Computer-assisted lesion classification and intervention planning for prostate cancer

Ryan M. Alfano, The University of Western Ontario

Supervisor: Ward, Aaron D., The University of Western Ontario
A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Medical Biophysics
© Ryan M. Alfano 2021

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

Part of the Medical Biophysics Commons

Recommended Citation
https://ir.lib.uwo.ca/etd/7831

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 wlsadmin@uwo.ca.
Abstract

Diagnosis of prostate cancer (PCa) is commonly performed using trans-rectal ultrasound (TRUS)-guided biopsy but is limited by the detection accuracy of 23-53%. Multi-parametric magnetic resonance imaging (mp-MRI) is emerging as a useful tool for classifying PCa, improving disease localization in comparison to TRUS-guided biopsy. However, mp-MRI suffers from two major limitations: (1) complex, multi-dimensional signals make interpretation challenging and (2) inter-observer variability of lesion classification between physicians. Critically needed are methods developed for augmenting the interpretability of mp-MRI to assist in lesion classification. To meet this need, we leveraged a patient cohort of men imaged prior to prostatectomy, where pathologist-annotated transverse histology was registered to in-vivo mp-MRI with a measured target registration error. We developed a radiomics-based machine learning model trained on mapped pathology annotations for PCa vs. non-PCa, and found that a 5-feature Naive Bayes classifier classified these two types of regions in leave-one-patient-out cross-validation with an area under the receiver operating characteristic curve of 0.80. We then investigated augmentation of mp-MRI interpretability via additional imaging using prostate-specific membrane antigen positron emission tomography (PSMA-PET). Using a patient cohort of 12 men imaged on a PET/MRI hybrid scanner, we investigated segmentation recommendations for classifying the dominant intraprostatic lesion (DIL). For focal therapy, we found that a threshold of 67% \(\text{SUV}_{\text{max}}\) and an 8.4 mm margin produced mean voxel-wise sensitivity of 95% with mean specificity 76%. For guided biopsy, we found that a threshold of 81% \(\text{SUV}_{\text{max}}\) and a 5.2 mm margin produced a mean sensitivity of 65% and mean specificity of 95%. Furthermore, using these DIL contours generated on PSMA-PET, we assessed the effect of boosting high-dose-rate brachytherapy treatment plans towards these segmentations in comparison to whole-gland plans. We found that dose to the underlying high-grade cancer was significantly increased in targeted plans, compared to whole-gland plans, while maintaining dose constraints to critical structures. To conclude, we classified PCa vs. non-PCa on mp-MRI using machine learning and produced segmentation recommendations for delineat-
ing PSMA-PET DILs for various clinical applications. Automatically generated PSMA-PET contours may be used to boost brachytherapy treatment plans to increase dose to underlying high-grade cancer.

**Keywords:** Prostate cancer, multi-parametric magnetic resonance imaging, machine learning, radiomics, prostate specific membrane antigen, positron emission tomography, dominant intraprostatic lesion, classification, high-dose rate brachytherapy, focal boosting
Summary for Lay Audience

Prostate cancer (PCa) is the most common non-skin cancer in men in Canada. PCa is commonly diagnosed by extracting prostate tissue through biopsy needles which are guided with ultrasound, known as trans-rectal ultrasound (TRUS)-guided biopsy. However, TRUS-guided biopsy isn’t very accurate, and many cancers may be missed. Multi-parametric magnetic resonance imaging (mp-MRI) is another form of imaging that has shown to improve detection of PCa. However, mp-MRI has two major challenges: (1) it is complex and thus even expert physicians have difficulty analyzing the images, and (2) there is a lot of variability between interpretations made by different physicians for the same image. Critically needed are methods assisting physicians on interpreting mp-MRI to assist in classifying PCa. To meet these needs, we use a dataset of men imaged with mp-MRI prior to surgical prostate removal. These prostates were sent to pathology where they were sliced, and cancer was annotated by the pathologists. These histology slices were mapped back to the pre-surgery mp-MRI so we could determine the true location of the underlying disease. We developed a machine learning model that was able to accurately distinguish PCa from non-PCa based on the appearance of the cancer on mpMRI. Furthermore, we investigated another method of assisting physicians with mp-MRI interpretation using additional imaging known prostate-specific membrane antigen positron emission tomography (PSMA-PET). Using a subset of 12 patients imaged on a PET/MRI scanner that acquired mp-MRI and PET images at the same time, we investigated recommendations for contouring the most threatening lesion in the prostate, known as the dominant intraprostatic lesion. Furthermore, using these contours generated on PSMA-PET, we aimed to determine the effect of focusing dose in brachytherapy treatment plans towards these contours in comparison to what is commonly performed in the clinic, namely whole-gland plans. We found that the highly aggressive and life-threatening cancer received a significantly increased dose in these targeted plans, compared to whole-gland plans, which may provide improved cancer treatment.
Co-Authorship

This thesis is presented in an integrated article format with each of the following chapters based on publications that are either in-preparation, under review or accepted for publication.


My contribution to this work included designing and conducting of all experiments, analyzing, and interpreting the experimental results and drafting the manuscript. Glenn Bauman was the principal investigator of the grant funding the Image Guidance for Prostate Cancer trial for which all imaging and histology data was obtained. Mena Gaed, Madeleine Moussa, and Jose A Gomez-Lemus contributed to processing and annotating all prostate histology. Joseph Chin and Stephen Pautler recruited all subjects to be part of the clinical trial and performed the radical prostatectomies for all patients. Aaron D. Ward contributed to defining the research question, analyzing and interpreting the results, and preparation and drafting of the manuscript. All authors helped in reviewing the manuscript. This work was performed under the supervision of Aaron D. Ward.


My contribution to this work included designing and conducting of all experiments, analyzing and interpreting the experimental results and drafting the manuscript. Glenn Bauman was the principal investigator of the grant funding the Image Guidance for Prostate Cancer trial for which all imaging and histology data was obtained. Wei Liu contributed to the manual contouring of prostate lesions on PET imaging used for analysis. Jonathan D. Thiessen contributed to the analysis and preparation of PET imaging data. Irina Rachinsky and William...
Pavlosky were responsible for acquiring PET images of all subjects. John Butler contributed to managing and organizing imaging data from all subjects. Mena Gaed, Madeleine Moussa, and Jose A Gomez-Lemus contributed to processing and annotating all prostate histology. Joseph Chin and Stephen Pautler recruited all subjects to be part of the clinical trial and performed the radical prostatectomies for all patients. Aaron D. Ward contributed to defining the research question, analyzing and interpreting the results, and preparation and drafting of the manuscript. All authors helped in reviewing the manuscript. This work was performed under the supervision of Aaron D. Ward.

Chapter 4: Ryan Alfano, Christopher Smith, Douglas Hoover, Kathleen Surry, David D’Souza, Jonathan D. Thiessen, Irina Rachinsky, John Butler, Jose A Gomez-Lemus, Mena Gaed, Madeleine Moussa, Joseph Chin, Stephen Pautler, Aaron D. Ward, “Prostate specific membrane antigen positron emission tomography (PSMA-PET) for lesion-directed prostatic high-dose-rate brachytherapy dose escalation” (under review)

This work was co-first-authored by myself and Christopher Smith. My contributions to this work included developing guidelines for segmentation, generating the segmentations from the PSMA-PET images, performing the PET/MRI to histology registrations, contributing to the PET/MRI to TRUS registrations, contributing to dose optimizations on some of the treatment plans, analyzing and interpreting the experimental results, and preparation and drafting of the manuscript. Christopher Smith contributed to performing the PET/MRI to TRUS registrations, performing dose optimizations on treatment plans, performing dose analysis following optimizations, analyzing, and interpreting the experimental results, and preparation and drafting of the manuscript. Douglas Hoover and Kathleen Surry contributed to the experimental design and analysis of results. David D’Souza contributed to the creation of brachytherapy treatment plans and the experimental design. Jonathan D. Thiessen contributed to the analysis and preparation of PET imaging data. Irina Rachinsky was responsible for acquiring PET images of all subjects. John Butler contributed to managing and organizing imaging data from all subjects. Mena Gaed, Madeleine Moussa, and Jose A Gomez-Lemus contributed to processing and an-
notating all prostate histology. Joseph Chin and Stephen Pautler recruited all subjects to be part of the clinical trial and performed the radical prostatectomies for all patients. Aaron D. Ward contributed to defining the research question, analyzing and interpreting the results, and preparation and drafting of the manuscript. All authors helped in reviewing the manuscript. This work was performed under the supervision of Aaron D. Ward.
Acknowledgements

The completion of this thesis was supported by many individuals over the past five years.

First and foremost, I would like to give thanks to my supervisor, Dr. Aaron Ward, for mentoring and supporting me throughout my graduate career. Aaron, your wisdom and leadership has shaped me into the scientist that I have always aspired to become. Words cannot begin to describe the impact that you have had on my life, both personally and professionally. I am forever indebted to you for the amount of time and effort you have spent helping to develop my academic and professional skills as I move forward in my career.

To my advisory committee, Dr. Glenn Bauman, Dr. Ali Khan, and Dr. Aaron Fenster, thank you for the support you have provided over the course of my PhD. Your insightful comments and suggestions have helped pave the way for the work in this thesis. To my examining committee, Dr. Stewart Gaede, Dr. Robert Bartha, Dr. Derek Cool and Dr. Dean Barratt, thank you for taking the time out of your day to examine me on my thesis.

To all collaborators from the Image Guidance in Prostate Cancer team from radiology, pathology, radiation oncology, surgical oncology, and genomics, thank you for your continued support and interest in this work. None of this would have been possible without your help. Special thanks goes out to Dr. José Gómez, Dr. Mena Gaed and Cathie Crukley, for making time spent performing specimen processing much more enjoyable.

My sincere gratitude goes out to all former and present members of the Baines Imaging Laboratory I had the privilege to work with: Dr. Sarah Mattonen, Dr. Peter Martin, Dr. Yiwen Xu, Dr. Wenchao Han, Dr. David Palma, Salma Dammak, Christopher Smith, Laurie Huang, David DeVries, Sean Pentinga, Derek Soetemans, Carol Johnson and Andrew Warner. I will never forget the friendship and support you have given me throughout the course of my PhD. It has been a pleasure to get to work with all of you, and I wish you the best in your future. To Christopher Smith, it has been a wonderful experience being able to work on a co-author project with you. Thank you for the time and effort you spent working with me to complete this project and demonstrating to others what successful collaboration can look like.

viii
There are many people outside of work life that have been a foundation of support throughout my graduate career. To my parents, Danny and Faye Alfano, my grandmother, Marion Frank, and my grandparents, Enzo and Mary Alfano, thank you, from the bottom of my heart, for the love and kindness you have always given me. Throughout all the ups and downs of graduate school, I took for granted the foundation of support that you have unconditionally provided. You have helped shape me into the person I am today. To my sister, Carly Alfano, I want to thank you for the short time we got to live together during the later half of my graduate career. Your hard-work and determination to succeed has always motivated me to push new boundaries in an effort to better myself. Thank you for putting up with me and I am so happy that we were able to share this experience with each other.

Last, but not least, I would like to give my utmost gratitude to my longtime girlfriend, Kristen Germani. For all of the support you provided me when times were tough, all of the long-distance car-ride visits, and all of the experiences we got to share in the past 7 years, you have had a more profound effect on my life than you know. You have always been my rock - the one I can count on for help and trust with anything in life. I am so excited that we get to share a workplace together in my next career step and I look forward to whatever the future has in store for us.

There have been many sources of funding that have supported the work in this thesis including the Canadian Institutes of Health Research, the Ontario Institute for Cancer Research, Cancer Care Ontario, the National Sciences and Engineering Research Council of Canada, and the Western Graduate Research Scholarship.
## Contents

Certificate of Examination ............................................. ii  
Abstract ........................................................................ ii  
Summary for Lay Audience ........................................ iv  
Co-Authorship ................................................................ vii  
Acknowledgements .......................................................... viii  
List of Figures ................................................................... xv  
List of Tables .................................................................... xix  
List of Appendices .......................................................... xxi  
List of Abbreviations, Symbols, and Nomenclature .......... xxii  

### 1 Introduction

1.1 Prostate cancer epidemiology ........................................ 1  
1.2 Imaging PCa ............................................................... 2  
  1.2.1 Applications of imaging in PCa ................................. 2  
  1.2.2 Multi-parametric magnetic resonance imaging ........ 3  
  1.2.3 Prostate specific membrane antigen positron emission tomography ......... 6  
  1.2.4 Ultrasound ............................................................ 7  
  1.2.5 Computed tomography ............................................ 8  

x
2.2 Methods and materials ................................................. 55
   2.2.1 Patient characteristics .......................................... 55
   2.2.2 MR imaging ....................................................... 55
   2.2.3 Preprocessing .................................................... 56
   2.2.4 Reference standard ............................................... 56
   2.2.5 Regions of interest ............................................... 57
   2.2.6 Feature extraction and scaling .................................. 60
   2.2.7 Feature selection ................................................ 60
   2.2.8 Classification .................................................... 61
   2.2.9 Inter-class area mismatch ....................................... 61
   2.2.10 Inter-class shape mismatch .................................... 63
   2.2.11 Inter-scanner differences ..................................... 63
   2.2.12 Inter-zonal differences ....................................... 63
2.3 Results ................................................................. 63
2.4 Discussion ............................................................ 69
2.5 Conclusions ........................................................... 74
References ................................................................. 75

3 Histologic validation of auto-contoured dominant intraprostatic lesions on $^{18}$F-
DCFPyL PSMA-PET imaging .................................................. 81
   3.1 Introduction .......................................................... 81
   3.2 Methods and materials .............................................. 82
      3.2.1 Patient Population .............................................. 82
      3.2.2 Registering Histology to Imaging ............................ 84
      3.2.3 Thresholding and Sensitivity/Specificity Analysis ....... 84
   3.3 Results ............................................................... 85
   3.4 Discussion ............................................................ 91
   3.5 Conclusions .......................................................... 95
4 \textbf{\textsuperscript{18}F-DCFPyL PSMA-PET for lesion-directed prostatic high-dose-rate brachytherapy dose escalation}} \hspace{1cm} 101

4.1 Introduction \hspace{1cm} 101

4.2 Methods and materials \hspace{1cm} 103

4.2.1 PSMA-PET/Histology Patient Population \hspace{1cm} 103

4.2.2 Histology to PET/MRI Registration \hspace{1cm} 104

4.2.3 Brachytherapy Population \hspace{1cm} 105

4.2.4 PET/MRI-TRUS Registration \hspace{1cm} 105

4.2.5 Standard Brachytherapy Planning \hspace{1cm} 105

4.2.6 Focal Targeting Treatment Planning \hspace{1cm} 106

4.2.7 Dose Analysis \hspace{1cm} 107

4.2.8 Statistics \hspace{1cm} 107

4.3 Results \hspace{1cm} 108

4.4 Discussion \hspace{1cm} 109

4.5 Conclusions \hspace{1cm} 114

References \hspace{1cm} 115

5 Conclusions and future work \hspace{1cm} 121

5.1 Project summary and conclusions \hspace{1cm} 121

5.1.1 Achievement of research objectives and testing hypotheses \hspace{1cm} 122

5.1.2 Advances in knowledge and technology \hspace{1cm} 125

5.2 Suggestions for future work \hspace{1cm} 128

5.2.1 Investigating the histopathology of MR-invisible lesions \hspace{1cm} 128

5.2.2 Determining optimal PSMA-PET DIL segmentation criteria for maximizing dose to underlying high-grade disease \hspace{1cm} 129

References \hspace{1cm} 131
List of Figures

1.1 Simple model of a neuron. .......................................................... 21
1.2 Architecture of a simple CNN with two output classes. ............... 22
1.3 Variability in lesion localization across three different pulse sequences (each column) of the same slice and same patient for three different radiologists (each row). ......................................................... 23

2.1 Hybrid pseudocode and graphical illustration of the algorithm used to select non-PCa tissue masks. The algorithm takes a list C containing PCa tissue masks cropped to the bounding box of the PCa tissue and iterates them over a list M of masks, in image space, containing non-PCa tissue to produce a list N of non-PCa tissue masks. This algorithm is run once in the PZ and once in the CG. This algorithm thus generates two lists that contain non-PCa tissue masks in each zone respectively. ............................................. 58

2.2 Flow diagram of the experimental pipeline. The red arrows indicate the pathology derived contours of PCa being mapped to the in-vivo imaging. No ROIs were manually drawn on the imaging sequences. Feature selection was performed on 15 patients (≈25%, 15/61) with cross-validation and results reporting being performed on the remaining 46 patients. ADC: apparent diffusion co-efficient map, T2W: T2-weighted, PCa: prostate cancer. ......................... 62
2.3 Receiver operating characteristic (ROC) curve (left) and a precision-recall ROC curve (right) for classification of PCa vs. non-PCa tissue in a LOPO cross validation experiment.  64

2.4 From left to right T2W images, ADC images and contoured histology. On imaging, red regions represent true negatives, green represent true positives, yellow represent false negatives and blue represent false positives. On histology, green contours represent Gleason 3+3 cancer, dark yellow represent Gleason 4+4 cancer, and gray represent prostatic intraepithelial neoplasia.  67

2.5 AUC calculated in a 10-fold cross validation experiment for shape-matched and non-shape-matched non-PCa tissue masks. Shape-matched non-PCa tissue masks refer to benign tissue masks that have the same shape as PCa tissue masks through the algorithm defined in Figure 2.1. Non-shape-matched non-PCa tissue masks refer to benign tissue masks that have been transformed into area-matched regular circles instead of their PCa shaped tissue masks. Boxplot displays the median and 10th to 90th percentile of AUC values.  69

3.1 Contour plots of average sensitivity (left) and average specificity (right) for segmentations on SUV-corrected PET corresponding to co-registered histology.  86

3.2 Qualitative results of segmentations using a threshold of 67% of SUVmax with an 8.4 mm margin. Each oblique histology plane is numbered by the patient number and the slice number separated by a hyphen. The dominant intra-prostatic lesion defined by histology is shaded in yellow in front of the SUV-converted PET image. The segmentation outline crossing the oblique histologic plane is coloured in red. The prostate boundary is outlined in a white dashed curve. Each image is intensity scaled based on the colour bar adjacent to it. Slices of the same patient are grouped by the black boxes.  88
3.3 Qualitative results of segmentations using a threshold of 81% SUVmax with a 5.2 mm margin. Each oblique histology plane is numbered by the patient number and the slice number separated by a hyphen. The dominant intra-prostatic lesion defined by histology is shaded in yellow in front of the SUV converted PET image. The segmentation outline crossing the oblique histologic plane is coloured in red. The prostate boundary is outlined in a white dashed curve. Each image is intensity scaled based on the colour bar adjacent to it. Slices of the same patient are grouped by the black boxes.

4.1 Flow diagram illustrating the PSMA-PET and corresponding histology lesions being registered into the simulated HDR-BT treatment plans and corresponding PSMA-PET targeted treatment plans. The targeted plans were designed to target the PSMA-PET lesions, with the optimizer blinded to the location(s) of the histologic lesion(s). The dose to the histology was calculated for the standard plans and targeted plans.

