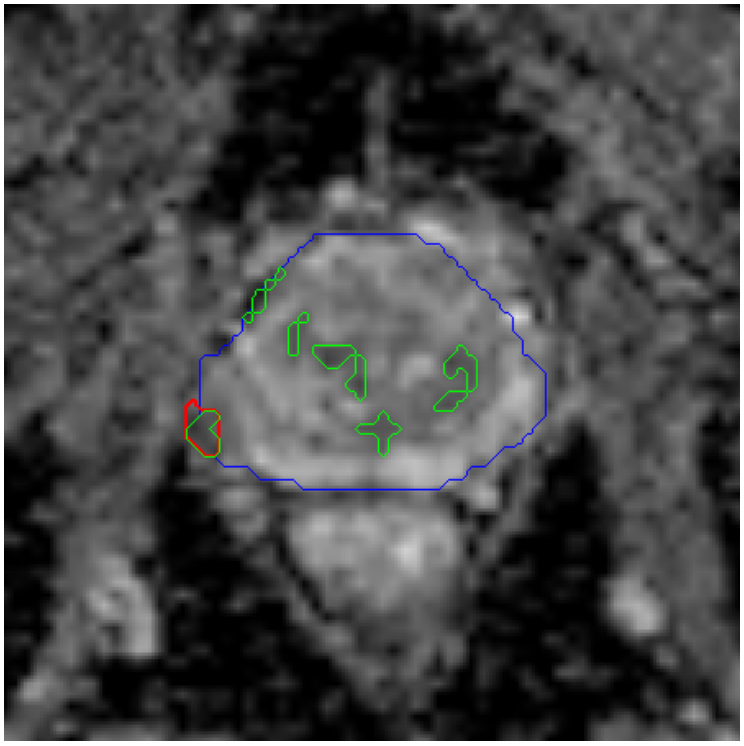


Figure 3.1: Example tumour candidate identification results, showing contours around the prostate (blue), true positive regions (red), and regions identified by the tumour candidate identification algorithm (green), overlaid on the ADC image (shown here interpolated to the same resolution as T2w). Although the proposed method successfully identifies the true positive region, it errs on the side of overidentifying healthy regions.

3.2.1 Morphology

The PI-RADS guidelines make mention of morphology in several areas; the appearance of a "focal mass", for instance, is an indicator for a higher PI-RADS score in several categories. Viewed with T2w, a "bulging" effect in the PZ is highly indicative of cancer, while in the TZ cancerous regions are "usually lenticular or water-drop shaped". Viewed with DWI, focal regions as well as "linear, triangular, or geographical [regions]" are important biomarkers [5]. In the proposed feature model, a set of features are computed to characterize the morphology of the tumour candidate region. The morphological features used in the MAPS model capture structural information about a candidate region by applying operations which smooth the shape of the region boundary. Regions with little morpho-

logical irregularity undergo little change with the smoothing operator, while regions with highly irregular shapes will see a drastic difference. The various smoothing operators used to compute morphology features capture different characteristics about the region shape. In total, three morphological features are computed from the region boundary.

The first feature in this group is the normalized difference in area between the morphological closing of the region and the morphological opening of the region (cf. Appendix A), using an identical disk structuring element for both operations [1]:

$$f_1^M = \frac{A_{closed} - A_{opened}}{A_{initial}} \tag{3.1}$$

Here, A denotes the area of a region. Peaks and valleys in the border of the region will cause the area to increase after closing, while it will decrease after opening (i.e., $A_{closed} \geq A_{opened}$); therefore, regions with very irregular borders will have a greater difference between these two values, and the feature value will be greater. Smaller feature values then correspond to regions with borders which do not feature sharp peaks and valleys. An example region is shown in Fig. 3.2a.

The second morphological feature compares the length of the region's perimeter before and after eliminating high-frequency components in the Fourier space (see Appendix A), and normalizing the difference:

$$f_2^M = \frac{|P_{initial} - P_{reconstruction}|}{P_{initial}}, \tag{3.2}$$

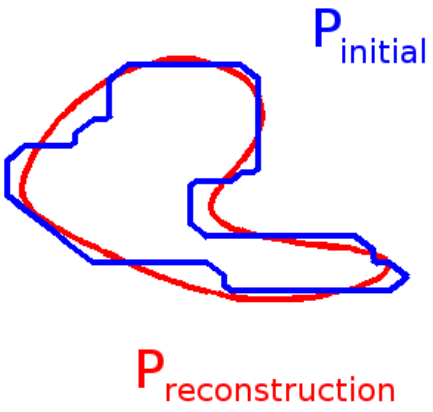
where each P is the perimeter of the region denoted by its subscript. Since high-frequency components capture rapid changes in the shape of the region, this feature will be greater for regions with rapidly-varying boundaries than for those with smooth, slowly-varying boundaries [1]. An example region is shown in Fig. 3.2b.

The third morphological feature is similar to the second, in that it examines the difference between two Fourier reconstructions of the region boundary, one at a low frequency and one at a higher frequency. However, it computes difference in area, rather than perimeter:

$$f_3^M = \frac{|A_{initial} \oplus A_{reconstruction}|}{|A_{initial} \cup A_{reconstruction}|}, \tag{3.3}$$

where $\oplus$ denotes the set symmetric difference, finding voxels which are in the low-frequency reconstruction or the high-frequency one, but not both. This value is normalized by the area of the union of both reconstructions, denoted by $\cup$.

(a) Morphological area feature



(b) Morphological perimeter feature

Figure 3.2: Illustrations of the morphological features. In (a), the area feature illustration, purple and orange contours denote the morphological opening and closing, respectively, of a candidate region. The large difference in areas between the contours results in a high feature value for this region.
In (b), the perimeter feature illustration, a candidate region contour, shown in blue, has a perimeter similar in length to that of the region's low-frequency Fourier reconstruction, shown in red, resulting in a low feature value for this region.

3.2.2 Asymmetry

The PI-RADS guidelines suggest that asymmetry in a lesions denotes an increased likelihood of cancerousness [5]. The Asymmetry feature group represents the degree of bilateral symmetry exhibited by a candidate region. In order to quantify the region's symmetry, the region is split in half along an axis passing through the region's center of mass. The halves to either side of this axis are then compared by taking the difference in their areas, and normalizing it:

$$f_{\cdot}^{A} = \frac{A_{large} - A_{small}}{A_{normalize}}, \tag{3.4}$$

where A_{large} and A_{small} represent the areas of the region halves, chosen such that $A_{large} \geq A_{small}$, and $A_{normalize}$ represents the area of the region used to normalize the difference. Four different Asymmetry features (illustrated in Figure 3.3) are computed in total by choosing either the major or the minor axis to split the region, and by choosing either the entire region area or the area of the smaller half-region as $A_{normalize}$. As regions may be symmetric about one axis but not the other, using both the major and minor axes