4.2 Box-and-whisker plot of the (left) $D_{98}$, (middle) $D_{90}$, and (right) mean doses delivered to the digital histology annotations within PSMA-PET targeted treatment plans and standard treatment plans. The whiskers represent the 10-90th percentiles.

4.3 Line graph displaying the percentage of treatment plans that delivered a given $D_{98}$ to the histology cancer. PET-targeted treatment plans are colored in blue and standard treatment plans are colored in magenta.
4.4 An oblique in-vivo image of T2W MRI in the background and PSMA-PET (red; SUV\textsubscript{max} = 27.62) in the foreground (a) for the corresponding (b) histologic section with primary grades listed in the legend. Presented subsequently is the corresponding dose matrix of this oblique slice in the TRUS imaging space using (c) focal targeting and (d) standard treatment. The prostate outline is delineated in purple. Histologically defined cancer is outlined and colored in blue while the PSMA-PET generated segmentation is outlined in black, both of which having been deformably registered to the TRUS space from MRI. White crosses represent the location of inserted catheters within the prostate. Color bar scale is in units of Gy scaled from 10-30 Gy.

5.1 An apparent diffusion coefficient map, T2W image and corresponding pathology annotations made on histology for a patient with low-grade disease with limited visibility on imaging.
List of Tables

1.1 PI-RADS v2.1 description for each score provided on a per lesion basis. 4
1.2 Summary of previous work done on PCa classification on mp-MRI. 25
1.3 Summary of previous work done on PCa classification on PSMA-PET. 27

2.1 Summary of error metrics for LOPO and 10-fold cross validation for a 5-feature NAIVEB classifier. 64
2.2 Features selected from the 15-patient hold-out set used in LOPO and 10-fold cross-validation. Features were sequentially added using greedy forward feature selection by taking the set with the largest measured sum of estimated mahalanobis distances. Directionality refers to the direction of the target voxel from the reference in GLCM calculation. GLCM and GLRLM direction vectors are as follows: 0°: [1,0], 45°: [1,1], 90°: [0,1], 135°: [-1,1]. Bolded feature values were found to be significantly different between PCa vs. non-PCa tissue (p<0.05). 65
2.3 Median AUC as a function of percent difference from baseline for multiple cross validation schemes (10-fold, 4-fold and LOPO). Experiment was performed by eroding non-PCa tissue masks (top) and PCa tissue masks (bottom) iteratively. 66
2.4 2-Fold cross validation experiment of PCa vs. non-PCa tissue with each fold being trained on data from one scanner and tested on the other (left: scanner split) and an equal ratio of scanner data in each fold (right: equal distribution). 69
2.5 10-Fold cross validation results for training and testing performed with regions of interest located only in the PZ (left) and CG (right). Error metrics displayed are reported with the median across all 10-folds and the interquartile range.

3.1 Patient information for subjects involved in this analysis.

3.2 Imaging information for each patient analyzed in this study.

3.3 Reported results for focal therapy and guided biopsy applications. Data are reported for each oblique histologic section labeled by patient ID and slice number separated by a hyphen.

4.1 Dose constraints followed during the standard and PSMA targeted HDR-BT treatment planning. Vxx% is the volume percentage of the structure that received xx% dose of the prescription dose, Dxx% is the prescription dose percentage that xx% of the structure received, and Dxxcc is the maximum prescription dose percentage permitted to xx cubic centimeters of volume in that structure.
List of Appendices

Appendix A: Imaging and histology data and PET/MRI to histology registration details
   for Chapter 3 ................................................................. 138

Appendix B: Imaging and histology data and PET/MRI to TRUS registration details for
   Chapter 4 ................................................................. 140
List of Abbreviations, Symbols, and Nomenclature

ADC  Apparent diffusion co-efficient
AUC  Area under the receiver operating characteristic curve
CAD  Computer aided diagnosis
CG   Central gland
CNN  Convolutional neural network
CT   Computed tomography
DCE  Dynamic contrast enhanced
DIL  Dominant intraprostatic lesion
DRE  Digital rectal examination
DWI  Diffusion weighted imaging
EBRT External beam radiotherapy
ECE  Extra-capsular extension
FDG  Fluorodeoxyglucose
FNR  False-negative rate
FPR  False-positive rate
GLCM Gray-level co-occurrence matrix
GLRLM Gray-level run-length matrix
GRPR Gastrin-releasing peptide receptor
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gy</td>
<td>Gray</td>
</tr>
<tr>
<td>HDR-BT</td>
<td>High-dose-rate brachytherapy</td>
</tr>
<tr>
<td>ISUP</td>
<td>International Society of Urological Pathology</td>
</tr>
<tr>
<td>KNN</td>
<td>K-nearest neighbour</td>
</tr>
<tr>
<td>LDA</td>
<td>Linear discriminant analysis</td>
</tr>
<tr>
<td>LOGL</td>
<td>Logistic linear</td>
</tr>
<tr>
<td>LOPO</td>
<td>Leave-one-patient-out</td>
</tr>
<tr>
<td>MCR</td>
<td>Misclassification rate</td>
</tr>
<tr>
<td>Mp-MRI</td>
<td>Multi-parametric magnetic resonance imaging</td>
</tr>
<tr>
<td>MRI</td>
<td>Magnetic resonance imaging</td>
</tr>
<tr>
<td>NAIVEB</td>
<td>Naive-Bayes</td>
</tr>
<tr>
<td>NCCN</td>
<td>National Comprehensive Cancer Network</td>
</tr>
<tr>
<td>Non-PCa</td>
<td>Non prostate cancer (benign tissue)</td>
</tr>
<tr>
<td>PCa</td>
<td>Prostate cancer</td>
</tr>
<tr>
<td>PET</td>
<td>Positron emission tomography</td>
</tr>
<tr>
<td>PI-RADS</td>
<td>Prostate imaging - reporting and data system</td>
</tr>
<tr>
<td>PSA</td>
<td>Prostate specific antigen</td>
</tr>
<tr>
<td>PSMA</td>
<td>Prostate specific membrane antigen</td>
</tr>
<tr>
<td>PZ</td>
<td>Peripheral zone</td>
</tr>
<tr>
<td>RF</td>
<td>Random forest</td>
</tr>
<tr>
<td>ROC</td>
<td>Receiver operating characteristic curve</td>
</tr>
<tr>
<td>ROI</td>
<td>Region of interest</td>
</tr>
<tr>
<td>SUV</td>
<td>Standardized uptake value</td>
</tr>
<tr>
<td>SVM</td>
<td>Support vector machine</td>
</tr>
<tr>
<td>T1W</td>
<td>T1-weighted</td>
</tr>
<tr>
<td>T2W</td>
<td>T2-weighted</td>
</tr>
<tr>
<td>TRUS</td>
<td>Transrectal ultrasound</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1 Prostate cancer epidemiology

Prostate cancer (PCa) is the most common non-cutaneous cancer amongst men in North America [1]. It is projected that there will be approximately 193,000 PCa cases by the end of 2020, accounting for 21% of all male cancer diagnoses [1]. Early diagnosis is crucial for survival as the 5-year survival rate for localized and/or regional disease is ≥99%, while the survival rate for distant disease drops to 30% [1].

PCa incidence is also highly variable depending on geographical location; variance in incidence rate amongst men prior to the introduction of the prostate specific antigen (PSA) screening test has been highly attributed to lifestyle factors associated with geographical regions [2]. Since PSA screening was introduced in 1986, an overall downward trend of mortality has been observed; relative reduction in risk of mortality was 20-21% compared to a control group with no screening [3, 4]. In combination with a digital rectal examination (DRE), the likelihood of finding disease with abnormal results from both tests was reported to be 20-25%, while the chance of missing cancer from a negative result was reported to be 10% [5]. The methodology of these tests is further described below in sections 1.3.1 and 1.3.2. Patients presenting with suspicious PSA and/or DRE findings are sent for 2D transrectal ultrasound (TRUS) guided
biopsy for diagnosis which targets biopsy needles primarily in the peripheral zone (PZ) as it has been shown to be the origin of 70% of all PCas (more information on TRUS-guided biopsy can be found in section 1.3.3) [6]. However, 2D TRUS biopsy has shown to miss clinically significant cancer, with a reported false-negative rate (FNR) in detection of 23% on a per-patient level [7]. Also, frequent repeat biopsies due to the low sensitivity of TRUS-guided biopsy have shown to potentially cause complications (e.g. post-biopsy hemorrhage) which increases as more biopsy cores are obtained [8]. The incorporation of alternative imaging during biopsy, such as magnetic resonance imaging (MRI), has improved cancer detection rates versus standard TRUS-guided biopsy (29.5% versus 9.8%, respectively) when trying to target cancer on each modality [8]. MRI-TRUS fusion systems have been shown to outperform 2D- and 3D-TRUS in terms of increased percent tumour involvement per core by 16% and 17%, respectively [9]. Thus, the rapid investigation of new techniques to further assist in improving PCa diagnosis began; with non-invasive imaging being one of the primary foci in recent years.

1.2 Imaging PCa

Imaging of PCa is an important field of investigation as it may provide a non-invasive method of disease detection, staging, treatment planning, and follow-up [10].

1.2.1 Applications of imaging in PCa

Imaging of PCa has been integrated into clinical practice with the role dependant on the risk-level of the patient. It has been shown that 1.6-38% of low-risk patients who undergo active surveillance (electing to not receive treatment) have been found to drop out and elect for curative treatment despite no signs of disease progression with primary reasons attributed to social and mental pressures [11]. This treatment option is discussed more thoroughly below in section 1.4.1. Routine imaging may provide more substantial evidence of indolent disease and further prevent unnecessary treatment, which has significant quality of life and economic
consequences [12]. Intermediate-risk patients with clinical intent to cure through radical radiation therapy or radical prostatectomy may benefit from the detection and/or localisation of extra-capsular extension (ECE), which refers to local tumour growth beyond the capsule of the prostate into local periprostatic soft tissues. This is because patients with cancer that has extended outside the capsule has been shown to have much worse 15-year overall survival (50%) than those with organ confined disease (63%) [13]. These treatment options are discussed more thoroughly in sections 1.4.2 and 1.4.3, respectively. Imaging of the prostate for intermediate- to high-risk disease has been shown to significantly predict extra-prostatic extension independent of other predictors such as Gleason score, clinical stage and prostate-specific antigen measurements [14]. Gleason score was a measure developed to rank PCa based on its aggressiveness by analyzing morphological patterns of the cellular components (i.e. stroma) (more information on Gleason score can be found in section 1.5.1). High-risk patients who undergo palliative treatment may benefit from imaging to detect distant skeletal or nodal metastases. Integration of new positron emission tomography (PET) techniques have shown to be more sensitive at detecting such metastases than the traditionally used bone scintigraphy [15].

1.2.2 Multi-parametric magnetic resonance imaging

Multi-parametric magnetic resonance imaging (mp-MRI) is an imaging procedure that involves the acquisition of multiple conventional and functional MR sequences to provide details on tissue anatomy and characteristics such as vasculature structure and diffusivity. MRI has the advantage of high soft tissue contrast when compared to other imaging modalities such as ultrasound and computed tomography (CT). Common MR sequences used in clinical practice for imaging PCa are T2-weighted (T2W) imaging, dynamic contrast enhanced (DCE) imaging, and diffusion weighted imaging (DWI) with standards set surrounding the acquisition of images from each pulse sequence [16]. The inclusion of mp-MRI in the clinic significantly increased the rate of disease localization in a preoperative setting when compared to TRUS guided biopsy and DRE [17]. Acquisition of mp-MRI prior to TRUS guided-biopsy has im-
proved detection of high-grade tumours in an equivalent or higher percentage of patients when compared to the clinical standard of systematic random sampling of tissue while also using fewer biopsy cores [18]. In order to diminish the variability in mp-MRI acquisition and interpretation, the Prostate Imaging - Reporting and Data System (PI-RADS) was developed to establish acceptable parameters for acquisition and a guideline for evaluating level of suspicion of disease presence [19]. Table 1.1 shows the description associated with each PI-RADS score that is assigned on a per lesion basis. After all lesions have been designated with a PI-RADS score, an overall patient score is assigned.

<table>
<thead>
<tr>
<th>PI-RADS Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PI-RADS 1</td>
<td>very low (clinically significant cancer is highly unlikely to be present)</td>
</tr>
<tr>
<td>PI-RADS 2</td>
<td>low (clinically significant cancer is unlikely to be present)</td>
</tr>
<tr>
<td>PI-RADS 3</td>
<td>intermediate (the presence of clinically significant cancer is equivocal)</td>
</tr>
<tr>
<td>PI-RADS 4</td>
<td>high (clinically significant cancer is likely to be present)</td>
</tr>
<tr>
<td>PI-RADS 5</td>
<td>very high (clinically significant cancer is highly likely to be present)</td>
</tr>
</tbody>
</table>

Table 1.1: PI-RADS v2.1 description for each score provided on a per lesion basis.

However, mp-MRI has some limitations in terms of sensitivity and specificity when detecting lesions, which has been shown to be 78-94% and 37-88% for detection of tumours within the whole gland, respectively [20, 21]. In terms of sensitivity, it has been shown that for T2W and ADC images there is no difference between the signal generated in sparse tumours and normal tissue [22]. Furthermore, it has been shown that 31% of clinically significant lesions (Gleason grade 3+4 and 4+4) were missed entirely upon second review by radiologists [23], suggesting that there exists a subset of lesions that are truly invisible. In terms of specificity, certain conditions within the prostate have been known to mimic the appearance of cancerous lesions on T2W imaging, such as chronic prostatitis, post-biopsy hemorrhage, and inflammation [24], contributing to a 37% specificity of lesion detection [21].
T2W proton MRI is a conventional imaging technique that interrogates the transverse relaxation of precessing protons after an emitted radio-frequency pulse. It is widely regarded as one of the most useful sequences for detecting ECE due to its rich anatomical detail [25]. T2W images also provide detail in the prostate zonal anatomy, which includes the PZ spanning the posterior side and the transition zone which spans the anterior side and includes the central gland (CG). Tumours located in the PZ tend to appear as ill-defined hypointense regions whereas tumours in the CG tend to be non-circumscribed, homogeneous regions with moderate hypointensity on this imaging sequence when compared to benign tissue [26]. Increased strength of this hypo-intensity has also been shown to be associated with higher Gleason grade [27]. Details on PCa pathology are described further below in section 1.5.1.

DCE MRI is a functional imaging technique that is acquired with T1W MRI. In a DCE-MRI scan, the patient is imaged using T1W MRI prior, during and after the injection of a paramagnetic contrast agent (typically a gadolinium-based contrast agent) that is visible on MRI. These images are typically acquired at a high temporal resolution of <15 seconds per acquisition to observe the "wash-in" and "wash-out" patterns of the contrast agent to provide an assessment of vascular permeability throughout the prostate [26]. The initiation of angiogenesis from PCa increases the vascular permeability of these regions in comparison to healthy tissue [28]. DCE-MRI for PCa can also be assessed from a quantitative standpoint with the goal of standardizing across various sequence parameters. These methods employ the use of pharmacokinetic models to mathematically fit the concentration-time curve of the "wash-in" and "wash-out" properties of the contrast agent at different locations in the gland. Several models have been proposed to create quantitative maps that have been shown to more accurately localize malignancy [29–31]. However, there is evidence of potentially toxic side-effects of intravenous contrast agents in MRI (such as gadolinium-based agents) [32] and thus research efforts focused towards evaluating the performance of lesion localization in PCa using much lower doses of contrast agent may be of interest [33]. Despite this, there is currently a push to remove the DCE sequence entirely from mp-MRI for screening cases since it has been desig-
nated a very small role in assessing lesions according to PIRADS documentation [19]. Benefits of removing DCE from the mp-MRI protocol include lower cost to the healthcare system (no payment for gadolinium based contrast agents), faster scan time, and easier workflow (no concerns about kidney function pre-scan).

DWI is a static imaging sequence that measures the level of restricted diffusion of water through cellular membranes. This provides a snapshot of the permeability of the cellular membrane and space between adjacent cellular tissue. Cancerous tissue typically exhibits cell proliferation, and thus an increased cellular density, which correlates with a decrease in magnitude of diffusion of water and thus a hyperintense region on DWI [34]. Apparent diffusion coefficient maps are commonly generated by taking multiple DWIs with various gradient amplitudes (known as "b-values") in order to eliminate T2-weighting that is inherent to DWI. Cancer typically displays a low ADC value and thus appears hypointense on ADC imaging. For imaging of PCa, the magnitude of ADC is negatively correlated with increased tumour aggressiveness or Gleason score [35–37].

1.2.3 Prostate specific membrane antigen positron emission tomography

PET is a functional imaging technique used to visualize changes in various metabolic pathways within the body. The intravenous injection of radioactive agents known as radiotracers emit positrons that rapidly undergo an annihilation reaction to produce two photons traversing opposite to one another. A stationary detector ring around the patient measures coincident photon interactions to create a line of response which is used in filtered-back projection methods to create a cross sectional map of source activity [38]. Radiotracers are comprised of naturally occurring chemicals that bind to different metabolic pathways within the body. A common radiotracer is fluorine-18-labeled sugar fluorodeoxyglucose (FDG) which is metabolized by cells instead of naturally occurring glucose. Since cancerous tissue is highly metabolically active, an increased uptake of FDG is present in these regions, which can be seen on imaging [38]. Since its discovery, research has been focused on developing recommendations for its use in
oncology in various disease sites [39]. However, in the context of PCa, FDG-PET has demonstrated little value in the setting of primary staging and biochemical recurrence as it has been shown there is no correlation between PSA levels and stage, grade, or volume of PCa and uptake of FDG along with a sensitivity of 4.0% for identifying clinically verified organ confined cases [40, 41].

The prostate-specific membrane antigen (PSMA) is a type-II membrane glycoprotein that is showing promise as an imaging biomarker that is specific to normal and cancerous tissue within the prostate [42]. PSMA has been found to be significantly overexpressed in PCa, with levels of expression positively correlating with increasing tumour aggressiveness or Gleason grade (described further in section 1.5.1) [43].

Since the inception of PSMA-based radiotracers, research groups focused on the development of novel radiotracers began to try to maximize the sensitivity and specificity of lesion detection on PSMA-PET imaging, with the most widely studied agent being $^{68}$Ga-PSMA [44]. However, $^{68}$Ga based agents suffer from short-half lives and non-ideal energies which has motivated the development of Fluorine based analogs [45]. One such fluorine-based radiotracers is $^{18}$F-DCFBC which has been investigated as a low-molecular weight probe for imaging of metastatic PCa [46]. Built off the existing $^{18}$F-DCFBC probe is $^{18}$F-DCFPyL which targets the same epitope on the PSMA molecule as $^{68}$Ga-PSMA with the advantage of increased spatial resolution and efficient cyclotron production [47]. These radiotracers all suffer from a common issue in urinary excretion causing signal leakage from the buildup of radiotracers in the bladder. A newer fluorine based probe known as $^{18}$F-PSMA-1007 has demonstrated comparable tumour uptake with the benefit of non-urinary background clearance [45].

### 1.2.4 Ultrasound

Ultrasound imaging is a non-ionizing, acoustic method of generating images of in-vivo tissues. These images are produced from transducers made of piezoelectric materials that expand and contract under an applied voltage to generate longitudinal sound waves with clinical fre-
quencies typically around 6-MHz [48]. These sound waves will create reflected echoes that travel back to the transducer upon encountering an interface between two bodily tissues. The acoustic impedance of a tissue structure will determine the fraction of the sound wave that is reflected to the transducer [38]. High acoustic impedance structures will reflect the most ultrasound waves and appear hyperechoic on imaging while the opposite is true for low acoustic impedance structures.

In the context of PCa imaging, one example of a hyperechoic structure is a calcification while one example of a hypoechoic structure is a fluid filled cyst. The most common application of ultrasound for the prostate is during 2D-TRUS guided biopsy, which aids in guiding the biopsy probe through the gland and is described in detail in section 1.3.3. Ultrasound images are also typically acquired during high-dose-rate brachytherapy (HDR-BT) procedures to guide catheter insertion described more rigorously in section 1.4.2. The biopsy probe and catheters exhibit high acoustic impedance making them hyperechoic on ultrasound imaging.

1.2.5 Computed tomography

CT imaging involves ionizing radiation in the form of x-rays being projected through the patient around a complete arc to generate a cross-sectional image slice. The bed moves through the donut shaped CT-machine to generate multiple slices across the patient to form a 3D volume. Similar to a 2D x-ray projection, a CT image displays the attenuation of high-energy photons traveling through various tissues, where each voxel represents the electron density in that region referred to as a Hounsfield unit. Thus, tissues with high electron density appear as hyperintense on imaging and tissues with low electron density appear as hypointense. Contrast agents with high electron densities, such as iodine, may be injected into the patient during the scan to provide additional contrast in areas where perfusion occurs.

In the context of PCa imaging, CT may be used to detect gross extra-capsular disease, nodal metastatic disease or visceral metastatic disease (patients classified as having T3 or T4 clinical stage) [49]. Due to the limited soft-tissue contrast generated by this imaging modality, it is
generally not suitable to evaluate the prostate gland itself.

1.3 PCa screening and diagnosis

PCa screening in contemporary practice is performed using the PSA test and the DRE. The goal of screening is to catch early onset of disease progression and to administer more effective treatment with the hope of better survival outcomes. Upon positive finding from screening, the patient is typically sent for TRUS guided biopsy for diagnosis to get a better estimate of cancer stage and grade.

1.3.1 Prostate specific antigen test

PSA is an androgen-regulated protease that is produced by prostate epithelial cells and PCa [50]. PSA levels in the blood have been found to positively correlate with PCa stage, suggesting its use as a viable screening technique [51]. Typically, a blood sample is acquired from the patient and sent to a lab to determine the concentration of PSA in nanograms/mL of blood acquired. Patients exhibiting PSA levels <10 ng/mL are classified as low risk, 10-20 ng/mL for intermediate risk, and >20 ng/mL for high risk disease [49]. PSA testing may be responsible for a decline in PCa mortality since its induction [52, 53]. However, PSA tests have found to be very non-specific (specificities as low as 36% [16]) causing frequent false positives and unnecessary biopsies in patients with indolent or no disease [54, 55]. Pre-existing benign conditions such as benign prostatic hyperplasia and prostatitis have been found to influence PSA concentration in the body [56]. To help alleviate concerns of high false-positive rates (FPR) associated with the PSA test, this screening method is performed in conjunction with the DRE.
1.3.2 Digital rectal examination

A DRE is the physical examination of the lower rectal tract. The procedure typically involves a physician manually palpating the prostate gland through the rectal wall for signs of malignant growth. The DRE has been shown to identify cases of PCa presence even in the absence of elevated PSA, but still suffers from a poor positive predictive value [57]. Also, the physician is only able to detect cancers residing in the posterior portion of the gland adjacent to the rectal wall; thus, transition zone tumours are commonly missed during this procedure. Due to the overall poor predictive performance of the DRE independently, it is often performed in conjunction with the PSA test.

1.3.3 2-D TRUS-guided biopsy

Upon a positive finding from the PSA test and/or DRE, the physician will typically recommend the patient undergo 2-D TRUS guided biopsy for more accurate diagnosis. In this procedure, an endorectal transducer is inserted to acquire ultrasound images of the prostate gland to guide the biopsy needle through either the rectal wall (transrectally) or the perineum (transperineally). However, since there is great difficulty in identifying PCa on ultrasound [58, 59], a systematic sextant (6-cores) or extended-sextant (12-cores) targeting scheme is used, with evidence suggesting that the acquisition of more cores provides no significant benefit [60]. A majority of cores in the targeting scheme are extracted from the PZ due to the increased frequency of cancer presence in this area compared to other prostate zones (nearly 74%) [61]. Prostate biopsy cores are sent to pathology where an analysis of the histology is performed to determine cancer presence and Gleason grade. Physicians tend to have difficulties with accurately detecting cancer on TRUS-guided biopsy with detection rates ranging from 23-53% across multiple centers [62]. Sensitivity and specificity ranges from 50-92% and 46-91% respectively, presenting a problem with disease undertreatment through misdiagnosis [63].
1.4 Localized PCa treatment

1.4.1 Active surveillance

Active surveillance, sometimes referred to as a "wait and watch" strategy, is the careful monitoring of a patient with indolent cancer with the intent to intervene with curative treatment if the cancer progresses. It was designed to combat over-diagnosis and over-treatment during the rise of non-specific PSA screening and has shown to be a safe alternative to immediate intervention [64].

The NCCN guidelines define the criteria for patients recommended to undergo active surveillance. Very low risk (T1c, Gleason score ≤ 6, PSA < 10 ng/mL, fewer than 3 positive biopsy cores) patients with life expectancy ≥ 20 years and low risk (T1-T2a, Gleason score ≤ 6, PSA < 10 ng/mL) patients with life expectancy ≥ 10 years are suitable candidates for active surveillance [49].

While some patients are moved from active surveillance to treatment due to early PSA doubling time (13-48% of patients), there is still evidence of patients that elect treatment with no signs of progression (7-13% of patients) [65, 66]. The anxiety of untreated cancer or fear of rapid progression into distant metastasis is often a common negative side-effect of active surveillance [67].

1.4.2 Radical radiation therapy

Radiation therapy in the context of PCa involves the irradiation of the prostate gland with high-energy photons with the intent of killing cancerous tissue. Typically, the entire gland is treated equally with ideally uniform dose - known as radical radiation therapy. Radiation therapy methods have shown similar patient outcomes to surgery [68] despite having improved overall toxicity profiles [69]. However, radical prostatectomy provides pathologists with histology of the entire gland allowing for a more complete picture of the extent of the disease which may inform physicians about patient prognosis. With radical radiation therapy, much of this
information is lost as the specimen remains in-vivo, causing physicians to rely on surrogates such as PSA doubling time or imaging to make judgments about prognosis.

PCA radiotherapy techniques for primary treatment are commonly external beam radiotherapy (EBRT) and brachytherapy. EBRT is the delivery of radiation from a source external to the patient in various configurations while brachytherapy involves implanting a radioactive source directly at the target (usually interstitially). HDR-BT involves physicians placing a high activity radioactive source at the site of the cancer for short periods of time before withdrawn through a catheter, whereas low-dose rate brachytherapy involves the placement of permanent, low-activity radioactive seeds in the prostate. Ideal candidates are patients with low risk cancers who may undergo EBRT with prescription doses of 75-79 Gy using conventional fractionation or low-dose rate brachytherapy [49]. Intermediate to high-risk patients may be treated with EBRT solely, with prescription doses up to 81 Gy using conventional fractionation; or a combination of high-dose rate or low-dose rate brachytherapy and EBRT, which has shown improved patient outcomes compared to EBRT alone [70].

Despite its many therapeutic benefits, radical radiation therapy has been associated with sexual dysfunction and urinary incontinence with rates better than for radical prostatectomy [69, 71]. More recently, studies have also suggested that PCA patients treated with EBRT have an elevated risk of developing primary bladder cancer [72], which is a consequence of collateral radiation induced toxicities to nearby organs at risk. Research is also focused around using imaging to boost doses to threatening PCA lesions as post-radiation therapy patients with a high probability of local failure are at a greater risk of distant metastasis [73].

1.4.3 Radical prostatectomy

Radical prostatectomy is a surgical procedure that involves the total removal of the prostate gland and its associated structures. This procedure may be performed in open-surgery or minimally invasive laparoscopic robotic surgery, which has shown to have improved short-term outcomes [74]. Radical prostatectomy from either procedure has been shown to reduce the risk
of PCa mortality and risk of distant metastasis 10-years after surgery compared to watchful waiting techniques [75].

Ideal candidates for radical prostatectomy are those with clinically localized PCa with life expectancies ≥10 years and no contraindications to surgery. Radical prostatectomy may also be used as a salvage treatment for patients with recurrence post radiation therapy but has shown to have significantly higher morbidities than when radical prostatectomy is used as initial treatment [76]. Patients with elevated risk of nodal metastasis will often undergo pelvic lymph node resection in conjunction with radical prostatectomy [49].

Radical prostatectomy has been associated with high rates of erectile dysfunction and urinary incontinence [77]. There are also complications associated with surgical operation including blood loss and/or transfusion effects and risks of anesthesia [49].

### 1.4.4 Lesion directed therapy

PCa that is unifocal and confined to the prostate gland while presenting favourable conditions for progression [78] and healthy tissue sparing may benefit from localized lesion directed therapy instead of a radical therapeutic approach [79]. In these procedures, treatment is delivered directly to the lesion with the intention of maximal healthy tissue sparing to improve patient quality of life. Typically, the lesion towards which treatment is directed is the dominant intraprostatic lesion (DIL), also known as the index lesion; most literature defines this as the largest lesion within the prostate [79, 80]. The pathological characteristics of this lesion will indicate the prognosis of the patient (i.e. Gleason grade or presence of ECE) [80]. Studies have demonstrated that these lesions are the drivers for local recurrence and treatment directed at these lesions may result in improved patient outcomes [81, 82].

There are many different methods for accomplishing lesion directed therapy in the context of PCa. Cryotherapy is a procedure that involves transporting very low temperature gases through a needle to the lesion to freeze and destroy the tissue [83]. High-intensity focused ultrasound uses acoustic frequencies much higher than diagnostic limits that are focused to-
wards the lesion to heat and ultimately ablate cancerous tissue [84]. Laser ablation therapy involves transporting high powered laser light through a fiber-optic cable via catheter to the lesion to burn the cancerous tissue; often this procedure is performed with MR image guidance [85, 86]. Photodynamic therapy involves the localized injection of photosensitizers which cause destructive reactions to nearby tissue when exposed to visible light [87]. Radiation treatment in the form of intensity modulated radiation therapy provides optimized dose coverage to target regions by controlling the radiation beam to conformal targets in the prostate [88]. Another form of targeted radiation treatment is in the form of stereotactic body radiotherapy which uses higher doses per fraction during standard treatment to take advantage of the resulting relative radiobiological effects [88]. Furthermore, brachytherapy has been used for lesion directed treatments by boosting doses in the form of adding additional catheters [89] or boosting dwell times [90] to suspicious regions.

One of the main issues with focal therapy treatment methods is the selection criteria for suitable candidates. Due to the highly multi-focal nature of the disease, there is often residual cancer in the gland outside of the main DIL. While ample evidence suggests that the DIL drives disease progression [91] - it is assumed that residual cancer does not compromise long-term disease control [92]. Thus, there is uncertainty in the community surrounding long-term focal therapy outcomes and more studies are necessary to determine its efficacy [93]. Lesion directed therapy is also reliant on image-guided techniques to create targets in-vivo; low sensitivity or specificity imaging methods may cause under or over-treatment of the lesion.

1.5 PCa histology

1.5.1 Histopathology of PCa

PCa is a very heterogeneous disease that presents with diverse morphological patterns on histology that provide insight towards its level of aggression. These tissue archetypes were categorized first by Dr. Gleason in the 1960s to classify different microscopic tissue com-
ponents based on growth patterns within the stromal tissue and differentiation of glandular structures [94]. Gleason’s grading system grouped the tissue structures into five grade groups, adding the primary and secondary most common patterns together to form a Gleason score ranging from 2-10. The International Society of Urological Pathology (ISUP) revised the Gleason grading system in 2005 and more recently in 2014 to form the system that is currently used [95, 96]. The ISUP grade groups range from grade groups 1-5 and make a distinction on between tumours that have the same Gleason score but with different primary Gleason patterns (i.e. Gleason score: 7 with Gleason grade: 4+3 vs. 3+4) and was shown to have more accurate grading stratification with this simplified system [97]. The ISUP grade groups stratified the original Gleason scoring and grading in the following way: ISUP grade group 1 is Gleason score ≤6, ISUP grade group 2 is Gleason grade 3 + 4, ISUP grade group 3 is Gleason grade 4 + 3, ISUP grade group 4 is Gleason score 8, and ISUP grade group 5 is Gleason scores 9-10 [97].

1.5.2 Clinical significance of pathology

Histologic analysis of prostate tissue is crucial in all steps of the clinical workflow from biopsy analysis to post-prostatectomy prognosis. Prostate biopsy is examined for a first definitive diagnosis after positive screening, during follow-ups for active surveillance candidates, and to confirm local recurrence in a post-treatment setting. The inclusion of the Gleason score for PCa in biopsy analysis has been rigorously validated as an important prognostic factor [98–100]. Gleason score from biopsy specimens has also been found to correlate with other molecular markers such as PSA concentration [101]. More definitively, Gleason scores ≥7 are associated with worse prognosis and Gleason scores <7 have much slower progression rates. Physicians commonly use the Gleason score predictor along with a series of other clinical variables in nomograms to predict disease progression after surgery or radiation therapy [102, 103].
1.5.3 Registration of histology to imaging

In clinical and research practices, there is often disagreement about the location of cancer presence on histology and in-vivo imaging. Registration of biopsy tissue provides only a small snapshot of the tissue landscape and may under-estimate cancer presence (See section 1.3.3). Whole-gland transverse sections analyzed post-prostatectomy provide much more information related to disease progression, but registration is often much more challenging.

Our group has developed a method for registration of annotated post-prostatectomy histology specimens to pre-operative in-vivo imaging with a quantified target registration error [104]. This procedure is performed in two major steps: registration of histology to ex-vivo imaging and registration of ex-vivo imaging to in-vivo imaging.

Prostate specimens acquired post-prostatectomy were mounted with seven external fiducial markers soaked in gadolinium on the surface of the gland. Three internal fiducials soaked in a mixture of blue dye and gadolinium were inserted into the gland in the superior-inferior direction with a pre-specified arrangement. External fiducials were made of porcine kidney to allow for the pathologist to easily cut the specimen. Internal fiducials were made of cotton thread and removed after imaging. The prostate specimen was then sliced in 4mm transverse sections while the apex and base were cut sagittally. Slices from each section were cut using a microtome and digitally scanned before being annotated by a pathology assistant. All contours were subsequently verified by a trained genitourinary pathologist. Registration between histology slices and corresponding ex-vivo imaging was performed using an affine transformation that attempted to find the best matching plane on ex-vivo imaging using the information from the external and internal fiducial strands [105].

Ex-vivo to in-vivo registration was performed semi-automatically using user defined homologous anatomical landmarks between both the image volumes. First, the ex-vivo volume was oriented to the space of the in-vivo volume using fiducial markers on the extremities of the gland. Next, anatomical landmarks, such as calcifications, were identified on both the ex-vivo and in-vivo specimens as pairs of fiducial markers. A thin-plate spline registration algorithm
deformably mapped the ex-vivo specimen to the in-vivo specimen based on the pairs of fiducial markers. Thus, the final result from the algorithm produces a registration scheme from histology to in-vivo prostate MRI with a measured target registration error of 1.7 mm on T2W imaging and 1.9 mm on ADC maps [104].

There are many challenges associated with accurate registration of prostatectomy histology to in-vivo imaging, some of which have been summarized by Meyer et al. [106]. First and foremost, by using an ex-vivo specimen to link histology to the in-vivo volume, the ex-vivo prostate specimen is subjected to many physical deformations including fixation from the formalin solution (required to keep tissue from degrading) as well as compression from paraffin wax embedding (required to hold the tissue for pathology processing). Also, if endorectal receive coils are used during the MRI, the prostate is notably deformed in the anterior direction, causing significant differences in the tissue geometry. Furthermore, histology sections cut by pathology are unlikely to be in the same orientation that the specimen was imaged which increases the degrees of freedom (translation and rotation transformations) for mapping the slide to a corresponding oblique plane in MRI. Our group proposed the use of external and internal fiducials strategically placed to assist in optimizing the best plane of fit, however there are inherent challenges in finding the correct material to use for the fiducials as they need to adhere to the gland through all the chemical processes involved in pathology while still being visible on MRI [105].

1.6 Artificial intelligence in PCa imaging

There are many applications of artificial intelligence in PCa imaging; namely: detection, segmentation and characterization (or classification), as examples. Disease detection is important for disease diagnosis at early time points as well as suggesting target areas for biopsy. Lesion segmentation is important for more precise biopsy measurements and radiotherapy targeting. Classification or characterization of lesions identified by machine learning or an expert
observer can help to provide a second opinion. These are just some of the many ways machine learning technologies are being investigated for improving PCa diagnosis, treatment and follow-up [107].

### 1.6.1 Radiomics and machine learning

The practice of radiomics involves the conversion of imaging information into quantitative features for use in decision support systems [108]. The process for radiomics analysis usually begins with acquiring an imaging data set, segmentation of regions of interest (ROIs), feature extraction and quantification, and analysis of quantitative data using decision models used to make predictions. Radiomics has been a prominent focus of research in the field of oncology for many disease sites such as breast, lung, and prostate [109–111].

However, there are inherent challenges involved with the incorporation of radiomics pipelines in computer aided diagnosis (CAD) systems; most commonly surrounding reproducibility for clinical translation. Imaging data sets are not often acquired with the same reconstruction parameters across different vendors and institutions, which may have significant impacts on the extracted quantitative information [112]. Localizing ROIs is often done with expert assistance to more accurately define the target; however, there are biases present in this process. Histology provides the gold-standard in terms of definitive disease localization, but accurate registration to imaging has many challenges of its own (see section 1.5.3). Prediction outcomes have demonstrated variability as a function of various factors such as lesion and/or patient size [113]. These topics highlight the challenges with radiomics frameworks incorporated into decision making models.

Prediction models are often founded on the concept of machine learning, which is the concept of computers making decisions and predictions based on prior data without a set of explicit instructions [114]. Essentially, the computer ”learns” from the data to make predictions - analogous to how humans learn from experiences to influence our future decisions. There are two types of machine learning: supervised and un-supervised learning. Supervised learning
involves generating a function that maps an independent variable to a set of outputs based on the trained set of predictors with the independent variable being exposed to the learning algorithm during training [115]. In un-supervised learning, the independent variable is not provided to the algorithm during training- the algorithm must infer clusters of the data on its own. In the context of classification, there have been a variety of mathematical methods developed to separate data. These models can be divided up into classical machine learning techniques (e.g. K-nearest neighbour (KNN)) and neural networks (e.g. artificial neural network).

### 1.6.2 Classical machine learning techniques

KNN is an algorithm that relies on mapping a label to an output by inferring from the labels of the closest neighbours in Euclidean space of the training set [116]. The number of neighbours that are inferred upon is decided a-priori as a parameter ”K”. This parameter is often made odd to ensure that there is no even split when deciding between labels.