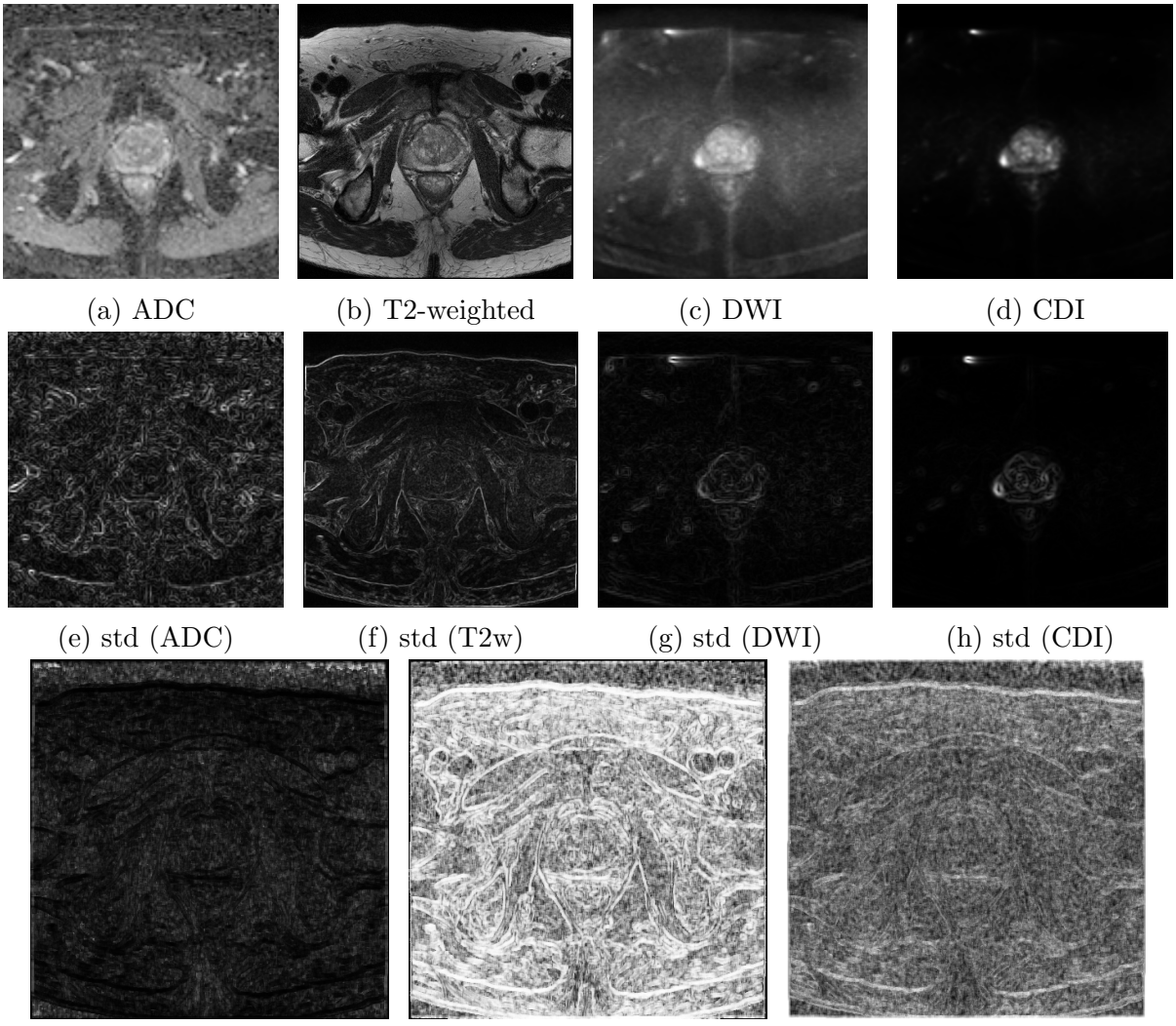
to measure region asymmetry allows the feature model to distinguish a greater variety of regions.

Figure 3.3: An example region contour is shown in blue, with the major axis drawn in black. The difference in area between the green and magenta regions is taken and then normalized two different ways, producing two different features. The process is then repeated using the minor axis (not shown).

3.2.3 Physiology

Cancerous tissue is often indicated by a number of physiological factors. PI-RADS includes several physiological MRI biomarkers as criteria, including "uniform high signal intensity", "homogeneous low signal intensity", or "heterogeneous [...] adenoma". The Physiology feature group attempts to capture such physiological factors. In addition to raw mpMRI signal values, many derived features are computed from the mpMRI data. Many of these features emulate various parts of the human visual system, such as second-order textural feature images, or Gabor filter responses. Raw mpMRI images as well as several derived images are shown in Figure 3.4.

While the features in the other groups can be computed from the region boundary, Physiology features are computed on a per-voxel basis. For each feature, the mean feature score was taken from all the voxels in the region, and the mean value used as the feature score for the region.



(a) ADC (b) T2-weighted (c) DWI (d) CDI

(e) std (ADC) (f) std (T2w) (g) std (DWI) (h) std (CDI)

(i) Contrast (j) Correlation (k) Homogeneity

Figure 3.4: Various Physiology feature images depicting the same region of tissue. Raw mpMRI values are shown in (a)–(d). First-order textural features (the standard deviation feature) are shown in (e)–(h), and various second-order texture images are shown in (i)–(k).

Raw mpMRI signal intensities

The most directly interpretable features consist of the MRI signal intensity values for the T2w, ADC, DWI, and CDI modalities. Important T2w biomarkers in the PI-RADS scheme include areas in the PZ of "uniform high signal intensity (SI)" and "homogeneous low signal [intensity] focus/mass", as well as "areas of [...] homogeneous low SI" in the TZ [5]. DWI measures the ability water to diffuse through tissue, and involves repeated imaging while varying several acquisition parameters, collectively referred to as the b-factor or b-value, to vary the degree of diffusion allowed to occur during acquisition. The DWI acquisitions at the various b-values are used collectively to estimate the ADC. Unlike DWI, which induces signal decay through diffusion (resulting in dark areas where water can diffuse), the ADC measures the rate of diffusion of tissue, so poorly-diffusing areas appear dark. Raw DWI and ADC intensity values are used in PI-RADS where areas of concern are indicated by "focal areas of reduced ADC" and "intense SI on high b-value images" [5]. CDI is not yet widely-used, but evidence suggests that CDI signal intensity is effective at delineating tumours from healthy prostate tissue [53].

Textural features

Texture features were included to capture the different textural characteristics between cancerous and healthy tissue [28]. They play an important role in PI-RADS, which uses properties including uniformity and homogeneity (or heterogeneity) in diagnosis of T2w imaging of the TZ [5]. A series of local statistical features were extracted from the mpMRI images in order to quantify textural characteristics of imaged tissue [28]. Feature values for each region were obtained by averaging texture image values over the region. Some of the texture feature images, before region-averaging, are shown in Figure 3.4.

For each candidate region, a set of low-level texture features were computed. Texture features quantify the occurrence of various patterns in an image. Whereas raw mpMRI values indicate the signal response from each respective voxel, texture images derived from the MR images indicate how the signal response at any given voxel relates to those around it. For example, the mean texture image can be computed using a $k \times k$ sliding window, where each voxel in the mean texture image consists of the mean value of the k^2 voxels in the $k \times k$ window centered around that voxel in the original image. As four different raw modalities were used (T2w, DWI, ADC, CDI), each texture operation described below constitutes four features in the Physiology group.

Texture features may be classified as first-, second-, or higher-order features, depending on the number of voxels in a particular spatial arrangement being compared in the

computation of each class of feature. That is to say, first-order features compare a single voxel to some global value; the median feature, for instance, is a first-order texture, as although several voxels are used to compute the median, their relationship to each other is not important except insofar as they all belong to the same sliding window around the feature voxel. By contrast, second-order features such as homogeneity are computed using a cooccurrence matrix (CM), which is a measure of the distribution of pairs of intensities in an image (see Appendix A for more information on CMs). Higher-order textural features compare the relationship between more than two voxels.

First-order features The first-order textural features in the MAPS model consist of the median, standard deviation, and mean absolute deviation. Each of these three features is computed on the T2w, ADC, DWI, and CDI modalities using a sliding window. Two window sizes were used for a total of 24 first-order texture features. The median and standard deviation feature images are straightforward to compute; each voxel (i, j) in the feature image is the result of evaluating the appropriate function (e.g. median) on $\mathcal{N}_{i,j}$, the set of voxel values belonging to the sliding window around voxel (i, j) in the original image I:

$$f_{median}(i,j) = \text{median}(\mathcal{N}_{i,j})\,, \tag{3.5}$$

$$f_{stdev}(i,j) = \sqrt{\text{var}(\mathcal{N}_{i,j})}\,; \tag{3.6}$$

while the mean absolute deviation feature image is simply the absolute value of the voxel-wise difference between the original image and the mean feature image:

$$f_{absdev}(i,j) = |I(i,j) - \text{mean}(\mathcal{N}_{i,j})|\,. \tag{3.7}$$

Second-order features The second-order features in the MAPS model consist of contrast, correlation, and homogeneity, previously used in [28], computed on T2w imagery at two scales each for a total of six features. Computed from a CM p (cf. Appendix A), the second-order textures are given by:

$$contrast = \sum_{i,j} |i-j|^2\, p(i,j)\,, \tag{3.8}$$

$$correlation = \sum_{i,j} \frac{(i-\mu_i)(j-\mu_j)p(i,j)}{(\sigma_i\sigma_j)}\,, \tag{3.9}$$

$$homogeneity = \sum_{i,j} \frac{p(i,j)}{1+|i-j|}\,. \tag{3.10}$$

3.2.4 Size

Region size is used by clinicians to assess mpMRI; the scheme proposed by Haider et al. uses the presence of nodules larger than 5 mm in the TZ as indicative of cancer and further uses size to distinguish between classifications of "Probably cancer" and "Definite cancer" [19], while PI-RADS suggests that if the size of the region's interface with the prostate surface is too large, cancer is indicated [5]. The Size feature group contains a single feature measuring the area of the candidate region. Since the resolution of a voxel changes with the modality and with the equipment, counting region sizes by number of voxels would be inconsistent. The size is therefore calculated in physical units, such as square millimeters. The size feature is shown in Figure 3.5.

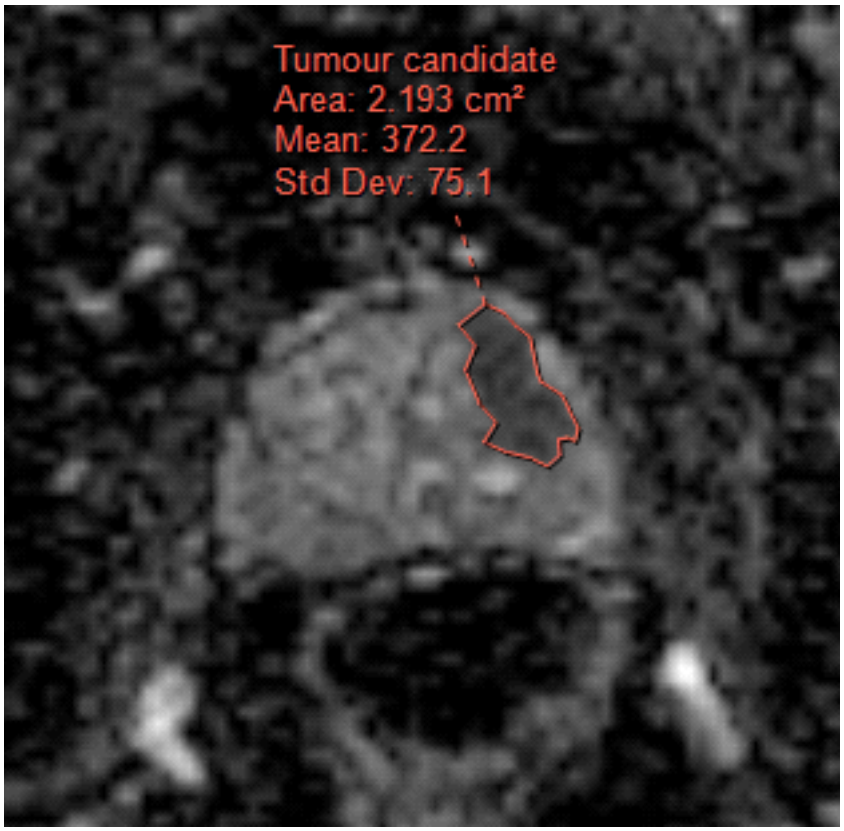


Figure 3.5: Size feature illustration. A tumour candidate is shown outlined in ClearCanvas Image Viewer software along with its area.

3.3 Summary

The proposed method introduces a novel means of segmenting candidate tumour regions by combining physiological and morphological information. Features are then extracted from the candidate regions in several high-level categories derived from clinical practice: Morphology, Asymmetry, Physiology, and Size. These feature groups are summarized along with the number of features in each group in Table 3.1. These feature groups capture information already commonly used by radiologists in the diagnosis of prostate cancer using mpMRI data, and groups this information in a way familiar to the diagnosticians. Features in the Morphology, Asymmetry, and Size groups describe properties of the region, while features in the Physiology group more traditionally describe properties of the tissue on a finer scale. In order to produce a single feature score for any given region, then, the feature values at the finer scale are averaged, and their mean value is used as the region's feature score. Collectively, this tumour candidate identification scheme and set of feature groups is the MAPS feature model. In the next chapter, the performance of this model at detecting prostate cancer will be evaluated and analyzed.

Feature group	Number of features
Morphology	3
Asymmetry	4
Physiology	34
raw mpMRI	4
textural (1st-order)	24
textural (2nd-order)	6
Size	1
total	42

Table 3.1: Summary of feature groups

Chapter 4

Experimental results

The proposed MAPS feature model was tested by using actual clinical mpMRI data to extract the MAPS features, which were then used to train automated classifiers. The performance of classifiers using the proposed feature set and various related feature sets was measured against ground truth diagnoses provided by an expert diagnostician. Various subsets of the entire model were independently evaluated to compare, for instance, the performance of the Morphology features on their own compared to the Asymmetry features. This chapter explains in detail how the clinical mpMRI data were acquired, how the data were processed for feature extraction, and how the classifiers were trained and evaluated. Numeric results are shown and interpreted in Section 4.3.

4.1 Data acquisition protocol

Fifteen patients with biopsy-confirmed tumours were imaged in total. Informed consent was obtained from all patients, and approval for this study was obtained from the ethics review board of Sunnybrook Health Sciences Centre. All acquisitions were made at Sunnybrook Health Sciences Centre in Toronto with a Philips Achieva 3.0 T device with SENSE-Cardiac coil. Patient ages and DWI acquisition parameters are summarized in Table 4.1. Images were reviewed and annotated by a trained radiologist with 18 and 13 years of experience interpreting body MRI and prostate MRI, respectively. Each annotation consists of a contour of the prostate gland as well as a contour around any cancerous regions identifiable by MRI.

Next, tumour candidates were identified by the proposed MAPS tumour candidate identification algorithm. Initially, all voxels with ADC values under $700 \times 10^{-6}\,\mathrm{mm}^2/\mathrm{s}$

Age	DFOV (cm^2)	Resolution (mm^3)	TE (ms)	TR (ms)
74	24×24	$1.667 \times 1.667 \times 3$	61	6692
61	20×20	$1.5625 \times 1.5625 \times 3$	61	6178
70	20×20	$1.5625 \times 1.5625 \times 3$	61	6178
53	20×20	$1.5625 \times 1.5625 \times 3$	61	6693
53	20×20	$1.5625 \times 1.5625 \times 3$	61	6178
74	24×24	$1.875 \times 1.875 \times 3$	63	8717
56	20×20	$1.786 \times 1.786 \times 3$	61	6667
57	24×24	$1.667 \times 1.667 \times 3$	61	6177
60	20×20	$1.5625 \times 1.5625 \times 3$	61	6178
75	20×20	$1.5625 \times 1.5625 \times 3$	61	6178
62	20×20	$1.5625 \times 1.5625 \times 3$	61	6178
70	20×20	$1.5625 \times 1.5625 \times 3$	61	6178
60	24×24	$1.667 \times 1.667 \times 3$	61	6692

Table 4.1: Summary of patient information for DWI acquisitions

were segmented as tumour candidates. Following this, connected component analysis was performed (cf. Appendix A) and those components with an area less than $1\,mm^2$ were discarded. An affine transformation (scaling, rotation, and translation) was applied to the DWI and ADC imagery as to register it to the T2w imagery according to the patient coordinates in the MRI metadata. The gland and region masks were likewise transformed. This ensured the various modalities could be meaningfully compared on a voxelwise basis by registering the imagery to a common coordinate system and interpolating to a common resolution. Finally, MAPS feature values were computed for each region, and the resulting feature vectors used to train and evaluate classifiers.