Naive-Bayes (NAIVEB) classifiers operate fundamentally under the Bayes’ formula for conditional probability. We are mapping a series of inputs to a set of probabilities that represent the likelihood of that input belonging to a specific class. If there are only two classes, the problem simplifies to binary classification. The ”naive” component of this classification method refers to the assumption made a-priori that all features for an input variable are independent of each other [117].

Random-forest (RF) classifiers are a subset of decision tree learning methods. Decision trees operate by mapping an input to a target label through a set of decisions made (nodes) that travel along the tree until the output is reached. These decisions are usually binary and, in the context of radiomics, often represent thresholds for certain feature values. RFs are an ensemble of decision trees that map a target to an input variable by using a voting scheme from the output from all decision trees.

Logistic linear (LOGL) classifiers are a subset of linear regression based classification. In LOGL classification models, a linear function has its slope and y-intercept fit to the data as
seen in equation 1.1.

\[ y = b_0 + b_1 \cdot x \]  

(1.1)

The output from this function is then passed as input into a sigmoidal function as seen in equation 1.2 (also known as a logit function) to map the output to a real value between 0 and 1.

\[ p = \frac{1}{(1 + e^{-y})} \]  

(1.2)

This is the form logistic regression takes when predicting binary variables with discrete outcomes instead of continuous (where regression is more suitable).

### 1.6.3 Neural networks and deep learning

Previously, we discussed methods of linear transformation functions to map a set of inputs to a desired set of outputs. Neural networks are a subset of non-linear models that take their inspiration from the human neuron. Figure 1.1 shows a simple model of a neuron with arbitrary numbers of inputs, \( x \), being mapped to an output, \( y \).

The formula for producing the output is:

\[ y = f(\sum_{i=1}^{n} x_i \cdot w_i + b) \]  

(1.3)

where \( x \) is the input, \( w \) is the weight specific to this input, \( b \) is the bias specific to the neuron, \( f() \) is the activation function and \( y \) is the output. Each input is multiplied by an associated weight along the edge of the neural network graph and added to the bias associated with the node of the graph. The weights and biases in these models are tunable parameters that are learned through training on a data set via backpropagation [118]. These values are summed and passed as input into a non-linear activation function, signifying the non-linearity of the model. An example activation function that is used in practice is the sigmoidal function (see equation 1.2).
The distinction of a deep learning model is the incorporation of one or more intermediate layers in the standard neural network model before converging to the final output layer. This is a model parameter that is tunable by the user, and provides the option for a significant increase in model complexity by increasing the total number of tunable weights and biases in the model. However, this may cause over fitting of the model, as the added complexity of the system allows for the model to learn too closely to the input data. Thus techniques, such as regularization, have been developed to help prevent over fitting of these highly complex models during training [119].

Another method used for classification is the convolutional neural network (CNN). The difference between a CNN and an artificial neural network is that the neuron is modeled as a convolutional kernel, with the inputs being the raw values of the pixels in the image captured by the kernel, the weights being associated to each element within the convolutional kernel, and the bias being associated with the overall kernel itself. Figure 1.2 depicts a typical CNN architecture, with the convolutional kernel being displayed in yellow at each layer, with the
Figure 1.2: Architecture of a simple CNN with two output classes.

final layer regressing to two outputs.

1.7 Clinical challenges and related works

Mp-MRI has been identified as a promising, non-invasive method for localizing prostate tumours and determining characteristics related to their size, shape and location within the gland (see section 1.2 for more information on mp-MRI) [25]. However, mp-MRI still has some inherent limitations in terms of lesion classification. First and foremost, there is significant inter-observer variability in lesion contouring between physicians [120, 121]. Figure 1.3 displays an example from data at our institution, where we provided three different radiologists with three different pulse sequences of the same slice in the same patient and asked them to contour the DIL. Across each column, there is large variability between the contours made for each physician, which may have significant impacts on the patient’s treatment management. Furthermore, mp-MRI for PCa detection has varied widely with sensitivities reported in a meta analysis by Futterer et al. to be between 44-87% [122]. De Rooij and colleagues found a pooled specificity and sensitivity for PCa detection on mp-MRI to be 88% and 74%, respectively [123].

More recently, Stabile et al. summarized the major improvements needed for mp-MRI, namely with respect to the development of computer-aided diagnosis systems and assessment
of imaging adjuncts to use in combination with MRI [124]. The work in this thesis is focused around augmenting the interpretation of mp-MRI for PCa classification surrounding these two areas.

### 1.7.1 Artificial intelligence for PCa classification

Many studies have worked on classifying prostate tissue present on mp-MRI with a more recent overview provided in the review article by Bardis et al [107]. Based on the materials of the article, we can divide the studies into two categories: (1) lesion detection and/or segmentation and (2) lesion characterization/classification. In the context of lesion classification, studies usually operate with selected ROIs on mp-MRI that have been labeled as cancerous or benign.
that is driven by some underlying validation scheme (i.e. biopsy report). Table 1.2 highlights some of the previous work surrounding PCa classification against benign tissue. Most of the studies presented in this list validated their models using biopsy results to try to pinpoint the location of the disease [125–129]. This has many limitations, with the most predominant being the high possibility of missed cancers with detection rates of 23-53\% (for more information on limitations of TRUS biopsy see section 1.3.3) [62]. Several authors developed their models based on a consensus mapping approach between radiology and pathology [111, 130–132]. In this method, the physicians will come to a consensus on where tumours and suspicious regions should be mapped by looking at the histology and relevant mp-MRI. These tumours are typically used as landmarks during the registration of histology to the mp-MRI space. However, this induces a bias as these studies are skewed towards the accurate localization of tumours, but not for healthy or confounding tissue. Litjens et al. [133] and Kwak et al. [134] used MR-guided biopsy to validate their models allowing for targets to be identified in the MR space and then subsequently validated for cancer presence by the targeted biopsy. The issue that arises from this method is that these studies are only training their models on tumours that can be identified by the radiologists. Tumours that are unidentifiable would not be targeted by the biopsy needle and thus may be missed entirely. Our system proposed in this thesis was validated on transverse prostatectomy whole slide images that were registered to in-vivo MRI with a measured target registration error [104]. This allows us to evaluate the performance of our models with all of the aforementioned biases eliminated in our experiments. Our work aims to fill the gap on validating radiomics driven machine learning models based on an unbiased reference standard compared to previous work in this space.
<table>
<thead>
<tr>
<th>Reference</th>
<th>Year</th>
<th>Patients</th>
<th>Algorithm</th>
<th>Imaging modalities</th>
<th>AUC</th>
<th>Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alfano et al. [Chapter 2]</td>
<td>2021</td>
<td>61</td>
<td>LOGLRF</td>
<td>T2W, ADC</td>
<td>0.80</td>
<td>Co-registered prostatectomy WSI</td>
</tr>
<tr>
<td>Chen et al. [126]</td>
<td>2019</td>
<td>204</td>
<td>Transfer learning</td>
<td>T2W, DWI, DCE</td>
<td>0.83</td>
<td>MR-guided biopsy</td>
</tr>
<tr>
<td>Song et al. [125]</td>
<td>2018</td>
<td>195</td>
<td>CNN</td>
<td>T2W, DWI</td>
<td>0.94</td>
<td>MR-guided biopsy</td>
</tr>
<tr>
<td>Wang et al. [127]</td>
<td>2018</td>
<td>360</td>
<td>CNN</td>
<td>T2W, ADC</td>
<td>0.95</td>
<td>MR-guided biopsy</td>
</tr>
<tr>
<td>Wang et al. [111]</td>
<td>2017</td>
<td>195</td>
<td>CNN</td>
<td>T2W, DWI</td>
<td>0.84</td>
<td>Consensus mapping</td>
</tr>
<tr>
<td>Liu et al. [135]</td>
<td>2017</td>
<td>341</td>
<td>CNN</td>
<td>T2W, ADC, DCE, Ktrans</td>
<td>0.80</td>
<td>MR-guided biopsy</td>
</tr>
<tr>
<td>Mehrtash et al. [128]</td>
<td>2017</td>
<td>344</td>
<td>CNN</td>
<td>T2W, ADC, DCE</td>
<td>0.84</td>
<td>MR-guided biopsy</td>
</tr>
<tr>
<td>Seah et al. [129]</td>
<td>2017</td>
<td>346</td>
<td>CNN</td>
<td>T2W, ADC, DCE</td>
<td>0.83</td>
<td>Consensus mapping</td>
</tr>
<tr>
<td>Sidhu et al. [130]</td>
<td>2017</td>
<td>67</td>
<td>Bivariate model</td>
<td>T2W, ADC, DCE</td>
<td>0.88</td>
<td>MR-guided biopsy</td>
</tr>
<tr>
<td>Litjens et al. [133]</td>
<td>2015</td>
<td>107</td>
<td>RF</td>
<td>T2W, DWI, DCE</td>
<td>0.89</td>
<td>MR-US fusion biopsy</td>
</tr>
<tr>
<td>Kwak et al. [134]</td>
<td>2015</td>
<td>244</td>
<td>SVM</td>
<td>T2W, DWI</td>
<td>0.95</td>
<td>Consensus mapping</td>
</tr>
<tr>
<td>Peng et al. [132]</td>
<td>2013</td>
<td>48</td>
<td>LDA</td>
<td>T2W, DWI</td>
<td>0.95</td>
<td>Consensus mapping</td>
</tr>
<tr>
<td>Niaf et al. [131]</td>
<td>2012</td>
<td>30</td>
<td>SVM</td>
<td>T2W, DWI, DCE</td>
<td>0.89</td>
<td>Consensus mapping</td>
</tr>
</tbody>
</table>

Table 1.2: Summary of previous work done on PCa classification on mp-MRI.
1.7.2 PSMA-PET for PCa classification

Previously, it was mentioned that the investigation of imaging adjuncts to use in combination with MRI may be important for improving mp-MRI interpretability [124]. Many studies have investigated the role that PSMA-PET may have as an adjunct to mp-MRI and shown concordance between regions of elevated max standardized uptake value ($SU_{\text{max}}$) and regions of biopsy proven disease [136–138]. Critically missing in the literature is a set of recommendations for segmenting the DIL within the prostate validated on a high-fidelity reference standard. A summary of previous work in this space can be found in Table 1.3. Lopci et al. [136] opted to use SUV cutoffs for delineating the DIL and validated their results off of prostate biopsy. Kostyszyn et al. opted to use deep learning based convolutional neural networks for classification of the DIL on PSMA-PET validated off of biopsy as well. The biases associated with biopsy based reference standards are outlined previously in section 1.7.1. Zamboglou et al. [139] and Spohn et al. [140] both use co-registered prostatectomy histology to PET/CT imaging to validate their analysis. However, this may create additional registration error between the two images as intra-prostatic tissue is not well defined on CT compared to MRI. Our study is the first to use a combination of $SU_{\text{max}}$ and margin expansion to find optimal criteria for focal therapy and guided biopsy based procedures validated against co-registered prostatectomy histopathology to the PET/MRI image volumes with a measured target registration error [104].
<table>
<thead>
<tr>
<th>Reference</th>
<th>Year</th>
<th>Patients</th>
<th>Tracer</th>
<th>Method</th>
<th>Recommendation</th>
<th>Results</th>
<th>Validation</th>
</tr>
</thead>
</table>
| Alfano et al. [Chapter 3] | 2020 | 12       | $^{18}$F-DCFPyL   | Percent $SUV_{\text{max}}$ and margin expansion | 1) $SUV_{\text{max}} = 67\%$ and 8 mm margin  
2) $SUV_{\text{max}} = 81\%$ and 5 mm margin | 1) 95% (sens.)  
2) 65% (sens.)  
95% (spec.) | Co-registered prostatectomy WSI |
| Spohn et al. [140] | 2020 | 10       | $^{18}$F-PSMA-1007 | Percent $SUV_{\text{max}}$ | $SUV_{\text{max}} = 20\%$ | 93% (sens.)  
96% (spec.) | Co-registered prostatectomy WSI |
| Kostyszyn et al. [141] | 2020 | 29       | $^{18}$F-PSMA-1007 | CNN | N/A | 100% (sens.)  
57% (spec.) | Biopsy |
| Lopci et al. [136] | 2018 | 45       | $^{68}$Ga-PSMA-HBED-CC | SUV cutoff | $SUV_{\text{max}} = 4.8$ | 82.4% (sens.)  
72.2% (spec.) | Biopsy |
| Zamboglou et al. [139] | 2018 | 20       | $^{68}$Ga-PSMA-11 | Percent $SUV_{\text{max}}$ | $SUV_{\text{max}} = 20\%$ | 100% (sens.)  
67% (spec.) | Co-registered prostatectomy WSI |

Table 1.3: Summary of previous work done on PCa classification on PSMA-PET.
1.7.3 Focal dose escalation towards PSMA-PET segmentations

Although many studies have examined boosting brachytherapy treatment plans using MRI derived contours of the DIL, to the best of our knowledge there exists no other study beyond ours that performs a similar experiment using PSMA-PET derived contours. Zamboglou et al. [142] performed a similar experiment by boosting dose to the gross tumour volume defined as 30% of $SUV_{\text{max}}$ for intensity modulated radiotherapy procedures which is a subset of EBRT (see section 1.4.2).

1.8 Thesis hypothesis and objectives

Mp-MRI may be augmented through artificial intelligence techniques as well as with additional imaging sequences. Machine learning models can help physicians make critical decisions about PCa presence by providing a second input on their suspected targets. PSMA-PET can also be used to identify DILs in the prostate, and these targets may be used to boost radiation therapy treatment plans with the hope of increasing dose to underlying disease. Based on these observations, we set out to test the following hypotheses related to this thesis.

1.8.1 Hypothesis

1. A radiomics and machine learning algorithm can classify cancer vs. non-cancer on mp-MRI validated on co-registered histology with an $AUC \geq 0.80$.

2. DILs on PSMA-PET imaging can be segmented using a threshold of $SUV_{\text{max}}$ and radial margin expansion with maximal sensitivity and specificity for focal therapy/boosting and guided biopsy applications.

3. Inverse-planning optimization of dwell times in HDR-BT targeted at segmentations of DILs on PSMA-PET result in a significantly increased $D_{98}$ to all underlying co-registered high-grade histology.
1.8.2 Research objectives

To answer these hypotheses, this thesis aims to explore the following research objectives using accurately co-registered prostatectomy histopathology to in-vivo multi-modality imaging.

1. To develop and test a machine learning model on radiomic features derived from an un-biased sampling of clinically significant malignant and benign lesions on mp-MRI validated with co-registered histology.

2. To investigate optimal combinations of threshold of $\text{SUV}_{\text{max}}$ and margin expansion for DIL segmentation on PSMA-PET imaging for focal therapy/boosting (Sensitivity $\geq 95\%$ with maximal specificity) and guided biopsy (Specificity $\geq 95\%$ with maximal sensitivity).

3. To investigate the change in dosimetry in HDR-BT to underlying PCa in focally boosted treatment plans using generated PSMA-PET segmentations compared to whole-gland treatment plans.
References


with an increased risk of death from prostate cancer after radical prostatectomy for patients with seminal vesicle invasion and negative lymph nodes,” in Urologic Oncology: Seminars and Original Investigations, vol. 32, pp. 26–e1, Elsevier, 2014.


[70] P. J. Hoskin, A. M. Rojas, P. J. Bownes, G. J. Lowe, P. J. Ostler, and L. Bryant, “Randomised trial of external beam radiotherapy alone or combined with high-dose-


43


Chapter 2

Prostate cancer classification using radiomics and machine learning on mp-MRI

2.1 Introduction

PCa is the most-common male non-skin cancer in North America [1]. Patients are screened using two routine methods: 1) a DRE and 2) a PSA test that is measured by taking a blood sample. Following suspicious results from either test, a physician will recommend the patient be sent for diagnosis via TRUS guided biopsy. In this procedure, 6-12 cores will be extracted from the prostate using a systematic targeting template which will be sent to pathology to perform an analysis on the histology. However, TRUS-guided biopsy is limited by its rate of misdiagnosis, as it has been shown to have a false negative rate of 23% [2]. This limitation is a major barrier towards personalized treatment; accurate staging and grading is crucial to determining the best treatment plan for the patient.

Mp-MRI has emerged as an additional diagnostic tool in the clinic for detecting PCa in patients [3]. More specifically, mp-MRI may have an important role in the setting of primary
staging as it has been shown that when fused with ultrasound during targeted biopsy, 67% more Gleason ≥4 + 3 tumours were identified, versus standard 12-core template biopsy [4]. Furthermore, it has been shown that mp-MRI/ultrasound fusion can identify more clinically significant cancer (38%) versus 12-core biopsy (26%) while also diagnosing fewer clinically insignificant cancers, potentially sparing unnecessary treatment [5]. To standardize radiological assessment of PCa on mp-MRI, PI-RADS was developed to provide technical guidelines on image acquisition and clinical guidelines for lesion assessment [6]. Despite these efforts to improve readers’ assessment of PCa status on mp-MRI, mp-MRI still suffers from two core limitations: (1) the high level of expertise currently required to interpret the complex, multidimensional mp-MRI signals makes accurate staging and grading challenging; and (2) there is a high level of inter-observer variability associated with lesion classification on mp-MRI [7, 8]. These issues present many clinical challenges that make accurate diagnosis and treatment management more difficult.

To overcome these limitations, CAD systems have been developed to leverage the power of artificial intelligence to assist readers in prostate mp-MRI interpretation. CAD models most commonly operate by generating radiomic features of imaging statistics, texture, shape and/or size and using machine learning to train models to correctly classify a given intraprostatic lesion (ROI) defined by a human observer on mp-MRI as harboring PCa versus benign tissue (non-PCa). CAD systems may provide a second opinion on suspicious regions identified by radiologists and have shown to improve reader sensitivity of detecting clinically significant cancer versus unassisted interpretation while significantly reducing reading time [9]. Fully automated CAD systems that both identify and classify lesions on mp-MRI have also been described [10, 11].

However, many PCa CAD models developed in the literature may be limited by inherent biases present in the study design. The most prominent is the selection bias induced by the choice in reference standard for validating the model. The most common strategies for validating PCa and non-PCa targets on mp-MRI are consensus mapping, MRI-guided biopsy, and
radiological assessment. Each of these strategies has its limitations; consensus mapping and radiological assessment approaches are biased towards lesions that are true positives on MRI, and thus are identifiable by human observers. MRI-guided biopsy approaches have shown high detection rates for high-grade disease (sensitivity: 91%) but suffer in detection of low-grade disease (sensitivity: 44%) creating bias towards user identified true-positive high-grade cancer [12]. Critically needed is a PCa CAD model validated on an accurate and unbiased fusion of histology to in-vivo imaging.