4.2 Classification and cross-validation

The features extracted from the clinical mpMRI data were used to train many naïve Bayes classifiers. Naïve Bayes classifiers assume that each feature is independent from one another. Formally, the joint probability of the region label and feature values is equal to the product of the prior probability and the likelihood of the assumed-independent feature values:

$$p(C|f_1^M, f_2^M, f_3^M, f_1^A, \cdots) = p(C) \prod_i f_i \,. \tag{4.1}$$

Although this produces an unsophisticated classifier, in practice they often perform quite well; furthermore, the simplicity of the model implies a lower number of parameters to train, which is helpful considering that the dataset used here had on the order of twice as many training samples as there are features in the model. Learning interaction effects would necessitate much more training data. Ultimately, the classifications performed for the sake of this thesis were done to compare the performance of different feature groups in a consistent way, rather than to maximize classification performance, so while a reduction in model complexity and classification accuracy may seem to be a trade-off for the sake of limited training data, it actually serves to facilitate analysis.

In total, 251 regions were used to train and evaluate classifiers. Of these, 53 were cancerous regions identified by an expert diagnostician. In order to produce non-cancerous ground truth regions, the automatic tumour candidate identification algorithm was run and any regions it identified which belonged to the cancerous region labels were removed. Altogether, 198 non-cancerous regions were identified in this way. Rather than using all 251 regions to train a classifier and then evaluating that classifier on the same regions, which leads to "memorization" of the training set, a cross-validation approach was used to ensure that the samples used to evaluate a classifier were not samples the same classifier had used in its training. Each classifier was trained using a subset of the data and evaluated using a disjoint subset in order to avoid overlearning. In particular, a single region was held out from training and evaluated using a classifier trained on all the other regions, and then the process was repeated until each region had been held out once in this way (leave-one-out cross-validation), producing 251 evaluation trials. The agreement (or disagreement) between the Bayesian classifiers' predictions and the expert's diagnoses for each region were used to compute measures of classification accuracy of the proposed feature model and various other feature sets for the sake of comparison. Details about the performance metrics used to evaluate classifiers are given in the next section along with the performance scores achieved using the various feature sets.

4.3 Evaluation

Classifier performance was evaluated using three metrics in common use in clinical practice: accuracy, sensitivity, and specificity. The following section explains the calculation and usage of these metrics. The final section presents and interprets the experimental results.

4.3.1 Performance metrics

Accuracy, sensitivity, and specificity were all calculated. Using TP as the number of correctly identified positive cases (true positives), TN as the number of correctly identified negative cases (true negatives), and FP and FN likewise for false positives and false negatives, respectively, these measures are given by:

$$acc = \frac{TP + TN}{TP + TN + FP + FN}, \tag{4.2}$$

$$sens = \frac{TP}{TP + FN}, \tag{4.3}$$

$$spec = \frac{TN}{TN + FP}. \tag{4.4}$$

In short, accuracy measures the proportion of predictions which are correct, sensitivity measures the proportion of positive cases correctly identified as such, and specificity measures the proportion of negative cases correctly identified as negative. Highly sensitive classifiers are typically used in the clinical process as screens, while highly specific ones are used to weed out false positives from the set of patients identified by the screen.

4.3.2 Results and analysis

Several feature groups were evaluated with the scheme described above: each component feature group of the MAPS model, the model as a whole, and two feature groups from a related approach. Results are shown in Figures 4.1–4.2. First, we will discuss the performance of the various MAPS feature groups relative to one another and to the model as a whole. Next, we will examine the performance of MAPS compared to the state of the art feature model from Madabhushi et al. [28].

Performance of the MAPS feature groups

Classification performance of each of the feature groups comprising the MAPS model is shown in Figure 4.1. Several trends are worth noting: some feature groups achieve consistent scores with each metric, such as the Physiology feature group, which achieves over 80% accuracy, sensitivity, and specificity. Other groups achieve a high sensitivity, but low specificity and low accuracy: while the Size group achieves a sensitivity of 91%, its accuracy is only 32% and its specificity is lower still. The Asymmetry group is similar, at about

85% sensitive and 45% accurate. It is logical that the Size feature group, consisting of a single measure of total region area, should fail to accurately classify samples, as there are many factors which can contribute to region size besides cancer, such as haemorrhaging. Likewise, while asymmetry may successfully detect the majority of cancerous cases, it fails to distinguish from healthy cases, producing many false positives.

The Morphology and Physiology feature groups, on the other hand, achieve scores better than a random guess (i.e., better than 50%) across all three performance metrics. The Morphology group achieves an accuracy of 71% and specificity of 69%, though over 80% sensitivity. Morphology would seem to be better at detecting true positives than true negatives; however, considering that there are only three Morphology features, their performance is better than one might expect given that the Asymmetry group has a similar number of features and nearly twice the misclassification rate. Clearly the inclusion of region shape adds diagnostic power.

Of all the individual feature groups evaluated, the Physiology group had the best performance, achieving no scores under 80%. In fact, the Physiology features alone outperformed the competing model evaluated below (cf. Figure 4.2). At first glance this may not seem surprising; with 31 features, the Physiology group was easily the largest component of the MAPS model in terms of number of features. However, the model proposed by Madabhushi et al. [28] has a comparable number of features, and uses the same first- and second-order texture operators as the Physiology features, as well as other filters not used by the MAPS model (e.g., Gabor filters). Why should the Physiology group perform so much better than the other model, given the similar features used by both? The answer lies in the MAPS model's use of mpMRI data: while the other model may use more filters, the Physiology features (and by extension the MAPS model as a whole) more than make up for this with the additional information captured by the different modalities of MRI, each providing additional information about the depicted tissues. The addition of DWI-, ADC-, and CDI-derived features significantly improves classification performance.

Unsurprisingly, the unabridged MAPS model performs better than any one of its feature groups. Taken together, each group provides complementary information, so that tissue information captured by the Physiology features and shape information captured by the Morphology features bolster each others' shortcomings; the sensitivity of the Size and Asymmetry groups can be exploited while using the strength of the other groups to reject more false positives. The MAPS feature model achieved approximately equal accuracy, sensitivity, and specificity scores of 89%.