Another common limitation is size bias when defining regions as input for radiomics feature extraction. For instance, consider a binary classifier that is to distinguish between ROIs of class A and ROIs of class B based only on texture features. Suppose that the class A ROIs tend to be larger than the class B ROIs, or vice-versa. The presence of a larger number of pixels or voxels in ROIs of one of the classes may influence the values of the computed texture features in a consistent way, allowing the classifier to indirectly detect the size differences between the ROIs of the different classes. This can lead to a spurious result where the classifier seems to be able to distinguish the two classes based on texture, but in reality, the classification resulted from an inadvertent bias in the texture feature values resulting from a consistent difference in ROI size between classes. This is, of course, only a problem in situations where ROI size is not germane to the classification problem at hand, as is the case for our problem of distinguishing cancerous ROIs within the prostate from equivalent non-cancerous tissue samples. This effect has been demonstrated in previous work [13]. Welch et al. attempted external validation on a radiomics model and found that the underlying tumour signature based on radiomic texture features was a merely a surrogate for tumour volume [14]. Yet to be investigated in the literature is the effect on reported system performance while controlling for size biases stemming from differences in lesion volume for PCa vs. non-PCa based classification. Furthermore, shape bias regarding the differences in geometry for PCa vs. non-PCa lesions have not been investigated in the literature and thus it is unknown whether these differences may impact system performance if not controlled. Critically needed is a report on the effect of size and shape biases on measured
CAD system performance to provide considerations for future CAD experimental design.

In this study, we developed a radiomics-based machine learning model for classifying ROIs on mp-MRI containing only PCa, vs ROIs on mp-MRI containing no PCa, as defined by accurately co-registered whole-mount prostatectomy histopathology. We attempted to control for selection, size, and shape biases in our development of the model. To combat selection bias, our lab used accurately co-registered transverse, mid-gland histology from post-prostatectomy patients to pre-surgery in-vivo mp-MRI with a measured target registration error [15]. This registration was performed in two parts: (1) mapping the annotated histology sections to ex-vivo MRI and (2) mapping the ex-vivo to in-vivo MRI. Thus, we are able to map transverse, mid-gland histology to an oblique plane in the MR space causing a reduction in selection bias for PCa lesions in this study. We controlled for size biases by developing a sampling algorithm which performed an optimization that maximized the number of patients and samples in our dataset while maintaining equal mean areas between PCa and non-PCa lesions to minimize volume confounding effects found with other sampling techniques [16]. Furthermore, our sampling algorithm also controlled for shape biases by sampling non-PCa tissue using masks identically shaped to PCa tissue. We also aimed to investigate trends in overall system performance based on PCa lesion size and shape by artificially inducing these biases into our experiments in order to provide considerations for other groups performing similar research, as it has been found that many of these biases have not been considered in previous work [17–23]. These studies may also contain a location bias, where benign tissue is preferentially sampled by mirroring the cancerous lesion to the contralateral side of the gland. In addition, tissue sampled in other areas of the gland labeled as benign that haven’t been verified via pathology may contain significant underlying disease. The rationale for these experiments is to shed light on potential sources of bias that may cause a classifier to produce a spurious result that may not be reproducible in external validation. Thus, this study aims to answer the following questions: (1) What is the highest measured CAD system performance for classifying PCa vs. non-PCa on mp-MRI when validated using accurate whole-mount radical prostatectomy histology fu-
sion to mp-MRI? (2) How do differences in shape and size between PCa and non-PCa lesions affect CAD system performance?

2.2 Methods and materials

2.2.1 Patient characteristics

We analyzed 61 male patients with biopsy-proven PCa as part of a larger multimodality image guided PCa therapy study (registration number NCT04009174). This research was approved by our Institutional Review Board (the Health Sciences Research Ethics Board), and we obtained informed consent from all patients. We obtained 261 prostate mid-gland tissue sections from these 61 patients post radical prostatectomy. The inclusion criteria were: (1) age ≥ 18, (2) biopsy-confirmed PCa. The exclusion criteria were: (1) prior PCa therapy, (2) use of 5 alpha reductase inhibitors within 6 months of study start, (3) inability to comply with pre-operative imaging, (4) contrast agent allergy, (5) sickle cell/other anemias, (6) hip prosthesis, (7) sources of artifact within the pelvis and (8) contraindications to MRI.

2.2.2 MR imaging

We obtained prostate mp-MRI 19 ± 8 weeks (mean ± SD) after biopsy and 2 ± 1 weeks before surgery using a Discovery MR750 (GE Healthcare, Waukesha, WI, USA) (n=41) or a Siemens Biograph mMR (Erlangen, Germany) (n=20) at 3 Tesla with an endorectal coil (Prostate eCoil; Medrad, Warrendale, PA). For the GE scanner, the clinical T2W protocol used repetition time 4000–13000 ms, echo time 156–164 ms, bandwidth 31.25 kHz, 2 averages, field of view 14 cm × 14 cm × 6.2 cm, slice thickness 3.3–3.6 mm, slice spacing 3.3–3.6 mm, 128 × 256 matrix size, 20–43 slices, flip angle 90°. The clinical DWI protocol used repetition time 4000 ms, echo time 70–77 ms, bandwidth 125 kHz, 3 averages, FOV 14 cm × 14 cm × 6.2 cm, slick thickness 2.2 mm, slice spacing 2.2 mm, 320 × 192 matrix size, 40 slices, flip angle 90°, b-
value 600—800. For the Siemens scanner, the clinical T2W protocol used repetition time 3500 ms, echo time 116 ms, bandwidth 115 Hz, 2 averages, field of view 18 cm × 18 cm, slick thickness 3.0 mm, slice spacing 3.0 mm, 24 slices, flip angle 120°. The clinical DWI protocol used repetition time 5200 ms, echo time 93 ms, bandwidth 1.158 kHz, 8 averages, FOV 26 cm × 26 cm, slick thickness 3.6 mm, 20 slices, b-value 100—800 s/mm².

2.2.3 Preprocessing

We adjusted for T2W spatial signal variation due to endorectal coil sensitivity using the N4ITK image bias correction algorithm in Slicer 4.2 (Surgical Planning Lab, Harvard Medical School, Boston, USA). We scaled MR image intensities such that the mean signal in manually segmented peri-prostatic fat regions matched an arbitrary reference value across all patients. This accounts for inter-patient differences in MR image intensity on T2W. Peri-prostatic fat was manually segmented as an additional pre-processing step.

2.2.4 Reference standard

Our group created the reference standard using fusion of mp-MRI with whole-mount digitized prostatectomy histology, with a mean target registration error of 1.7 mm for the T2W images and 1.9 mm for the ADC maps [15]. After whole-mount paraffin embedding of the prostate specimen post-radical prostatectomy with apex and seminal vesicles removed, a pathologist sliced the mid-gland into 4.4 mm transverse sections. Using a microtome, a pathology technician cut 4 µm slices from each section and stained with hematoxylin and eosin. Using a Scan Scope GL (Leica Biosystems, Concord, CA) bright field scanning system, the pathology team digitized the slides at 0.5 µm/pixel (24-bit color). A pathology assistant trained in PCa morphology annotated histology sections (intra-prostatic lesions graded and segmented), supervised by a genitourinary pathologist. One of two genitourinary pathologists at our institution verified and corrected all annotations, if necessary. The registration algorithm mapped
histology slides to the ex-vivo specimen using internal and external fiducials via an affine transformation. The algorithm then mapped ex-vivo to in-vivo MRI using fiducials corresponding to user-identified homologous anatomical landmarks (e.g. benign prostatic hyperplasia nodules, calcifications) via a thin-plate spline deformable registration. The PZ and CG were manually delineated on mp-MRI. PCa tissue masks $\geq 0.2 \text{ cm}^3$ (assuming spherical tumour shape) were considered clinically significant [24] and were retained for analysis.

2.2.5 Regions of interest

In determining the ROIs to use for this experiment, we aimed to develop a method to rigorously sample all the tissue on histology in an unbiased fashion. Hand drawing regions of healthy tissue is subject to selection bias (i.e. ROIs may be selected to contain no confounding tissue) and spatial bias (i.e. ROIs may be subconsciously drawn in similar areas for each slice). This may be overcome by using a simple method for iterating pre-drawn ROIs over every voxel of the histology section and making selections when the ROI fully intersects with pathology verified healthy tissue. However, if hand-crafted, radiomic texture features correlate with size and shape, then biases may be induced if all ROIs from one class have a significantly higher or lower average area of tissue analyzed compared to the other class. Thus, an algorithm to combat this problem must also be able to select ROIs to use from each class such that the average area of tissue analyzed is identical between classes (PCa ROIs and non-PCa ROIs).

Intuitively, the algorithm operates by using PCa tissue masks made by pathology and iterates them throughout all the histologic sections while making selections when it intersects with pathology verified non-PCa tissue. When a selection is made, the algorithm then iterates another PCa tissue mask, in descending order of size. Thus, the same PCa tissue mask may be used multiple times to generate non-PCa tissue masks. At this point, we are left with non-PCa ROIs that have been selected without any selection or spatial biases, however we have not addressed the size bias. We aimed to correct for this by balancing the mean area of tissue analyzed between PCa and non-PCa tissue. The mean was selected to compare the ROI areas
due to its sensitivity to outliers such as abnormally large or small ROIs that may have more significant impacts on classifier performance. Thus, after all tissue has been sampled and no more ROIs can be selected, the algorithm selects a fixed number of non-PCa ROIs to retain using an optimization function that maximizes the number of patients in the dataset while still trying to maintain an equal average area between PCa and non-PCa tissue ROIs.

Figure 2.1: Hybrid pseudocode and graphical illustration of the algorithm used to select non-PCa tissue masks. The algorithm takes a list C containing PCa tissue masks cropped to the bounding box of the PCa tissue and iterates them over a list M of masks, in image space, containing non-PCa tissue to produce a list N of non-PCa tissue masks. This algorithm is run once in the PZ and once in the CG. This algorithm thus generates two lists that contain non-PCa tissue masks in each zone respectively.

More specifically, the algorithm used to generate non-PCa tissue masks is illustrated in
Figure 2.1. We first ran this algorithm using only PZ tissue and then using only CG tissue. Once we generated all the non-PCa tissue masks, we grouped them according to the PCa tissue mask that was used to generate that set. To ensure area matching between classes, the algorithm determined a threshold to define the maximum number of masks used for each set of non-PCa tissue masks. If a set of non-PCa tissue masks contained more masks than an arbitrary threshold $x$, then we selected $x$ masks from that set of non-PCa tissue masks at random and kept them in the dataset. If a set of non-PCa tissue masks had fewer masks than an arbitrary threshold $x$, then we removed all masks from that set of non-PCa tissue masks from the dataset along with its associated PCa tissue mask. Thus, the number of non-PCa tissue masks retained can be modeled as a function of the threshold selected, $f(x)$. Furthermore, the number of unique patients retained in the dataset can be modeled as a function of the threshold selected, $g(x)$. For each threshold ranging from 1 to the global maximum number of masks created across all sets of non-PCa tissue masks, $X$, the algorithm maximized a linear cost function (Equation 2.1) to produce a final threshold $T$, defined as:

$$T = \text{arg max}_x \{ f(x) + \lambda g(x) | x \in \mathbb{Z} \land 1 \leq x \leq X \}$$  \hspace{1cm} (2.1)$$

where $x$ is the threshold, $f(x)$ is the total number of non-PCa tissue masks retained, $g(x)$ is the total number of unique patients retained, $\lambda$ is a constant and $T$ is the threshold that maximizes this linear equation. For each zone, we set $\lambda$ to be the average number of PCa tissue masks across all patients. This process ensured that the malignant PZ/CG PCa tissue masks and PZ/CG non-PCa tissue masks had evenly matched mean areas of tissue processed. This resulted in a total of 245 PZ non-PCa tissue masks with an average area of 60.7 mm$^2$ and 390 CG non-PCa tissue masks with an average area of 78.8 mm$^2$ with corresponding equal mean areas of PCa tissue masks.
2.2.6 Feature extraction and scaling

We extracted 21 first and 31 second-order texture features from each ROI in MATLAB 9.2 (The Mathworks, Natick, Massachusetts, USA). We generated gray level co-occurrence matrices (GLCM) with gray level quantization of 128 bins with a bin size of 64 gray levels. We calculated GLCMs using four neighbouring voxel pair directions in the 2D plane \([-1, 0), (-1, -1), (0,1), (1, -1)\] and each was treated as an independent feature and normalized. We also generated gray level run-length matrices (GLRLM) using a run threshold of 32 in four orientations \([-1, 0), (-1, -1), (0,1), (1, -1)\]. Before training and testing, we scaled features via standardization to have zero-mean and unit variance (Equation 2.2).

\[
x' = \left( \frac{x - \bar{x}}{\sigma} \right)
\]  

(2.2)

2.2.7 Feature selection

Out of 61 patients, we selected 15 candidates who best represented the entire dataset to perform feature selection. The 15 candidates were selected to contain an approximately equal distribution of patients imaged on the GE and Siemens scanners. For each scanner in the 15 patient subset, we obtained an approximately even ratio of Gleason grade 3+3, 3+4, 4+3, and 4+4 tumours as well as the corresponding healthy tissue samples generated from the algorithm described in Figure 2.1. These patients were not used in any cross-validation experiment. We performed feature selection using PRTools5 (Delft Pattern Recognition Research, Delft, The Netherlands) using the sum of estimated Mahalanobis distances as a feature-ranking metric to account for the variance of each feature and the covariance between features. We used greedy forward feature selection to select 1–10 features generating 10 feature sets in total.
2.2.8 Classification

We used PRTools5 for classification of the remaining 45 patients using four classifiers: a RF, LOGL, KNN and NAIVEB with 1–10 features selected for each, producing a total of 40 different classifiers. We validated our models using leave-one-patient-out (LOPO) and 10-fold cross validation. Patients in the testing set were not used in the training set in any fold. We reported the AUC, MCR, FNR, FPR, precision and recall at the ROC point with the highest true positive rate and lowest false positive rate, depicting ideal classification. We derived the reported error metrics from the classifier that had the highest AUC in the testing set. Figure 2.2 illustrates the complete pipeline from data acquisition and pre-processing to results reporting. Note that the ROIs used in these experiments are once again derived from pathology annotations of PCa, and not manually delineated by an imaging observer.

2.2.9 Inter-class area mismatch

In the Regions of Interest section, we discussed the development of a method for sampling non-PCa tissue without any selection or size bias. To further explore the effect size bias may have on system performance, we artificially induced a deviance in the evenly matched mean areas for PCa and non-PCa tissues as result of applying the algorithm defined in the Regions of Interest section. This was performed by iteratively eroding all the non-PCa tissue masks without touching the PCa tissue masks (and vice-versa) and performing multiple cross validation experiments- namely, 4-fold, 10-fold and LOPO cross validation. We eroded all tissue in MATLAB 9.2 using a disk-shaped structured element with radius of N voxels, where N increased by one voxel at each iteration until the stopping criterion was met. We defined the stopping criterion for this procedure as the iteration at which a mask disappeared due to the erosion process. This operation will primarily modify the size of the tissue mask, with shape preserved to the extent possible by this morphological operation. Thus, we aimed to determine if there were any positive or negative effects on classifier performance at each iteration of the erosion pro-
Figure 2.2: Flow diagram of the experimental pipeline. The red arrows indicate the pathology derived contours of PCa being mapped to the in-vivo imaging. No ROIs were manually drawn on the imaging sequences. Feature selection was performed on 15 patients (∼25%, 15/61) with cross-validation and results reporting being performed on the remaining 46 patients. ADC: apparent diffusion co-efficient map, T2W: T2-weighted, PCa: prostate cancer.

cess in comparison to baseline- potentially indicating a spurious result. We also performed the same experiment with the only difference being that non-PCa tissue was manually contoured on the contralateral side of the prostate. This was done to simulate the methodology that is common in other CAD papers [17–23]; classifier performance would then be evaluated with the mean-area mismatch and compared to the results from our erosion experiment.
2.2.10 Inter-class shape mismatch

We investigated the effects of inter-class shape mismatch by transforming the non-PCa tissue masks into area-equivalent regular circles, under the assumption that the degree of similarity between PCa tissue masks and regular circles is small. Subsequently, we performed a 10-fold cross-validation experiment to interrogate differences in shape that may affect classifier performance and to assess the variance in the resulting AUCs.

2.2.11 Inter-scanner differences

Since our dataset is comprised of mp-MRI acquired on two different scanners, we aimed to evaluate the robustness of the features and classifiers to scanner differences by training them on data from one scanner and testing on the other. We performed this experiment in two passes. In the first pass, we separated the data from each scanner into two folds and performed cross-validation. In the second pass, we distributed the data randomly such that each of the two folds had an equal ratio of data from each scanner; primarily due to the unequal distribution of patients imaged on each scanner (n=41 on the GE scanner, n=20 on the Siemens scanner). This allowed us to make a fair comparison of our systems robustness to scanner differences.

2.2.12 Inter-zonal differences

We repeated the original experiment but separated lesions located in the PZ and CG into two different cross-validation experiments. This was performed to interrogate system classification of PCa vs. non-PCa tissue on a per zone basis.

2.3 Results

We evaluated the results of a LOPO and 10-fold cross-validation experiment using four different classifiers each trained on 1 to 10 selected features. A 5-feature NAIVEB classifier
produced an AUC of 0.80 for LOPO cross-validation (Figure 2.3) and a median [IQR] AUC of 0.82 [0.73–0.91] for 10-fold cross validation (Table 2.1).

<table>
<thead>
<tr>
<th>Error metric</th>
<th>LOPO cross validation</th>
<th>Median</th>
<th>10-fold cross validation</th>
<th>Median</th>
<th>IQR</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUC</td>
<td>0.80</td>
<td></td>
<td>0.82</td>
<td>[0.73-0.91]</td>
<td></td>
</tr>
<tr>
<td>FNR</td>
<td>0.19</td>
<td>0.25</td>
<td>[0.00-0.28]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FPR</td>
<td>0.29</td>
<td>0.17</td>
<td>[0.12-0.30]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCR</td>
<td>0.28</td>
<td>0.18</td>
<td>[0.12-0.30]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Precision</td>
<td>0.23</td>
<td>0.32</td>
<td>[0.21-0.45]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recall</td>
<td>0.81</td>
<td>0.75</td>
<td>[0.71-1.00]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.1: Summary of error metrics for LOPO and 10-fold cross validation for a 5-feature NAIVEB classifier.

Figure 2.3: Receiver operating characteristic (ROC) curve (left) and a precision-recall ROC curve (right) for classification of PCa vs. non-PCa tissue in a LOPO cross validation experiment.

Features used in both LOPO and 10-fold cross-validation are described in Table 2.2. We found features 2 and 5 were found to be significantly different in a Mann-Whitney U test as they did not pass a normality test (p <0.05). Two features selected were first-order statistical features and three were second-order GLCM based features. Four of the five selected features
were derived from the ADC images.