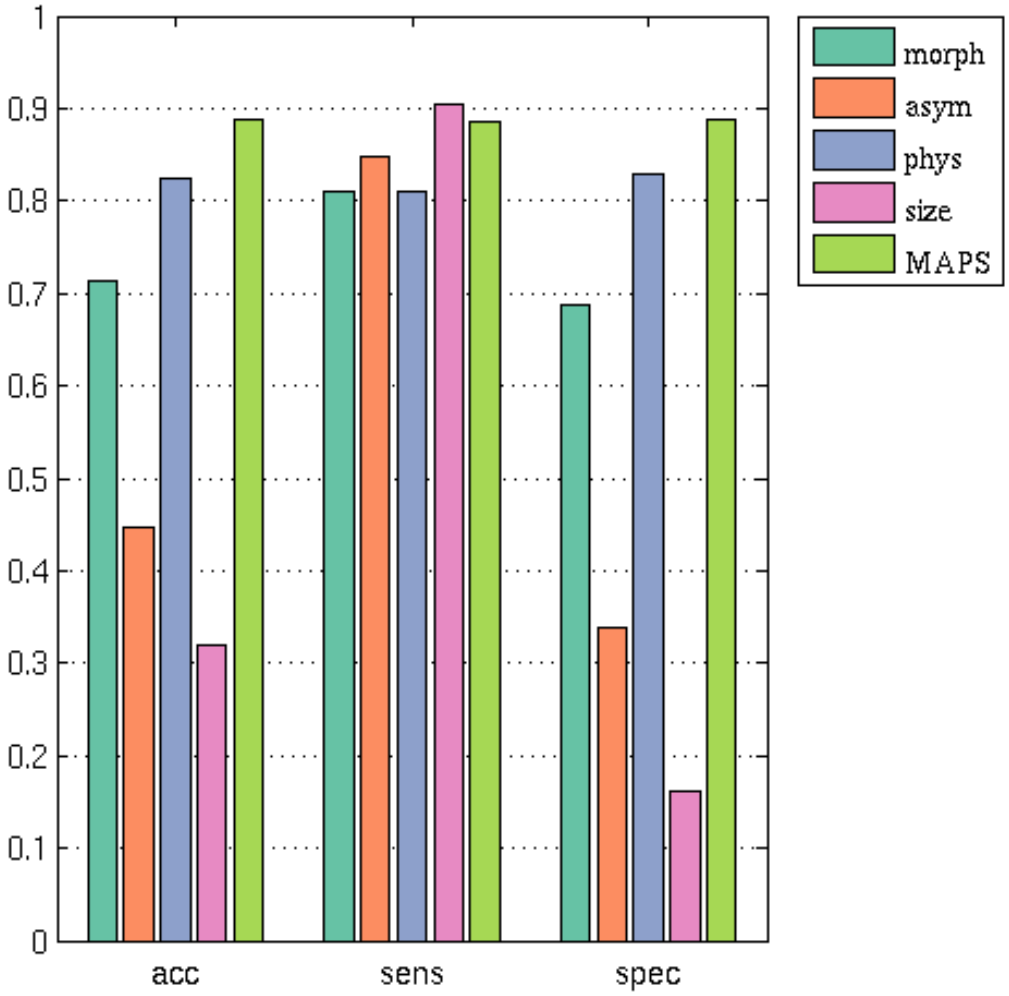


Figure 4.1: Classification performance (accuracy, sensitivity, and specificity) of each MAPS feature group independently (Morphology, Asymmetry, Physiology, Size), and the MAPS feature model as a whole. The performance of the MAPS model dominates the scores for each feature group except Size, which achieves a slightly higher sensitivity than MAPS, though abysmally low accuracy and specificity.

MAPS vs. state-of-the-art

The performance of the MAPS feature model is also shown in Figure 4.2, where it is compared against other feature models. In particular, the feature model proposed by Madabhushi et al. [28] based on T2w imagery was evaluated, along with the subset of that model which corresponds to Gabor filter responses. Interestingly, these two feature groups had comparable performance: they both achieved 47% accuracy, $78 \pm 1\%$ sensitivity, and 38–39% specificity, with the Gabor feature set being slightly more sensitive and the full feature set slightly more specific, although these differences are too subtle to be statistically significant (overlapping 95% Wilson confidence intervals are shown in Table 4.2). This suggests that adding textural features to Gabor filter responses did not improve cancer detection power, and that both feature groups characterizing strictly T2w imagery cannot differentiate between cancerous and noncancerous tissue. The many false positives produced by both approaches are a byproduct of the similar T2w characteristics between healthy and tumorous tissue, suggesting that T2w alone is not sufficient for good classification. Indeed, adding more information in the form of additional MRI modalities resulted in a much greater improvement, as shown with the performance of the MAPS model, which includes features derived from several mpMRI modalities in the Physiology feature group. Therefore, in addition to being easier to interpret by radiologists, the MAPS feature model achieves higher classification performance.

Feature set	sensitivity		specificity	
	lower bound	upper bound	lower	upper
Gabor	0.665	0.880	0.319	0.453
[28]	0.645	0.866	0.324	0.458
MAPS	0.774	0.947	0.838	0.926

Table 4.2: 95% confidence estimates on feature set performance. The significant degree of overlap in the confidence intervals between the first two feature sets suggests that the difference in performance of the two feature sets is not statistically significant. The MAPS set appears significantly more specific than the others.

Summary of observations

Although the components of the MAPS model may not be strong predictors individually, the experimental results make it clear that their combination is competitive with state of the art approaches. It was also observed that adding information in the form of additional

mpMRI modalities was more effective than additional filter response images on existing modalities at improving classification outcomes. Overall, the MAPS model performed better than any other feature group evaluated, achieving approximately equal classification accuracy, sensitivity, and specificity scores of 89%.

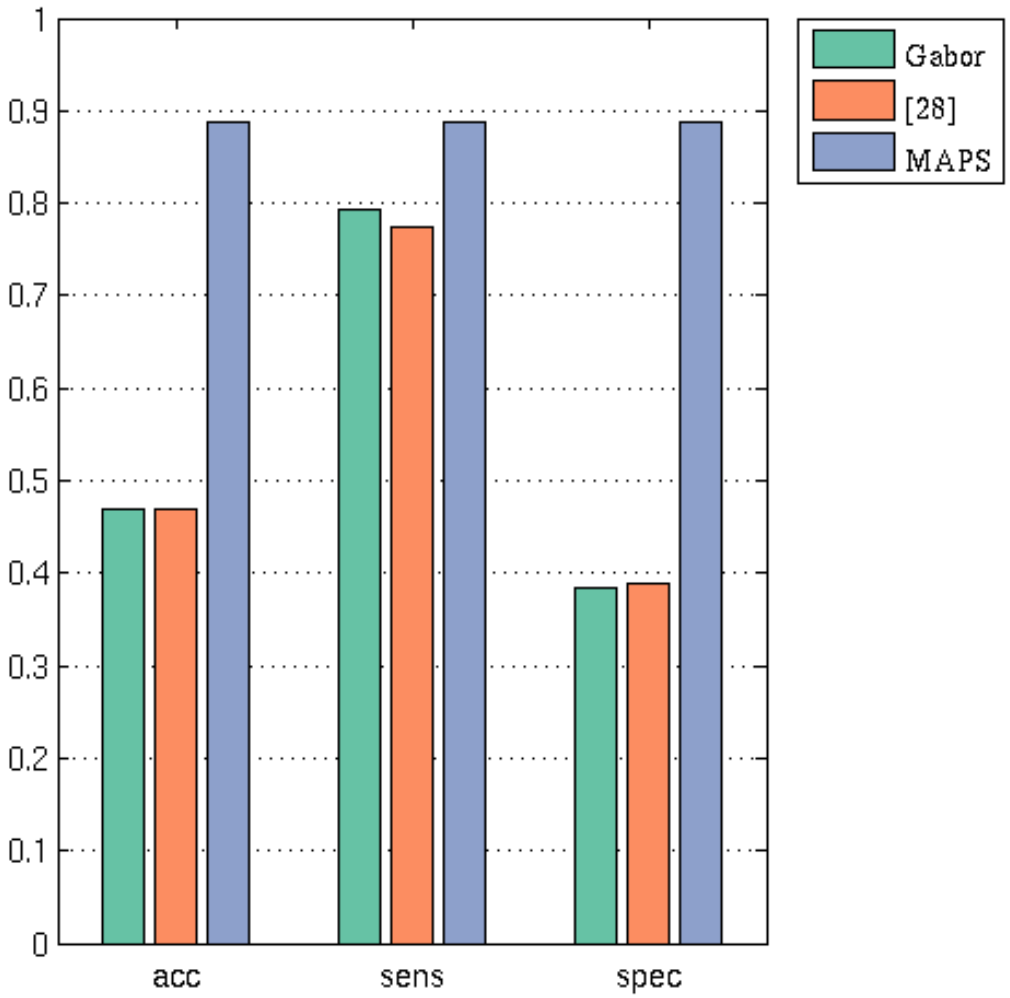


Figure 4.2: Classification performance (accuracy, sensitivity, and specificity) of: Gabor features on their own; Gabor features with several T2w-derived textural features as published by Madabhushi et al. [28]; and the MAPS feature model.