<table>
<thead>
<tr>
<th>Feature Selection Order</th>
<th>Feature Class</th>
<th>Feature Name</th>
<th>Directionality/Modality</th>
<th>Mean (±std.) Cancer Feature Value</th>
<th>Mean (±std.) Non-cancer Feature Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1st Order Statistical</td>
<td>Kurtosis</td>
<td>-</td>
<td>ADC</td>
<td>3.16 ± 1.14</td>
</tr>
<tr>
<td>2</td>
<td>Gray-level-co-occurrence</td>
<td>Difference Average</td>
<td>90°</td>
<td>T2W</td>
<td>0.45 ± 0.26</td>
</tr>
<tr>
<td>3</td>
<td>Gray-level co-occurrence</td>
<td>Difference Average</td>
<td>0°</td>
<td>ADC</td>
<td>0.60 ± 0.34</td>
</tr>
<tr>
<td>4</td>
<td>Gray-level co-occurrence</td>
<td>Joint Entropy</td>
<td>0°</td>
<td>ADC</td>
<td>0.05 ± 0.03</td>
</tr>
<tr>
<td>5</td>
<td>1st Order Statistical</td>
<td>10th Percentile</td>
<td>-</td>
<td>ADC</td>
<td>914.4 ± 319.9</td>
</tr>
</tbody>
</table>

Table 2.2: Features selected from the 15-patient hold-out set used in LOPO and 10-fold cross-validation. Features were sequentially added using greedy forward feature selection by taking the set with the largest measured sum of estimated mahalanobis distances. Directionality refers to the direction of the target voxel from the reference in GLCM calculation. GLCM and GLRLM direction vectors are as follows: 0°: [1,0], 45°: [1,1], 90°: [0,1], 135°: [-1,1]. Bolded feature values were found to be significantly different between PCa vs. non-PCa tissue (p<0.05).

Figure 2.4 displays qualitative results from the LOPO cross-validation experiment. ROIs outlined in red are true negatives, green are true positives, blue are false positives and yellow are false negatives. Displayed are the T2W images, ADC images and relevant corresponding histology sections from left to right. Histology tissue contains annotations made by the pathologist with color coding depicting a specific disease pattern.
We analyzed the effect of training systems with average area mismatch between PCa and non-PCa tissue masks. Table 2.3 shows the AUCs calculated from a 10-fold, 4-fold and LOPO cross-validation experiment versus the percent difference of average area from baseline for erosion of PCa tissue masks (bottom) and non-PCa tissue masks (top). We chose these three validation schemes as they allow us to contextualize our results among several commonly reported validation schemes in the radiomics literature. Overall, we observed a trend of increasing median AUC as the average area mismatch between class labels increased across all cross-validation schemes. We also evaluated system performance using the non-eroded PCa tissue masks and manually contoured non-PCa tissue masks on the contralateral side of the prostate. The average area of the PCa and non-PCa tissue masks in the dataset was 127.5 mm² and 69.3 mm² respectively, which corresponds to a 59.1% difference. After performing a 10-fold cross validation experiment, we recorded a median AUC of 0.86.

<table>
<thead>
<tr>
<th>Cross validation scheme</th>
<th>Percent difference of average area from baseline (%)</th>
<th>Median AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>19.6</td>
</tr>
<tr>
<td>Eroded non-PCa tissue masks</td>
<td>10-Fold</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>4-Fold</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>LOPO</td>
<td>0.80</td>
</tr>
<tr>
<td>Eroded PCa tissue masks</td>
<td>10-Fold</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>4-Fold</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>LOPO</td>
<td>0.80</td>
</tr>
</tbody>
</table>

Table 2.3: Median AUC as a function of percent difference from baseline for multiple cross validation schemes (10-fold, 4-fold and LOPO). Experiment was performed by eroding non-PCa tissue masks (top) and PCa tissue masks (bottom) iteratively.

We calculated the results from 10-fold cross-validation experiments using both non-PCa tissue masks that matched PCa tissue mask shapes (referred to as “shape-matched non-PCa tissue masks”) and non-PCa tissue masks transformed to area-matched regular circles (referred to as “non shape-matched non-PCa tissue masks”). We aimed to examine if there was a depen-
Figure 2.4: From left to right T2W images, ADC images and contoured histology. On imaging, red regions represent true negatives, green represent true positives, yellow represent false negatives and blue represent false positives. On histology, green contours represent Gleason 3+3 cancer, dark yellow represent Gleason 4+4 cancer, and gray represent prostatic intraepithelial neoplasia.
dency on shape between masks of both classes even with no mean area mismatch. Figure 2.5 displays the resulting AUCs for shape-matched non-PCa tissue masks (left) and non-shape-matched non-PCa tissue masks (right). The median AUCs were 0.82 for both cases, and thus we did not observe any dependency on shape during classification of PCa vs. non-PCa tissue.

Of interest was the robustness of the features selected having been trained on data from one scanner and tested on data from another. We performed a folded experiment using the same features listed in Table 2.2 by inserting all the data from scanner one into fold one and all the data from scanner two into fold two. This was compared to another folded experiment where the data from both scanners were evenly distributed among both folds. Results from the 2-fold cross validation for each scanner split are listed in Table 2.4. We observed a drop in performance compared to the results from our experiments with classifiers trained on features from both scanners (AUC: 0.80 vs. 0.71).

By splitting up the ROIs from the PZ and CG, we determined the best performing systems for each zone listed in Table 2.5. AUC of the PZ was significantly lower than the AUC of the CG (AUC: 0.75 vs. 0.95 respectively). We selected features independently for each zone and did not use the same five features used in the combined model located in Table 2.2, suggesting that the features used in this study are much more sensitive to CG lesions than PZ lesions. In total, there were 20 PZ PCa tissue masks and 245 PZ non-PCa tissue masks as well as 30 CG PCa tissue masks and 390 CG non-PCa tissue masks used for this experiment (non-PCa tissue masks generated by the algorithm defined in Figure 2.1).
Figure 2.5: AUC calculated in a 10-fold cross validation experiment for shape-matched and non-shape-matched non-PCa tissue masks. Shape-matched non-PCa tissue masks refer to benign tissue masks that have the same shape as PCa tissue masks through the algorithm defined in Figure 2.1. Non-shape-matched non-PCa tissue masks refer to benign tissue masks that have been transformed into area-matched regular circles instead of their PCa shaped tissue masks. Boxplot displays the median and 10th to 90th percentile of AUC values.

<table>
<thead>
<tr>
<th>Error Metric</th>
<th>Scanner split</th>
<th>Equal distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUC</td>
<td>0.71</td>
<td>0.78</td>
</tr>
<tr>
<td>FNR</td>
<td>0.29</td>
<td>0.24</td>
</tr>
<tr>
<td>FPR</td>
<td>0.36</td>
<td>0.25</td>
</tr>
<tr>
<td>MCR</td>
<td>0.36</td>
<td>0.25</td>
</tr>
<tr>
<td>Precision</td>
<td>0.17</td>
<td>0.25</td>
</tr>
<tr>
<td>Recall</td>
<td>0.71</td>
<td>0.76</td>
</tr>
</tbody>
</table>

Table 2.4: 2-Fold cross validation experiment of PCa vs. non-PCa tissue with each fold being trained on data from one scanner and tested on the other (left: scanner split) and an equal ratio of scanner data in each fold (right: equal distribution).

2.4 Discussion

In this study, we aimed to evaluate the performance of classical machine learning techniques for classifying malignant vs. benign tissue on prostate mp-MRI validated on a dataset with ac-
Table 2.5: 10-Fold cross validation results for training and testing performed with regions of interest located only in the PZ (left) and CG (right). Error metrics displayed are reported with the median across all 10-folds and the interquartile range.

<table>
<thead>
<tr>
<th>Error metric</th>
<th>Median PZ</th>
<th>IQR PZ</th>
<th>Median CG</th>
<th>IQR CG</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUC</td>
<td>0.75</td>
<td>[0.60-0.88]</td>
<td>0.95</td>
<td>[0.92-0.99]</td>
</tr>
<tr>
<td>FNR</td>
<td>0.00</td>
<td>[0.00-0.17]</td>
<td>0.00</td>
<td>[0.00-0.00]</td>
</tr>
<tr>
<td>FPR</td>
<td>0.21</td>
<td>[0.03-0.39]</td>
<td>0.04</td>
<td>[0.00-0.15]</td>
</tr>
<tr>
<td>MCR</td>
<td>0.19</td>
<td>[0.05-0.34]</td>
<td>0.04</td>
<td>[0.00-0.14]</td>
</tr>
<tr>
<td>Precision</td>
<td>0.31</td>
<td>[0.00-0.68]</td>
<td>0.58</td>
<td>[0.25-0.91]</td>
</tr>
<tr>
<td>Recall</td>
<td>1.00</td>
<td>[0.83-1.00]</td>
<td>1.00</td>
<td>[1.00-1.00]</td>
</tr>
</tbody>
</table>

accurate fusion of prostate mid-gland histology to in-vivo imaging [15]. In our experiments, we discovered that a system using a NAIVEB classifier trained on five features produced an AUC of 0.80 in a LOPO cross-validation experiment. We explored the effect of eroding PCa and non-PCa tissue masks to create an average area mismatch between class labels. We observed a trend of increasing AUC as the average area mismatch increased, suggesting a size dependency when performing classification experiments with these hand-crafted features and classifiers for all cross-validation schemes. Furthermore, we explored the effect of creating an artificial shape mismatch between tissue masks by transforming all non-PCa tissue masks into area-matched regular circles. We found that there was no difference in performance after changing the shape of the ROI, suggesting that there is no shape dependency when performing classification experiments with these hand-crafted features and classifiers. We evaluated system robustness to different scanners in a 2-fold cross-validation experiment and found that there was a decrease in performance when one scanner was used for training and the other for testing. We also found that system performance was much better when trained and tested only on CG ROIs compared to PZ ROIs.

Characterization of prostate lesions independently from other clinical variables on mp-MRI has been summarized more recently by Bardis et al [25]. Most work was focused on application of machine learning and deep learning techniques to the PROSTATEx dataset. Several
authors have applied modified CNN to classify PCa vs. non-PCa tissue including multiple paired-CNNs [10], deep-layered CNNs [26, 27], and 3D-CNNs [28], with AUCs ranging from 0.80-0.96. These models were validated using MR-guided biopsy which may be biased towards tumours that can be detected by radiologists. Kwak et al. employed a support vector machine on a proprietary dataset of 244 patients and recorded an AUC of 0.89 [29]. This model was based on prostate biopsy specimens analyzed by a pathologist using an MRI-US fusion targeted biopsy machine. These datasets may be biased towards lesions that are only visible to the radiologist during the MRI-US fusion biopsy; complete information about cancer presence from the whole gland is missing. One of the more common approaches is consensus mapping where the radiologist delineates ROIs based on information from the pathologist on an approximate corresponding whole-mount transverse section [17, 18, 30]. These techniques allow for better whole gland mapping of disease; however, they are limited by their lack of quantified spatial error in registration between the fusion of histology to MRI. This produces a potential bias towards radiological false negative areas; underlying lesions that are not visible on MRI need to be mapped to their corresponding location within the gland. In our experiments, we addressed the potential biases with the aforementioned reference standards using a technique for highly accurate fusion of in-vivo mp-MRI to whole-mount digitized transverse histology post-prostatectomy with a measured target registration error [15]. Thus, we were able to evaluate the performance of a radiomics based machine learning model on contours of PCa and non-PCa on mp-MRI containing no bias that may be present from an imaging observer’s interpretation of the MRI signal.

The results of classification from four example patients in our dataset are illustrated in Figure 2.4. Patients in rows one and two displayed results where the lesion had poor visibility on mp-MRI. This is important, because it demonstrates the importance of having accurately co-registered prostatectomy histology as a reference standard for these versus user interpretation of where the MR lesion is located. A research group proposing to use MR-guided biopsy or consensus mapping based approaches to generating reference standards would have missed
these lesions, thus positively biasing their classification results because they are provided with ROIs that are more easily identifiable. The patient in row three had lesions that were all classified as benign tissue, with one false negative. The false negative was a lesion of high-grade disease (Gleason 4+4) located on the anterior right side of the prostate. The patient in row four had one lesion on this slice that was classified as a false positive. This appears to be in a region with benign prostatic hyperplasia which presented itself as a large hypointense region on imaging. This is a limitation of the specificity of MRI for cancer detection, benign confounders such as BPH have the same textural patterns as cancer on imaging which may explain why the system had difficulty identifying it as healthy tissue. Future work in this area would be tuning the system to be more specific towards regions of BPH.

We analyzed the effect on system performance as a function of inter-class area and shape mismatch. As mean area-mismatch increases between classes in our dataset, we saw a trend of increasing median AUC. This is supported by recent work published by Traverso et al. which demonstrated a large percentage of radiomic features to be dependent on tumour volume [13]. Our previous work demonstrated a similar trend of increasing AUC versus eroded non-PCa tissue but was limited by the use of feature selection at each round of cross-validation thus causing overfitting of the model [31]. In this experiment, the same five features extracted in Table 2.2 were used at each stage in the erosion process, effectively eliminating the bias. Yet, we still observed a similar trend using a separate tuning set (n=15) for feature selection and applying it to the experimental set (n=46). Classification using the non-eroded PCa tissue masks and manually contoured non-PCa tissue masks (percent difference of mean area between PCa and non-PCa was 59.1%) produced an AUC of 0.86. This coincides with the results we listed in Table 2.3, where we observed an AUC of 0.86 from a 10-fold cross-validation experiment with a percent difference of mean area between PCa and non-PCa tissue of 56.9% - similar to the aforementioned experiment (59.1% vs. 56.9%). Thus, the two systems reported the same resulting AUC while having roughly the same mean-area differences between PCa and non-PCa, further suggesting that there may be some volume dependency on the handcrafted radiomic-
features used in this experiment. System performance as a function of shape was analyzed by converting tumour shaped non-PCa tissue masks to area matched regular circular ROIs. In our experiments, we found no difference in system performance between models trained on each of the two non-PCa tissue mask shapes (AUC: 0.82 vs. 0.82); thus, suggesting that ROI shape has no influence on classification of malignant vs. benign regions. These findings are important for clinical translation; users prompted to delineate a suspicious region need not consider the shape of their contour but only the size, which is a much more manageable task.

Our dataset is comprised of mp-MRI acquired on two different scanners with different imaging parameters as described in the materials and methods section. We were interested in the robustness of imaging features derived from our tuning set by training our experimental set on features from one scanner and testing on the other. We recorded an AUC of 0.71 [0.70-0.72] (median [IQR]), which signified an expected drop in performance in comparison to our experiments with models trained on features from both scanners (AUC: 0.78 vs 0.71). Classification on separate zones within the prostate instead of the whole gland has been previously suggested to provide improved sensitivity [32]. Previous studies have explored the development of models limited to different zonal anatomy [18, 31, 33]. We performed a similar analysis by dividing ROIs into two categories: PZ and CG. We recorded an AUC of 0.75 [0.60-0.88] in the PZ and 0.95 [0.92-0.99]. These findings are similar to those found in Viswanath et al. [33] who recorded an AUC of 0.86 in the CG and 0.73 in the PZ using imaging texture features derived from each zone and classifying using Quadratic Discriminant Analysis.

The results of our study should be interpreted in the context of its limitations. Our dataset was derived from a cohort of patients that went to prostatectomy and thus a selection bias was induced. Although our use of co-registered histopathology as a reference standard for the presence or absence of cancer in mp-MRI ROIs eliminates bias related to variable radiological saliency of positive and negative regions, an important next step is for our system’s performance to be tested on ROIs input by radiologist observers. All prostate whole-mount specimens were analyzed and contoured by one of two pathologists at our local institution.
which is known to be an observer-dependent process and thus may affect the accuracy of the pathological assessment of malignant tissue. We did not perform classifier hyperparameter tuning in these experiments opting to use default hyperparameters in the PRTools5 package; optimization of hyperparameters may produce better results. Error metrics dependent on the choice of operating point on the ROC curve were selected to have the highest true positive rate and lowest false positive rate and may not reflect actual performance of the system until external validation is performed to ensure system generalizability. No deep learning methods have been investigated up to this point, it is unknown whether these non-linear systems can overcome the biases in size mismatch between classes.

2.5 Conclusions

We developed and validated a radiomics based machine learning system using five features with a NAIVEB classifier to classify PCa vs. non-PCa prostate tissue on mp-MRI with an AUC of 0.80 using accurately co-registered mid-gland histology with a measured target registration error and histologically derived PCa ROI delineation. Non-PCa tissue masks were systematically sampled, and size and shape matched to eliminate any potential selection and/or spatial biases. We found a trend of increasing median AUC as inter-class size mismatch increased, which suggests that experiments that do not account for area of tissue analyzed may have positively biased results. There was found to be no dependency on shape for non-PCa tissue sampled in our experiments. System performance dropped when training on data from one scanner and testing on the other. System performance was much higher when trained strictly on lesions from the CG compared to the PZ. Once fully validated, the system may improve lesion characterization and treatment selection for PCa patients in the clinic.
References


t1-w, and dw automatically registered images,” in Medical Imaging 2013: Computer-
Aided Diagnosis, vol. 8670, p. 86703E, International Society for Optics and Photonics,
2013.

J. L. Chin, S. Pautler, and A. D. Ward, “Development of a computer aided diagnosis model
for prostate cancer classification on multi-parametric mri,” in Medical Imaging 2018:
Computer-Aided Diagnosis, vol. 10575, p. 1057515, International Society for Optics and
Photonics, 2018.

[21] X. Min, M. Li, D. Dong, Z. Feng, P. Zhang, Z. Ke, H. You, F. Han, H. Ma, J. Tian,

enberg, A. Mühlberg, M. S. Michel, S. O. Schoenberg, et al., “Multiparametric mri for

[23] S. Monti, V. Brancato, G. Di Costanzo, L. Basso, M. Puglia, A. Ragozzino, M. Salvatore,
and C. Cavaliere, “Multiparametric mri for prostate cancer detection: New insights into
the combined use of a radiomic approach with advanced acquisition protocol,” Cancers,

[24] J. I. Epstein, P. C. Walsh, M. Carmichael, and C. B. Brendler, “Pathologic and clinical findings to predict tumor extent of nonpalpable (stage t1 c) prostate cancer,” Jama,

Bui, M. Rupasinghe, C. G. Filippi, and D. S. Chow, “Applications of artificial intelligence


Chapter 3

Histologic validation of auto-contoured dominant intraprostatic lesions on $^{18}$F-DCFPyL PSMA-PET imaging

3.1 Introduction

Accurate, non-invasive delineation of areas of highest suspicion for biologically significant PCa is necessary for proper risk stratification and further treatment planning. Focal therapies and focal radiation boosting techniques are becoming an attractive option due to their improved morbidity profiles, while guided biopsy is showing promise towards improved sensitivity of lesion identification as opposed to systematic TRUS guided biopsy [1, 2]. Limitations in delineating intra-prostatic foci by conventional imaging like CT and ultrasound are barriers to such applications.

More recently, mp-MRI has become available for identification of intra-prostatic foci, with standards around mp-MRI acquisition and guidelines for delineation being introduced to aid in clinical implementation [3]. Despite advances in mp-MRI PCa detection, there are still known limitations including: the invisibility of some clinically significant lesions [4], significant intra-
and inter-observer variability [5], and significant false-positive rates in confounding regions (e.g. benign prostatic hyperplasia) [6]. In addition, accuracy of cancer detection of extraprostatic pelvic disease such as lymphadenopathy is similar to conventional imaging like CT.

PSMA-PET is emerging as a clinically useful tool for imaging PCa [7]. Previous work investigating the efficacy of PSMA targeted imaging with $^{68}$Ga-labeled PSMA ligands suggested higher sensitivity and specificity rates compared to alternative techniques [8]. $^{18}$F-DCFPyL targets PSMA with the added advantage of efficient cyclotron-based production and higher spatial resolution than Gallium-based agents [9]. Usefulness of PSMA based imaging for the restaging of men with suspected recurrent PCa has been demonstrated [8]; however, the use of PSMA based imaging for intra-prostatic lesion detection and delineation has been explored to a lesser extent.

In the setting of primary treatment and biopsy, high rates of identification of intra-prostatic foci have been noted and high concordance of lesions with elevated maximum SUV$_{\text{max}}$ with involved sextants or positive biopsies have been found [10–12]. However, the ability of PSMA-PET to delineate boundaries of DILs within the prostate, validated against a high-fidelity reference standard, remains poorly described. Using an existing database of men enrolled prospectively on a trial of multi-modality imaging prior to radical prostatectomy, we sought to investigate recommendations for threshold and margin expansions surrounding focal therapy/boosting and guided biopsy applications using $^{18}$F-DCFPyL PET images.

### 3.2 Methods and materials

#### 3.2.1 Patient Population

We analyzed 81 male patients with biopsy-proven PCa as part of a larger multimodality image guided PCa therapy study (registration number NCT04009174). Research was approved by the Institutional Review Board (the Health Sciences Research Ethics Board), and we obtained informed consent from all patients. A subset of 20 patients underwent $^{18}$F-DCFPyL PET imaging
on a hybrid PET/MRI scanner and of these, 12 had lesions suitable for this analysis. For this study, we were interested in patients with a DIL in the mid-gland with corresponding PET focus within the prostate, as our registration framework is optimized for registration of transverse and not the sagittal histologic sections taken from the base and apex at our institution. The DIL was defined as being the largest and most threatening (i.e. highest Gleason score) clinically significant lesion (volume >0.2cc assuming spherical tumour shape) within the prostate [13]. Other sparse malignant disease was excluded from the analysis. As noted, 12 patients had pathology with confirmed cancer within the mid-gland section; the remaining patients had extensive disease in the apex and/or base regions, with one patient having considerable misalignment between PET and MRI that was unable to be corrected through manual rigid registration. Patients involved in this study were enrolled from March 2016–January 2018. Mean (± SD) patient age was 62 ± 4 years (median 63, interquartile range 58–68) and mean (±SD) PSA at biopsy was 9.3 ± 6.1 ng/mL (median 7.3, interquartile range 0.1–14.5) (Table 3.1).

<table>
<thead>
<tr>
<th>Patient number</th>
<th>Pre-biopsy PSA (ng/mL)</th>
<th>Clinical staging</th>
<th>Gleason Score (DIL)</th>
<th>Pathological staging (pT)</th>
<th>Pathological staging (pN)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8.2</td>
<td>cT1c</td>
<td>4+4</td>
<td>pT3a</td>
<td>pN0</td>
</tr>
<tr>
<td>2</td>
<td>25.6</td>
<td>cT2a</td>
<td>3+3</td>
<td>pT3a</td>
<td>pN0</td>
</tr>
<tr>
<td>3</td>
<td>12.9</td>
<td>cT2a</td>
<td>3+4</td>
<td>pT3a</td>
<td>pN0</td>
</tr>
<tr>
<td>4</td>
<td>3.7</td>
<td>cT1c</td>
<td>4+3</td>
<td>pT2c</td>
<td>pN0</td>
</tr>
<tr>
<td>5</td>
<td>8.9</td>
<td>cT2c</td>
<td>4+3</td>
<td>pT3b</td>
<td>pN0</td>
</tr>
<tr>
<td>6</td>
<td>5.2</td>
<td>cT2b</td>
<td>3+4</td>
<td>pT2c</td>
<td>pN0</td>
</tr>
<tr>
<td>7</td>
<td>12.1</td>
<td>cT1c</td>
<td>3+4</td>
<td>pT2c</td>
<td>pN0</td>
</tr>
<tr>
<td>8</td>
<td>12.9</td>
<td>cT2b</td>
<td>4+4</td>
<td>pT3a</td>
<td>pN0</td>
</tr>
<tr>
<td>9</td>
<td>4.2</td>
<td>cT1c</td>
<td>3+3</td>
<td>pT2c</td>
<td>pN0</td>
</tr>
<tr>
<td>10</td>
<td>6.1</td>
<td>cT3a</td>
<td>3+4</td>
<td>pT2c</td>
<td>pN0</td>
</tr>
<tr>
<td>11</td>
<td>6.4</td>
<td>cT2a</td>
<td>4+4</td>
<td>pT3a</td>
<td>pN0</td>
</tr>
<tr>
<td>12</td>
<td>5.4</td>
<td>cT2a</td>
<td>4+3</td>
<td>pT2c</td>
<td>pN0</td>
</tr>
</tbody>
</table>

Table 3.1: Patient information for subjects involved in this analysis.

Study inclusion criteria were patients with pathologically confirmed PCa on previous biopsy who had elected to undergo radical prostatectomy for their primary treatment. Exclusion criteria included: (1) prior therapy for PCa, (2) use of 5-alpha reductase inhibitors within 6
months of study start, (3) inability to comply with pre-operative imaging, (3) allergic reactions to contrast agents, (4) contraindications to MRI, (5) insufficient renal function (eGFR < 60 mL/min/1.73 m²).

3.2.2 Registering Histology to Imaging

A detailed description of the in-vivo imaging parameters and histology processing can be found in Appendix A. Registration was performed using our group’s previously validated method in a two-step process by registering histology to ex-vivo MRI and ex-vivo MRI to in-vivo MRI with a quantified target registration error of 1.7 mm on T2W imaging. The registration of MRI to PET arises from the simultaneous acquisition of the two modalities; thus we have a complete pipeline for registering histology to PET [14].

3.2.3 Thresholding and Sensitivity/Specificity Analysis

Whole prostate volume was manually segmented on MRI approximately 5 mm inferior to the base to mitigate SUV signal bleeding from the bladder, before being transferred to the PET image. Segmented PET volumes were then thresholded from 1–100\% of SUV\textsubscript{max} in 1\% intervals. At each interval, a margin of 0–30 voxels (0–19.32 mm at 0.644 mm/voxel) was applied in one-voxel increments using morphological dilation with a spherical structuring element (limited by the prostate boundary), for a total of 3,000 different segmentations for each patient. All thresholding and margin expansion was performed in the in-vivo 3D MRI space by up-sampling the PET images to match the voxel resolution of the MRI (see Appendix A) in MATLAB 9.4 (The Mathworks, Natick, Massachusetts, USA). We calculated sensitivity and specificity on the 2D oblique histologic planes that intersected with the 3D segmentation on a voxel-by-voxel basis for each patient, which was then averaged across all patients. Sensitivity was calculated as the number of voxels in the segmentation that intersected with the DIL divided by the number of voxels in the segmentation on the 2D slice. Specificity was calculated as the number of voxels
outside of the segmentation divided by the number of voxels outside the DIL on the 2D slice. We determined the thresholds and margin expansions that produced (1) an average sensitivity $\geq 95\%$ with maximal specificity (for lesion-directed therapy) and (2) an average specificity $\geq 95\%$ with maximal sensitivity (for guided biopsy). Furthermore, we compared the results of the automatic segmentation against manual DIL contours using window/level of minimum to maximum SUV from 0-5 by a radiation oncology resident at our institution.

### 3.3 Results

Patient imaging characteristics are listed in Table 3.2. The mean ($\pm$SD) dose administered and imaging time from injection to the patients in this subset were $326.6 \pm 12.2$ MBq and 1:54:36 $\pm$ 0:09:09 (HH:MM:SS) respectively. Median (IQR) DIL SUV$_{\text{max}}$ was 9.5 (4.3–14.8). Median (IQR) DIL lesion area was 284.0 (77.0–491.4) mm$^2$ measured on the 2D oblique histology slices in the mid-gland for each patient.

<table>
<thead>
<tr>
<th>Patient number</th>
<th>Total dose (MBq.)</th>
<th>Imaging time from injection (HH:MM:SS)</th>
<th>Weight (kg.)</th>
<th>DIL SUV$_{\text{max}}$</th>
<th>DIL diameter (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>327</td>
<td>1:44:20</td>
<td>85</td>
<td>27.6</td>
<td>9.6</td>
</tr>
<tr>
<td>2</td>
<td>328</td>
<td>1:45:20</td>
<td>82</td>
<td>8.3</td>
<td>15.9</td>
</tr>
<tr>
<td>3</td>
<td>341</td>
<td>1:57:49</td>
<td>80</td>
<td>11.1</td>
<td>8.8</td>
</tr>
<tr>
<td>4</td>
<td>311</td>
<td>2:14:17</td>
<td>65</td>
<td>4.3</td>
<td>10.0</td>
</tr>
<tr>
<td>5</td>
<td>302</td>
<td>1:47:50</td>
<td>110</td>
<td>9.4</td>
<td>15.9</td>
</tr>
<tr>
<td>6</td>
<td>321</td>
<td>1:47:18</td>
<td>98</td>
<td>9.7</td>
<td>9.1</td>
</tr>
<tr>
<td>7</td>
<td>333</td>
<td>2:02:59</td>
<td>103</td>
<td>6.6</td>
<td>9.2</td>
</tr>
<tr>
<td>8</td>
<td>333</td>
<td>1:57:20</td>
<td>84</td>
<td>48.3</td>
<td>11.0</td>
</tr>
<tr>
<td>9</td>
<td>344</td>
<td>1:48:58</td>
<td>91</td>
<td>11.2</td>
<td>8.8</td>
</tr>
<tr>
<td>10</td>
<td>335</td>
<td>2:00:55</td>
<td>87</td>
<td>5.2</td>
<td>7.4</td>
</tr>
<tr>
<td>11</td>
<td>317</td>
<td>2:02:12</td>
<td>85</td>
<td>5.2</td>
<td>10.8</td>
</tr>
<tr>
<td>12</td>
<td>327</td>
<td>1:53:51</td>
<td>91</td>
<td>9.8</td>
<td>8.0</td>
</tr>
</tbody>
</table>

Table 3.2: Imaging information for each patient analyzed in this study.

The average sensitivity and average specificity versus SUV$_{\text{max}}$ threshold and margin expansion is shown in Figure 3.1. With minimum average sensitivity $\geq 95\%$, a threshold of 67%
of SUV\textsubscript{max} with an 8.4 mm margin achieved an average sensitivity and specificity of (mean ± SD) 95.0 ± 7.8 % and 76.4 ± 14.7 %, respectively. With minimum average specificity ≥95%, a threshold of 81% SUV\textsubscript{max} with a 5.2 mm margin achieved an average sensitivity and specificity of (mean ± SD) 65.1 ± 28.4 % and 95.1 ± 5.5 %, respectively.

Figure 3.1: Contour plots of average sensitivity (left) and average specificity (right) for segmentations on SUV-corrected PET corresponding to co-registered histology.

Segmentation characteristics for each oblique slice per patient are detailed in Table 3.3. Patient ID is indicated by patient number and slice number, separated by a hyphen. Segmentation and prostate areas were measured on the 2D oblique histologic planes that intersect with their respective 3D volumes.

For sensitivity ≥95%, ratio of the segmentation to whole-gland areas encompasses more than a third of the prostate. Out of 12 patients, 2 out of 12 had segmentations that expanded beyond 50% of the prostate. One patient had underlying histology with average area across all slices approaching 50% of the whole-gland (49.5 ± 7.0 %); the other had an average area less than a third of the prostate (16.4 ± 1.9 %) indicating an over-segmentation. Median area of segmentations was approximately double that of underlying histology.

For specificity ≥95%, ratio of the segmentation to whole-gland areas was very much similar
to the size of the lesions (15.0 % [1.0–29.1 %]). Out of 12 patients, 2 out of 12 had no overlap between the segmentation and the histologic lesion on one or more slice(s). Median area of the segmentations was approximately the same as that of underlying histology on each oblique slice (170.0 mm$^2$ vs. 159.3 mm$^2$).

Qualitative results of the segmentations on each oblique histologic plane are presented in Figure 3.2 and Figure 3.3. Each image is identified by patient number and slice number separated by a hyphen corresponding to the patient ID in Table 3.3. For example, patient ID: 1-1 (indicated in row one of Table 3.3) is illustrated in row one column two in Figure 3.2 and Figure 3.3; as identified by the patient ID in the upper left corner inside the white box. To make the patterns of enhancement clear, each subfigure was window and levelled from 0–SUV$_{\text{max}}$ within the image. Thus, each subfigure presents a different range of SUV values within the image.

To simulate testing on external unseen data we conducted a LOPO cross validation experiment by selecting the best parameters in the training set under both the focal therapy/boost criteria and the guided biopsy criteria. In the focal therapy/boosting setting with minimum mean sensitivity $>95\%$ in the training set, we recorded a mean (± std.) sensitivity of 93.4 ± 10.8% and a mean (± std.) specificity of 76.5 ± 14.9% on the test set. In the guided biopsy setting with minimum mean specificity $>95\%$ in the training set, we recorded a mean (± std.) sensitivity of 64.3 ± 30.1% and a mean (± std.) specificity of 94.5 ± 7.2% on the test set.

Using the same analysis, segmentations made by a radiation oncology resident produced a sensitivity of (mean ± SD) 62.8 ± 26.6% and specificity of 91.2 ± 10.4%. Out of the twelve patients the resident contoured, three patients (patient 2, 4, and 9) had the target DIL missed (sensitivity and specificity averages reflect the absence of those patients).
Figure 3.2: Qualitative results of segmentations using a threshold of 67% of SUVmax with an 8.4 mm margin. Each oblique histology plane is numbered by the patient number and the slice number separated by a hyphen. The dominant intra-prostatic lesion defined by histology is shaded in yellow in front of the SUV-converted PET image. The segmentation outline crossing the oblique histologic plane is coloured in red. The prostate boundary is outlined in a white dashed curve. Each image is intensity scaled based on the colour bar adjacent to it. Slices of the same patient are grouped by the black boxes.
Figure 3.3: Qualitative results of segmentations using a threshold of 81% SUVmax with a 5.2 mm margin. Each oblique histology plane is numbered by the patient number and the slice number separated by a hyphen. The dominant intra-prostatic lesion defined by histology is shaded in yellow in front of the SUV converted PET image. The segmentation outline crossing the oblique histologic plane is coloured in red. The prostate boundary is outlined in a white dashed curve. Each image is intensity scaled based on the colour bar adjacent to it. Slices of the same patient are grouped by the black boxes.
<table>
<thead>
<tr>
<th>Patient ID</th>
<th>Lesion area (mm²)</th>
<th>Segmentation area (mm²)</th>
<th>Prostate total area (mm²)</th>
<th>Ratio of lesion to whole-gland area (%)</th>
<th>Ratio of segmentation to whole-gland area (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1</td>
<td>147.23</td>
<td>329.72</td>
<td>739.06</td>
<td>19.92</td>
<td>44.61</td>
</tr>
<tr>
<td>1-2</td>
<td>175.85</td>
<td>403.95</td>
<td>1085.78</td>
<td>16.20</td>
<td>37.20</td>
</tr>
<tr>
<td>1-3</td>
<td>107.42</td>
<td>410.17</td>
<td>1116.47</td>
<td>9.62</td>
<td>36.74</td>
</tr>
<tr>
<td>2-1</td>
<td>103.27</td>
<td>182.90</td>
<td>1312.64</td>
<td>7.87</td>
<td>13.93</td>
</tr>
<tr>
<td>2-2</td>
<td>379.48</td>
<td>398.56</td>
<td>1386.88</td>
<td>27.36</td>
<td>28.74</td>
</tr>
<tr>
<td>2-3</td>
<td>479.02</td>
<td>588.93</td>
<td>1472.31</td>
<td>32.54</td>
<td>40.00</td>
</tr>
<tr>
<td>3-1</td>
<td>146.82</td>
<td>257.55</td>
<td>978.36</td>
<td>15.01</td>
<td>26.32</td>
</tr>
<tr>
<td>3-2</td>
<td>175.85</td>
<td>340.91</td>
<td>1155.04</td>
<td>15.22</td>
<td>29.52</td>
</tr>
<tr>
<td>3-3</td>
<td>121.93</td>
<td>379.48</td>
<td>1247.53</td>
<td>9.77</td>
<td>30.42</td>
</tr>
<tr>
<td>4-1</td>
<td>190.78</td>
<td>326.40</td>
<td>879.66</td>
<td>21.69</td>
<td>37.11</td>
</tr>
<tr>
<td>5-1</td>
<td>374.51</td>
<td>493.12</td>
<td>630.40</td>
<td>59.41</td>
<td>78.22</td>
</tr>
<tr>
<td>5-2</td>
<td>453.72</td>
<td>564.04</td>
<td>848.96</td>
<td>53.44</td>
<td>66.44</td>
</tr>
<tr>
<td>5-3</td>
<td>459.53</td>
<td>584.78</td>
<td>996.61</td>
<td>46.11</td>
<td>58.68</td>
</tr>
<tr>
<td>5-4</td>
<td>476.53</td>
<td>584.78</td>
<td>1007.81</td>
<td>47.28</td>
<td>58.02</td>
</tr>
<tr>
<td>5-5</td>
<td>416.81</td>
<td>584.78</td>
<td>1007.81</td>
<td>41.36</td>
<td>58.02</td>
</tr>
<tr>
<td>6-1</td>
<td>126.49</td>
<td>367.04</td>
<td>872.60</td>
<td>14.50</td>
<td>42.06</td>
</tr>
<tr>
<td>6-2</td>
<td>155.94</td>
<td>463.67</td>
<td>1038.50</td>
<td>15.02</td>
<td>44.65</td>
</tr>
<tr>
<td>7-1</td>
<td>159.26</td>
<td>292.80</td>
<td>1276.14</td>
<td>12.48</td>
<td>22.94</td>
</tr>
<tr>
<td>7-2</td>
<td>126.49</td>
<td>292.80</td>
<td>1293.15</td>
<td>9.78</td>
<td>22.64</td>
</tr>
<tr>
<td>8-1</td>
<td>229.35</td>
<td>552.43</td>
<td>1260.80</td>
<td>18.19</td>
<td>43.82</td>
</tr>
<tr>
<td>9-1</td>
<td>85.44</td>
<td>65.53</td>
<td>935.64</td>
<td>9.13</td>
<td>7.00</td>
</tr>
<tr>
<td>9-2</td>
<td>145.57</td>
<td>223.96</td>
<td>987.90</td>
<td>14.74</td>
<td>22.67</td>
</tr>
<tr>
<td>10-1</td>
<td>104.93</td>
<td>292.80</td>
<td>941.04</td>
<td>11.15</td>
<td>31.12</td>
</tr>
<tr>
<td>11-1</td>
<td>220.64</td>
<td>749.84</td>
<td>1246.28</td>
<td>17.70</td>
<td>60.17</td>
</tr>
<tr>
<td>11-2</td>
<td>202.39</td>
<td>902.88</td>
<td>1348.72</td>
<td>15.01</td>
<td>66.94</td>
</tr>
<tr>
<td>12-1</td>
<td>111.98</td>
<td>283.26</td>
<td>1308.08</td>
<td>8.56</td>
<td>21.66</td>
</tr>
<tr>
<td>12-2</td>
<td>121.52</td>
<td>321.01</td>
<td>1593.83</td>
<td>7.60</td>
<td>20.1</td>
</tr>
</tbody>
</table>

Table 3.3: Reported results for focal therapy and guided biopsy applications. Data are reported for each oblique histologic section labeled by patient ID and slice number separated by a hyphen.
3.4 Discussion