Chapter 5

Conclusions

In Chapter 3, we developed a comprehensive feature model for prostate cancer detection consisting of several feature groups in order to foster user acceptance and facilitate interpretability of classification results. However, the results must not only be interpretable, they must be accurate. To that end, the proposed feature model was evaluated on clinical multiparametric MRI data, annotated by a radiologist with thirteen years of experience diagnosing prostate MRI. By using some of the data to train a classifier and the rest of the data to test that classifier, we can estimate the performance of the feature model on novel datasets.

The strength of the results presented in Chapter 4 is limited due to the size of labeled dataset available for training; however, initial results are promising. Collectively, the features in the MAPS model achieved an accuracy of 89%, compared with 47% by the feature model from [28], achieving higher performance while providing a structured framework for interpretation.

5.1 Future Work

Although the initial results are promising, further work is needed to support the use of the MAPS model in the clinical decision-making environment.

5.1.1 Feature Selection

Further labeled MRI data would be invaluable in supporting the experimental results shown thus far. Additional data would provide stronger evidence of the performance of the MAPS feature model at detecting prostate cancer. In particular, with enough data, the model could be refined through feature selection in order to discard features shown to have little predictive power.

5.1.2 New features and changes to existing features

Furthermore, there are many additional features that could be considered for inclusion in the model. Although MAPS currently tries to capture many of the features used by radiologists when they are searching for cancer, additional ones can no doubt be designed, which should be evaluated along with the existing features for inclusion into the model. The Size feature group in particular, consisting of a single feature measuring region area, could be expanded with several different measures of size, such as the major axis length of the region, or the diameter of the largest circle which can be inscribed therein, for example.

5.1.3 User testing

Finally, and perhaps most importantly, the MAPS model needs to be evaluated in a clinical decision-making environment. Although initial results show promise for the framework's ability to automatically detect prostate cancer, no amount of automated testing can stand in for user experience testing. In order to validate MAPS as being easy to interpret, it must be evaluated by practitioners. Studies should be run to measure the impact (if any) MAPS has on observer variability, on the accuracy of the diagnostician accuracy, and on the speed with which diagnosticians can review patient cases.

APPENDICES

Appendix A

Image processing tools

This chapter explains in more depth some of the image processing tools used at various points throughout this thesis. A basic understanding of signal theory is assumed. Although the word "pixel" is used here to denote the discrete units ("picture elements") comprising an image, the methods described herein apply equally to volumetric data, in which case any mention of pixels should be taken to refer to volumetric picture elements, or "voxels".

A.1 Morphological operations

Morphological operators are used to perform spatial processing of images [18]. Though morphological operators can be applied to grayscale images, this thesis only considers the case of a binary input image. We can represent a region contour as such a binary image, with true-valued pixels inside the region and false ones elsewhere. Morphological operators pass a smaller binary image known as a "structuring element" over the input image. Each output pixel depends on the interaction of the structuring element with the input image. There are two basic morphological operators: erosion and dilation. There are also two more complex operators, opening and closing, but these are simply obtained by using an identical structuring element to perform first a dilation followed by an erosion (in the case of morphological opening) or vice versa (in the case of morphological closing), so it is sufficient to study dilation and erosion.

The morphological dilation of a binary image with a certain structuring element is straightforward to understand: the structuring element is successively centred at each true pixel in the original image, and at each location, any pixels reachable by the structuring

element are set to true-valued in the output image. Repeated application of the dilation operator will tend to cause a region to increase in size. If we take A to be the set of pixel coordinates of true-valued pixels in our binary input image, each of which lies on a two-dimensional integer grid $\mathbb{Z}^2$, and B to likewise represent the structuring element, then the dilation of A by B is given by:

$$A \oplus B = \left\{ z \in \mathbb{Z}^2 \mid (\hat{B})_z \cap A \neq \varnothing \right\} , \tag{A.1}$$

where $\hat{S}$ is the set of coordinates in S reflected through the origin (i.e., $\hat{S} = \{-s \mid s \in S\}$), S_t is the set S translated by the vector t (i.e., $S_t = \{s + t \mid s \in S\}$), and $\varnothing$ is the empty set.

Morphological erosion, on the other hand, removes area from a binary region mask. Instead of moving the centre of the structuring element through the true pixels and considering all pixels covered by the structuring element, the converse is true: the structuring element is only allowed where all the true pixels in the structuring element (not only the centre pixel) lie within the input mask, and while the structuring element is shifted around the input image in this manner, only those pixels visited by the centre pixel of the structuring element are set true in the morphologically eroded output image. The morphological erosion of A by B is given by:

$$A \ominus B = \left\{ z \in \mathbb{Z}^2 \mid B_z \subseteq A \right\} . \tag{A.2}$$

A.2 Fourier analysis of closed contours

For features relying on Fourier-domain filtering of region contours, it is assumed that the reader is familiar in general terms with frequency-domain concepts. In order to apply Fourier analysis to a closed contour, we must find an appropriate representation for the contour [18]. First, we consider the set of K contiguous pixels which comprise the region boundary: $\{(x_0, y_0), (x_1, y_1), \ldots, (x_{K-1}, y_{K-1})\}$. We reinterpret these coordinates as the real and imaginary components of a series of complex numbers, $s_i = x_i + \jmath y_i \, , i \in \{0, 1, \ldots, K-1\}$. Since the contour is closed, continuing from the last boundary pixel returns to the first boundary pixel; in other words, $s_K = s_0$ and in general $s_n = s_{n \bmod K}$, making the boundary signal periodic. We compute the one-dimensional Fourier series coefficients of this complex periodic sequence:

$$a(u) = \frac{1}{K} \sum_{k=0}^{K-1} s_k \exp\left(-\jmath 2\pi u k / K\right) . \tag{A.3}$$

The K frequency-domain coefficients comprising a may be used to extract information about the region's shape; for instance, it is clear that when $u = 0$, (A.3) reduces to the arithmetic average of the elements s_i, and thus the zero-frequency Fourier coefficient $a(0)$ stores the region centroid. Filtering may also be performed on a before using the inverse transform to reconstruct the filtered spatial-domain signal. For example, a low-pass filtered reconstruction of the contour may be obtained by summing only the first P frequency-domain coefficients:

$$s(k) = \sum_{u=0}^{P-1} a(u) \exp\left(\jmath 2\pi uk/K\right). \tag{A.4}$$

A.3 Co-occurrence matrices

Co-occurrence matrices (CMs) are similar to image intensity histograms: they measure the distribution of various intensity values appearing in an image. However, while histograms count the frequency of individual pixels taking on certain intensity values, CMs count the frequency of pairs of intensity values occurring at a specified offset. In particular, certain horizontal and vertical displacements are chosen, and pixel pairs separated by this offset are counted for inclusion in the matrix. For a CM p measuring image I using offset $(\Delta x, \Delta y)$, the entries are given by:

$$p_{\Delta x,\Delta y}(i,j) = \sum_{\forall m} \sum_{\forall n} \begin{cases} 1, & \text{if } I(m,n) = i \text{ and } I(m+\Delta x, n+\Delta y) = j \\ 0, & \text{otherwise} \end{cases} \tag{A.5}$$

Depending on the number of gray levels in the image, p may be a very large matrix. As a result, rather than interpreting p directly, it is common to compute metrics on the entries of p (cf. Subsection 3.2.3).

A.4 Connected component analysis

Connected component analysis is a method of analyzing binary images. Each contiguous region of true-valued pixels in an image is considered as one "connected component". Subsets of connected components can then be selected based on arbitrary criteria, such as region size, major axis orientation, or any other metric which can be computed on a closed 2D region.

A.5 Gabor filters

Gabor filters capture frequency information about the image. However, unlike the Fourier transform, which captures global frequency information, Gabor filtering captures periodic patterns over a smaller scale, producing an image where each pixel value corresponds to the local frequency response around that pixel in the original image. Any one Gabor filter applies at a specific scale and orientation, responding to a particular frequency. More precisely, a Gabor filter kernel is a sinusoidal plane wave (which detects frequency response) with a Gaussian envelope (which imposes locality on the frequency response):

$$g(x,y) = \exp\left(-\frac{1}{2}\left(\left(\frac{x'}{S_x}\right)^2 + \left(\frac{y'}{S_y}\right)^2\right)\right)\cos\left(2\pi f x'\right) , \tag{A.6}$$

where $x' = x\cos\theta + y\sin\theta$ and $y' = -x\sin\theta + y\cos\theta$ rotate the coordinates such that the line normal to the wavefronts lies at an angle of θ from the horizontal; S_x and S_y are the variances along the x' and y' axes, respectively; and f is the frequency. Several such filters are used in combination to complement each others' information. In particular, Gabor filters were applied to T2w imagery at a combination of frequencies and orientations totalling 18 filter response images, corresponding to the frequencies $f = \left\{16\sqrt{2}, 32\sqrt{2}, 64\sqrt{2}\right\}$ and the orientations $\theta = \{n\pi/6 \mid \forall n \in \{0, 1, \ldots, 5\}\}$.

References

[1] Robert Amelard, Alexander Wong, and David A. Clausi. Extracting Morphological High-Level Intuitive Features (HLIF) for Enhancing Skin Lesion Classification. In *34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4458 – 4461, San Diego, 2012.

[2] American Cancer Society. *Cancer Facts & Figures 2012*. American Cancer Society, Atlanta, 2012.

[3] American Cancer Society. *Cancer Facts & Figures 2013*. American Cancer Society, Atlanta, 2013.

[4] Gerald L Andriole, E David Crawford, Robert L Grubb, Saundra S Buys, David Chia, Timothy R Church, Mona N Fouad, Edward P Gelmann, Paul a Kvale, Douglas J Reding, Joel L Weissfeld, Lance a Yokochi, Barbara O'Brien, Jonathan D Clapp, Joshua M Rathmell, Thomas L Riley, Richard B Hayes, Barnett S Kramer, Grant Izmirlian, Anthony B Miller, Paul F Pinsky, Philip C Prorok, John K Gohagan, and Christine D Berg. Mortality results from a randomized prostate-cancer screening trial. *The New England journal of medicine*, 360(13):1310–9, March 2009.

[5] Jelle O Barentsz, Jonathan Richenberg, Richard Clements, Peter Choyke, Sadhna Verma, Geert Villeirs, Olivier Rouviere, Vibeke Logager, and Jurgen J Fütterer. ESUR prostate MR guidelines 2012. *European radiology*, 22(4):746–57, April 2012.

[6] Peter J Basser and Derek K Jones. Diffusion-tensor MRI: theory, experimental design and data analysis - a technical review. *NMR in biomedicine*, 15(7-8):456–67, 2002.

[7] Matthew D Blackledge, Martin O Leach, David J Collins, and Dow-Mu Koh. Computed Diffusion-weighted MR Imaging May Improve Tumor Detection. *Radiology*, 261(2):573–581, 2011.

[8] Roger M Bourne, Nyoman Kurniawan, Gary Cowin, Timothy Stait-Gardner, Paul Sved, Geoffrey Watson, Supriya Chowdhury, and William S Price. Biexponential diffusion decay in formalin-fixed prostate tissue: Preliminary findings. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, 000:1–6, December 2011.

[9] Canadian Cancer Societys Advisory Committee on Cancer Statistics. *Canadian Cancer Statistics 2013*. Canadian Cancer Society, Toronto, ON, 2013.

[10] Canadian Cancer Societys Steering Committee on Cancer Statistics. *Canadian Cancer Statistics 2012*. Canadian Cancer Society, Toronto, ON, 2012.

[11] Peter R. Carroll. Early detection of prostate cancer. *Commun Oncol*, 7:25 – 27, 2010.

[12] Ian Chan, William Wells, Robert V. Mulkern, Steven Haker, Jianqing Zhang, Kelly H. Zou, Stephan E. Maier, and Clare M. C. Tempany. Detection of prostate cancer by integration of line-scan diffusion, T2-mapping and T2-weighted magnetic resonance imaging; a multichannel statistical classifier. *Medical Physics*, 30(9):2390, 2003.

[13] Y Chen, W Guo, Q Zeng, X Yan, F Huang, H Zhang, G He, B C Vemuri, and Y Liu. Estimation , Smoothing , and Characterization of Apparent Diffusion Coefficient Profiles from High Angular Resolution DWI. *Engineering*, 00(C):0–5, 2004.

[14] Angelo M De Marzo, Elizabeth a Platz, Siobhan Sutcliffe, Jianfeng Xu, Henrik Grönberg, Charles G Drake, Yasutomo Nakai, William B Isaacs, and William G Nelson. Inflammation in prostate carcinogenesis. *Nature reviews. Cancer*, 7(4):256–69, April 2007.

[15] Scott Doyle, Mark Hwang, Kinsuk Shah, and Anant Madabhushi. Automated grading of prostate cancer using architectural and textural image features. In *4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1284–1287, 2007.

[16] Scott Doyle, Anant Madabhushi, Michael Feldman, and John Tomaszeweski. A boosting cascade for automated detection of prostate cancer from digitized histology. *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, 9(Pt 2):504–11, January 2006.

[17] JAD Farrell and SA Smith. High bvalue qspace diffusionweighted MRI of the human cervical spinal cord in vivo: Feasibility and application to multiple sclerosis. *Magnetic*, 59(5):1079–1089, 2008.

[18] Rafael C. Gonzales and Richard E. Woods. *Digital Image Processing*. Prentice Hall, Upper Saddle River, NJ, 3rd edition, 2007.

[19] Masoom A. Haider, Theodorus H van der Kwast, Jeff Tanguay, Andrew J Evans, Ali-Tahir Hashmi, Gina Lockwood, and John Trachtenberg. Combined T2-weighted and diffusion-weighted MRI for localization of prostate cancer. *AJR. American journal of roentgenology*, 189(2):323–8, August 2007.

[20] L O Hall, a M Bensaid, L P Clarke, R P Velthuizen, M S Silbiger, and J C Bezdek. A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain. *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council*, 3(5):672–82, January 1992.

[21] WK Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[22] Keyanoosh Hosseinzadeh and Samuel David Schwarz. Endorectal diffusion-weighted imaging in prostate cancer to differentiate malignant and benign peripheral zone tissue. *Journal of magnetic resonance imaging : JMRI*, 20(4):654–61, October 2004.

[23] Kazuhiro Katahira, Taro Takahara, Thomas C Kwee, Seitaro Oda, Yasuko Suzuki, Shoji Morishita, Kosuke Kitani, Yasuyuki Hamada, Mitsuhiko Kitaoka, and Yasuyuki Yamashita. Ultra-high-b-value diffusion-weighted MR imaging for the detection of prostate cancer: evaluation in 201 cases with histopathological correlation. *European radiology*, 21(1):188–96, January 2011.

[24] Chan Kyo Kim, Byung Kwan Park, and Bohyun Kim. High-b-value diffusion-weighted imaging at 3 T to detect prostate cancer: comparisons between b values of 1,000 and 2,000 s/mm2. *AJR. American journal of roentgenology*, 194(1):W33–7, January 2010.

[25] Piotr Kozlowski, Silvia D Chang, Edward C Jones, Kenneth W Berean, Henry Chen, and S Larry Goldenberg. Combined diffusion-weighted and dynamic contrast-enhanced MRI for prostate cancer diagnosis–correlation with biopsy and histopathology. *Journal of magnetic resonance imaging : JMRI*, 24(1):108–13, July 2006.

[26] Deanna L. Langer, Theodorus H van der Kwast, Andrew J Evans, John Trachtenberg, Brian C Wilson, and Masoom A. Haider. Prostate cancer detection with multiparametric MRI: logistic regression analysis of quantitative T2, diffusion-weighted imaging, and dynamic contrast-enhanced MRI. *Journal of magnetic resonance imaging : JMRI*, 30(2):327–34, August 2009.

[27] Xin Liu, DL Langer, Masoom A. Haider, and Yongyi Yang. Prostate cancer segmentation with simultaneous estimation of Markov random field parameters and class. *Medical Imaging,*, 28(6):906–915, 2009.

[28] Anant Madabhushi, Michael D Feldman, Dimitris N Metaxas, John Tomaszeweski, and Deborah Chute. Automated detection of prostatic adenocarcinoma from high-resolution ex vivo MRI. *IEEE transactions on medical imaging*, 24(12):1611–25, December 2005.

[29] Anant Madabhushi, Jianbo Shi, and Mark Rosen. Graph embedding to improve supervised classification and novel class detection: application to prostate cancer. *Image Computing and*, (i):729–737, 2005.

[30] S E Maier, P Bogner, G Bajzik, H Mamata, Y Mamata, I Repa, F a Jolesz, and R V Mulkern. Normal brain and brain tumor: multicomponent apparent diffusion coefficient line scan imaging. *Radiology*, 219(3):842–9, June 2001.

[31] W Malina. On an extended Fisher criterion for feature selection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, C(5):611–614, 1981.

[32] F H Martini, M J Timmons, and R B Tallitsch. *Human Anatomy*. Pearson Benjamin Cummings, San Francisco, 7th editio edition, 2012.

[33] Jhimli Mitra, Robert Martí, Arnau Oliver, Xavier Lladó, Soumya Ghose, Joan C Vilanova, and Fabrice Meriaudeau. Prostate multimodality image registration based on B-splines and quadrature local energy. *International journal of computer assisted radiology and surgery*, 7(3):445–54, May 2012.

[34] T Morineau, X Morandi, N Le Moëllic, and P Jannin. A cognitive engineering framework for the specification of information requirements in medical imaging: application in image-guided neurosurgery. *International journal of computer assisted radiology and surgery*, 8(2):291–300, March 2013.

[35] Shivang Naik, Scott Doyle, and Shannon Agner. Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology. *: From Nano to*, (c):284–287, May 2008.

[36] Shivang Naik, Scott Doyle, Michael Feldman, and John Tomaszewski. Gland segmentation and computerized gleason grading of prostate histology by integrating low-, high-level and domain specific information. *MIAAB*, pages 1–8, 2007.

[37] Richard Nock and Frank Nielsen. Statistical region merging. *IEEE transactions on pattern analysis and machine intelligence*, 26(11):1452–8, November 2004.

[38] S Ozer, Masoom A. Haider, and DL Langer. Prostate cancer localization with multispectral MRI based on relevance vector machines. In *2009 IEEE International Symposium on Biomedical Imaging From Nano to Macro*, pages 73–76, 2009.

[39] Sedat Ozer, Deanna L. Langer, Xin Liu, Masoom a. Haider, Theodorus H. van der Kwast, Andrew J. Evans, Yongyi Yang, Miles N. Wernick, and Imam S. Yetik. Supervised and unsupervised methods for prostate cancer segmentation with multispectral MRI. *Medical Physics*, 37(4):1873–1883, 2010.

[40] Y Peng, Y Jiang, C Yang, JB Brown, T Antic, I Sethi, C Schmid-Tannwald, ML Giger, SE Eggener, and A Oto. Quantitative Analysis of Multiparametric Prostate MR Images: Differentiation between Prostate Cancer and Normal Tissue and Correlation with Gleason ScoreA Computer-aided Diagnosis Development Study. *Radiology*, 267(3):787–796, 2013.

[41] M Röthke, D Blondin, H-P Schlemmer, and T Franiel. [PI-RADS classification: structured reporting for MRI of the prostate]. *RöFo : Fortschritte auf dem Gebiete der Röntgenstrahlen und der Nuklearmedizin*, 185(3):253–61, March 2013.

[42] S Ruan, S Lebonvallet, and A Merabet. Tumor segmentation from a multispectral MRI images by using support vector machine classification. *Imaging: From Nano*, 2(1):1236–1239, 2007.

[43] Fritz H Schröder, Jonas Hugosson, Monique J Roobol, Teuvo L J Tammela, Stefano Ciatto, Vera Nelen, Maciej Kwiatkowski, Marcos Lujan, Hans Lilja, Marco Zappa, Louis J Denis, Franz Recker, Antonio Berenguer, Liisa Määttänen, Chris H Bangma, Gunnar Aus, Arnauld Villers, Xavier Rebillard, Theodorus van der Kwast, Bert G Blijenberg, Sue M Moss, Harry J de Koning, and Anssi Auvinen. Screening and prostate-cancer mortality in a randomized European study. *The New England journal of medicine*, 360(13):1320–8, March 2009.

[44] Anjali Sehrawat, Kenji Shimada, and Yoed Rabin. Generating prostate models by means of geometric deformation with application to computerized training of cryosurgery. *International journal of computer assisted radiology and surgery*, pages 301–312, 2012.

[45] H S Seo, K-H Chang, D G Na, B J Kwon, and D H Lee. High b-value diffusion (b = 3000 s/mm2) MR imaging in cerebral gliomas at 3T: visual and quantitative comparisons with b = 1000 s/mm2. *AJNR. American journal of neuroradiology*, 29(3):458–63, March 2008.

[46] D Shier, J Butler, and R Lewis. *Hole's Essentials of Human Anatomy and Physiology*. WCB McGraw-Hill, Boston, 6th editio edition, 1998.

[47] G N Srinivasan and G Shobha. Statistical Texture Analysis. 36(December):1264–1269, 2008.

[48] Ian M. Thompson, Donna K. Pauler, Phyllis J. Goodman, Catherine M. Tangen, M. Scott Lucia, Howard L. Parnes, Lori M. Minasian, Leslie G. Ford, Scott M. Lippman, E. David Crawford, John J. Crowley, and Charles A. Coltman Jr. Prevalence of prostate cancer among men with a prostate-specific antigen level ≤ 4.0 ng per mililiter. *N Engl J Med*, 350:2239 – 2246, 2004.

[49] Simon Walker-Samuel, Matthew Orton, Jessica K R Boult, and Simon P Robinson. Improving apparent diffusion coefficient estimates and elucidating tumor heterogeneity using Bayesian adaptive smoothing. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, 65(2):438–47, February 2011.

[50] Simon Walker-Samuel, Matthew Orton, Lesley D McPhail, and Simon P Robinson. Robust estimation of the apparent diffusion coefficient (ADC) in heterogeneous solid tumors. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, 62(2):420–9, August 2009.

[51] Y C Weon, J H Kim, J S Lee, and S Y Kim. Optimal diffusion-weighted imaging protocol for lesion detection in transient global amnesia. *AJNR. American journal of neuroradiology*, 29(7):1324–8, August 2008.

[52] Alexander Wong. Constrained Bayesian streak artifact reduction approach for contrast enhanced computed tomography imaging of the intervertebral disc. *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine*

and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference, 2011:8487–90, August 2011.

[53] Alexander Wong, Jeffrey Glaister, Andrew Cameron, and Masoom Haider. Correlated diffusion imaging. *BMC medical imaging*, 13:26, January 2013.

[54] Alexander Wong, Akshaya Mishra, Justin Yates, Paul Fieguth, David a Clausi, and Jack P Callaghan. Intervertebral disc segmentation and volumetric reconstruction from peripheral quantitative computed tomography imaging. *IEEE transactions on bio-medical engineering*, 56(11 Pt 2):2748–51, November 2009.

[55] Alexander Wong, Akshaya Mishra, Wen Zhang, Paul Fieguth, and David a. Clausi. Stochastic image denoising based on Markov-chain Monte Carlo sampling. *Signal Processing*, 91(8):2112–2120, August 2011.