In the setting of focal therapy/boosting, we intended to produce a sensitivity of lesion segmentation >95%; and therefore, aimed to maximize specificity. This coincides with the main goal of focal therapy and focal boosting treatments; target the entire lesion of interest and spare healthy tissue. We found a threshold of 67% of SUV$_{\text{max}}$ with an 8.4 mm margin produced segmentations with an average sensitivity and specificity of 95.0 ± 7.8% and 76.4 ± 14.8% respectively. Given that our threshold analysis was based on means, our results are susceptible to outliers. For example, applying this threshold + margin expansion in patient 11 resulted in a gross over-segmentation of the underlying histology, largely due to the presence of two distinct foci more noticeable in slice two. The pathology report showed that this patient had high-density prostatic intraepithelial neoplasia in this area, which has been reported to over-express PSMA and thus represents a false-positive [15]. Another notable outlier is patient 2, containing a Gleason Score 6 lesion with the largest volume (1702 mm$^3$), and low to moderate uptake within the prostate. We investigated further by performing the same analysis with the removal of patient 2 in the dataset. In the focal therapy/boosting setting, we saw the mean specificity rise (78.4% vs. 76.4%) with a reduction in standard deviation for both sensitivity (6.2% vs. 7.81%) and specificity (13.8% vs. 14.7%). In the guided biopsy setting, we saw the mean sensitivity rise (69.1% vs. 65.0%) with a reduction in standard deviation for both sensitivity (24.6% vs. 28.4%) and specificity (4.4% vs. 5.51%). This case pushed the threshold and margin expansion parameters to produce a larger segmentation, further diminishing the average specificity attained. Beyond conventional focal therapy strategies, boosting techniques focused at dominant tumour locations may opt to relax the constraints on sensitivity while choosing to be more specific as physicians often incorporate planning margins. In this experiment, high sensitivity and high specificity criteria represent two extrema in DIL segmentation; what remains to be seen is optimal criteria that lie between these points which produce highest dose to the DIL while minimizing dose to surrounding tissue. One could then determine the optimal criteria to yield feasible targets for radiation dose delivery in all techniques (e.g. EBRT,
In the setting of guided biopsy, we intended to produce a specificity of lesion segmentation >95%; and therefore, aimed to maximize sensitivity. This coincides with the main goal of guided biopsy: ensure needle targeting is in a region of verified cancer to not miss any disease. Through our analysis we found a threshold of 81% SUV$_{\text{max}}$ within the prostate with a 5.2 mm margin to produce segmentations with an average sensitivity and specificity of 65.1 ± 28.4 % and 95.1 ± 5.5 % respectively. Previous work investigating the probability of hitting lesions in one or two attempts found that the upper bound of the 99% prediction interval for the probability of one positive sample in a biopsy core ≥95% to be 1.05 cm$^3$ [16]. We found that one out of the 12 patients had a DIL lesion size that was less than the upper bound prediction interval; suggesting that all patients but one could have their DIL biopsied in two attempts. This is a promising result especially in the setting of biochemical failure for characterizing patterns of recurrence where MRI post radiation therapy tends to be confounded. However, PSMA-PET is showing promise towards becoming an effective imaging tool for early detection of PCa recurrence [17].

Manual contours performed by the radiation oncology resident by adjusting window/level from minimum to maximum SUV of 0-5, had high specificity, similar to the criteria provided for guided biopsy. Out of the 12 patients assessed, the resident’s contours did not intersect with the identified DIL on histology for three patients. Despite this, the resident had average sensitivity and specificity of 62.8 ± 26.6% and 91.2 ± 10.4%, respectively, for the remaining nine patients. These results are similar to those reported by Zamboglou et al., where teams of expert observers tended to produce contours with high median specificities and lower median sensitivities [18]. In comparison, our threshold and margin expansion recommendations for guided biopsy produced an average sensitivity of 65.1 ± 28.4% and an average specificity of 95.1 ± 5.5% and enabled hitting the target DIL on all twelve patients. Differences in sensitivity and specificity may be due to the analysis performed at a patient level as well as their
stratification of disease being clinically significant cancer (Gleason score >7) vs. other. Our experiments were primarily focused at segmenting the DIL regardless of pathological grade. This work suggests that a combination of our recommended threshold and margin expansions along with user refinement and assessment may be the best method for clinical implementation. Future work involving the contouring of associated multi-parametric MRI alongside PET with more observers is required to assess the incremental benefit of semi-automatic contouring of the PET images of intraprostatic lesion delineation.

In the context of lesion contouring, radiotracer selection plays a large role for primary staging as well as metastatic involvement. Gallium-based PSMA radiotracers have shown to exhibit weaker PET tumour-to-background uptake ratios than 18F based radiotracers (e.g. 18F-DCFPyL and 18F-PSMA-1007) [19]. Thus, we hypothesize more accurate segmentations would be made on 18F based tracers; however, there is a paucity of literature comparing accuracy of lesion contouring for primary staging within the gland between Ga based and 18F based radiotracers. In the context of fluorine based radiotracers, 18F-DCFPyL and 18F-PSMA-1007 have shown to exhibit excellent image quality with no statistical difference between SUV_max values in lesions between images taken with either tracer [20]. Thus, we hypothesize that the same contouring strategies we suggest through these experiments could be applied to PET imaging taken with other 18F based radiotracers; however, validation through another study is required. Furthermore, our SUV_max thresholds of 67% and 81% for the focal therapy/boosting and guided biopsy setting respectively are at the high end compared to those reported in the literature (typically 40-50%). We repeated our experiments and limited our search domain to have an upper bound of 50% SUV_max and noted that a drop in mean specificity for the focal therapy/boosting case (63.6% vs. 76.4%) as well as a drop in mean sensitivity for the guided biopsy case (57.8% vs. 65.1%).

To the best of our knowledge, we are the first group to analyze DIL segmentation performance using a combination of percentage of SUV_max in the prostate and margin expansion for 18F-DCFPyL based PSMA-PET imaging, validated on co-registered whole mount digi-
tized histology [14]. Previous studies have recommended segmentation methods using static methods such as SUV cutoffs and dynamic methods such as percentage of SUV\textsubscript{max}. Lopci et al. suggested an SUV cutoff of 4.8 to differentiate cancer from benign tissue with 82.4% sensitivity and 72.2% specificity validated on software based fusion biopsy [10]. Lower sensitivity and specificity may be due to the use of a Gallium-based PSMA radiotracer, which has been shown to exhibit weaker PET tumour-to-background uptake ratios than its \(^{18}\text{F}-\text{DCFPyL}\) counterpart [19]. Zamboglou et al. suggested a threshold of 20% SUV\textsubscript{max} to achieve a \(>95\%\) median sensitivity with median specificity of 67%, validated on a registration method between histology to in-vivo CT [18]. Based on our results, a 20% SUV\textsubscript{max} threshold with no margin expansion resulted in an average sensitivity of 71% and an average specificity of 69%. Differences in these results from our study may be attributed to registration error between histology and CT as intra-prostatic tissue is not well defined on CT. Our previously validated registration pipeline offers an unbiased approach to validation against the gold-standard whole mount histology with a measured target registration error [14]. Overall, there was very little change in the resulting mean sensitivities and specificities when we performed a cross validation experiment on the data; however the strength of this result is limited until external validation is performed on an unseen dataset. In the context of investigating parameters for PSMA-PET lesion segmentation with a histologic reference standard, we are the first group to perform such cross-validation in our analysis.

Our results should be interpreted in the context of the limitations of our study. First is the relatively small number of patients in our study; these segmentation recommendations require validation on a larger external cohort. The high uptake of \(^{18}\text{F}-\text{DCFPyL}\) in the bladder may be a confounder for mid-gland lesions located near the base; newer PET imaging agents such as PSMA-1007 are not excreted through the urine and might facilitate the use of PET for intra-prostatic delineation by reducing interference from signal from the bladder and urethra [20]. One other limitation is intra-patient heterogeneity in PSMA uptake, which has been demonstrated and may confound the intensity of signal displayed by some DILs [21]. Our experiments
were limited to PET data acquired on a PET/MRI scanner, which may differ from acquisition on a PET/CT machine, which is much more readily available for radiotherapy planning. In principle, the use of PET/CT should not affect the results demonstrated in this study (given an accurate segmentation of the prostate); however in practice these guidelines require validation on multi-scanner and/or multi-institutional data. This study only investigated DILs lying with the prostate mid-gland, excluding DILs in the apex and base; generalization of the results of this study to the whole gland involves the assumption that PSMA uptake, and its relationship to the presence of cancer, is the same in the apex and base as it is in the mid-gland. All analysis was performed in the in-vivo MR imaging space at the corresponding MRI voxel resolution and therefore would require PET images to be resampled accordingly in a clinical environment without a PET/MRI scanner. It remains to be seen how our method would perform if PET images were used without resampling. Lastly, our dataset consisted predominantly of patients with clinical staging ≤cT2; further testing against patients with higher grade clinical stage is needed.

3.5 Conclusions

This study used accurate co-registration of PSMA-PET/MRI and surgical histology to determine optimal SUV thresholds and margin expansions for high sensitivity, supporting focal therapy/boosting, and high specificity, supporting guided biopsy. Focal therapy/boosting recommendations (SUV ≈ 70% of SUV$_{\text{max}}$ with an ≈ 8mm margin) produced an average lesion sensitivity of 95.0 ± 7.8% and an average specificity of 76.4 ± 14.8%. Guided biopsy recommendations (SUV ≈ 80% of SUV$_{\text{max}}$ with a ≈ 5mm margin) produced an average lesion sensitivity of 65.1 ± 28.4% and an average specificity of 95.1 ± 5.5%. Importantly, no DIL was missed by our threshold and margin expansions in the setting of guided biopsy and previous literature indicated all lesions would be hit by typical guided biopsy protocols including multiple attempts per lesion. The thresholds and margin expansions determined by this study
can be used in a larger validation study supporting clinical translation.
References


Chapter 4

$^{18}$F-DCFPyL PSMA-PET for lesion-directed prostatic high-dose-rate brachytherapy dose escalation

4.1 Introduction

PCa recurrence most often originates from the largest and/or highest-grade site of disease [1–3]. These site(s) are termed DILs. To improve patient outcomes and reduce disease recurrence, treatment methods for PCa that target DILs have begun to emerge. One method is focal dose escalation during radiation therapy. Many studies have confirmed the feasibility of radiation therapy to escalate dose to identified DILs [4–6]. Due to its highly focal nature, brachytherapy techniques such as HDR-BT may be particularly well-suited to escalate doses to DILs while abiding by dose constraints to surrounding organs at risk [4–8].

Previous studies investigating HDR-BT for focal dose targeting have used mp-MRI to localize the DIL [9–11]. While mp-MRI has been confirmed to be a useful tool in locating the site(s) of disease [12], there has been recent evidence that mp-MRI is not able to delineate all of the cancer, and that interobserver variability could be a confounding issue [13]. An
alternative imaging method that has shown promise for prostatic DIL localization is PSMA-PET [14]. Recent work has demonstrated that use of a Gallium-labeled PSMA based radiotracer led to improved sensitivity and specificity for lesion detection confirmed by histopathology, compared to alternative imaging techniques [15]. More recently, a prospective trial reported higher specificities and positive predictive values using PSMA-PET for locating tumour foci within the prostate, compared to mp-MRI [16]. A promising radiotracer for PSMA-PET imaging of the DIL is (2-(3-carboxy-5-[(6-18F-fluoro-pyridine-3-carbonyl)-amino]-pentylureido)-pentanedioic acid) or more commonly referred to as [18F]DCFPyL, [17] which has demonstrated higher spatial resolution than previously-developed Gallium based agents and has the advantage of efficient cyclotron-based production [15].

Challenging the use of PSMA-PET imaging in the clinic is the lack of established guidelines regarding SUV thresholds for delineating lesions. Previous studies have attempted to define standardized guidelines through voxel- and region-based analyses. However, they are often limited by their reference standards (e.g. targeted biopsy cores [18]) and consensus mapping approaches [19, 20]. Whole-mount histology registered to the in-vivo imaging volume serves as the gold standard. However, most studies that have attempted to register histology to in-vivo imaging are limited by the uncertainty of their registration error [21]. Previously, a semi-automated registration algorithm was developed by Gibson et al. to map whole-mount transverse prostate histology to in-vivo MRI with a quantified target registration error [13]. Using this technique, we obtained images from patients using a hybrid PET/MRI scanner and analyzed them retrospectively using an optimal threshold and margin expansion algorithm for lesion delineation developed by Alfano et al. [22].

In this paper, we investigated the use of this PSMA-PET thresholding algorithm to design focal dose-escalated HDR-BT treatment plans. Through the confirmation of registered digital histology, we aim to answer the following question: does targeting PSMA-PET defined contours increase the dose to all histologically-defined ground truth cancer while meeting dose constraints to nearby critical structures, in comparison to the clinically standard whole-gland
treatment planning method? The use of co-registered whole-mount histology with clinical HDR-BT plans to answer this question is unique to our study. We hypothesize that targeting dose to PSMA-PET defined DIL contours will result in a significant increase in dose to all histologically defined cancer within the prostate, compared to standard-of-care whole-gland treatment plans.

4.2 Methods and materials

4.2.1 PSMA-PET/Histology Patient Population

Under an approved Health Sciences Research Ethics Board study, patients were recruited as part of a multimodality, pre-operative PCa imaging study (n=20). Patients were eligible for this study if they had biopsy-confirmed clinical stage T1c or T2 PCa and had elected to undergo radical prostatectomy for their primary treatment. Exclusion criteria included: prior therapy for PCa, use of 5-alpha reductase inhibitors within 6 months of study start, inability to comply with pre-operative imaging, allergic reactions to contrast agents, contraindications to MRI, and insufficient renal function (eGFR <60mL/min/1.73m²). MR imaging was performed at 3T on a Siemens Biograph mMR (Erlangen, Germany) hybrid PET/MRI scanner. PET imaging was performed using the [18F]DCFPyL molecular probe targeting PSMA, manufactured with Health Canada Quality (Chemistry and Manufacturing) Certification and prepared and supplied by the Centre for Probe Development and Commercialization (CPDC; Hamilton, Ontario, Canada). For further details regarding MRI and PET imaging please refer to Appendix B.

A pathology assistant trained in PCa morphology annotated (i.e. graded and segmented all intra-prostatic lesions) all histology sections, under the supervision of a genitourinary pathologist. All histology annotations were given a Gleason score by the observing clinician, and all were confirmed (or corrected as necessary) by the pathologist [23]. For the purpose of this study, only annotations that had a Gleason score of 7 or greater were evaluated. Differentia-
tion of PCa by a Gleason score 7 (3+4) or greater and Gleason score 6 or less is concordant with the Kasivisvanathan et al. definition of clinically significant and clinically insignificant disease [24]. More details of histological examination can be found in Appendix B.

4.2.2 Histology to PET/MRI Registration

All annotations of cancerous tissue from the mid-gland of the prostate were registered to the in-vivo PET/MRI using a registration procedure that was previously developed by Gibson et al. [13]. The registration algorithm aligned the histology to the 3D ex-vivo MRI of the prostate post-surgery using external and internal fiducial markers via an affine transformation. These fiducial markers were soaked in a solution of gadolinium and blue dye to be identifiable on both digitized histology and MRI. The ex-vivo MRI was then registered to the in-vivo MRI scan using user identified homologous anatomical landmarks (e.g. calcifications, benign prostatic hyperplasia nodules) via a thin-plate spline. The mean target registration error between histology and in-vivo MRI was reported to be 1.7 mm [13]. The contemporaneous acquisition of both modalities (PET and MRI) produced an intrinsic co-registration between PET and MRI and thus we had a registration scheme between histology and in-vivo PSMA-PET.

From the population of 20 patients imaged, we analyzed a cohort of 12 patients who had a DIL in the mid-gland with corresponding focus of high uptake on PET imaging within the prostate. This is due to our registration framework being highly optimized and robust for registration of axial sections; the remaining 8 patients had cancer that lied in the apex and/or base. At our institution, the apex and base were sectioned sagittally to prevent under-sampling in these regions with less tissue; thus, we were unable to register these sections to the in-vivo imaging. Two additional cases from this analysis were excluded: one due to an abnormally small sized gland preventing a suitable registration to any of the HDR-BT patients in our dataset, and one due to a lesion enveloping over half the gland, preventing dose constraints to the urethra to be met, and therefore this patient is not a suitable candidate for prostatic HDR-BT focal dose escalation. Thus, a total of 10 patients were able to be used for this analysis.
4.2.3 Brachytherapy Population

Between July 2015 and December 2016, 30 patients received a single-fraction TRUS guided HDR-BT procedure at our institution. A Profocus 2202 ultrasound machine and an 8848 bi-planar probe (BK Medical, Boston, MA) operating at 9 MHz were used to capture all TRUS images. Using a custom manual stepper installed with Vitesse 2.5 (Varian Medical Systems, Palo Alto, CA) intra-operative software, axial slices of ultrasound images with 5 mm spacing were acquired as part of the HDR-BT planning.

4.2.4 PET/MRI-TRUS Registration

Each PET/MR image and corresponding histology images were deformably mapped to two different HDR-BT plans in the 30-patient dataset by matching prostates that were close in terms of anatomical similarity (size and shape) (see Figure 1, red arrows). Of the 30 HDR-BT patients, 9 cases had prostate anatomy similar to our PET/MRI cohort and were selected as suitable matches. This registration procedure yielded the mapping of PSMA-PET DILs and the corresponding digital histology annotation to the intraprocedural TRUS imaging space producing a total of 20 treatment plans (10 patients each mapped to two different treatment plans, where each plan is a unique combination of catheter insertions and dose distributions). PSMA-PET DILs from different patients may be mapped to the same HDR-BT patient in order to meet size and shape constraints for the matching. Further details of registration are explained in Appendix B.

4.2.5 Standard Brachytherapy Planning

Standard clinical HDR-BT treatment plans were designed using BrachyVision 13.6 planning software (Varian Medical Systems, Palo Alto, CA) with the intent to deliver the prescription dose to the entire gland as equally as possible while abiding by dose constraints to organs at risk (Figure 1, third column). Treatment plan dose constraints were based on the techniques
used by Morton et al. [25] and guidelines outlined by the American Brachytherapy Society, as described by Yamada et al. [26]. Details of the treatment planning dose constraints and objectives are described in Table 1. All standard treatment plans were created using the volume optimization application that is built into BrachyVision.

4.2.6 Focal Targeting Treatment Planning

All HDR-BT PSMA-PET targeted treatment plans were designed using BrachyVision 13.6. Each plan was generated by using the volume optimization application that is directly built into BrachyVision. Since these plans were designed with the goal to deliver a focally targeted dose to the PSMA-PET lesion, an adjusted set of dose constraints and dose goals were used and are listed in Table 1. The target dose was 20.3 Gray (Gy) to 98% of the PSMA-PET segmentation (Figure 1, right column) which was generated using a threshold of 81% of the SUV$_{\text{max}}$, along with a margin expansion of 5.2 mm. The segmentation parameters, yielding $\geq 95\%$ specificity with maximal sensitivity on a voxel-by-voxel basis, were determined in previous work by Alfano et al. to be a threshold of 81% of the maximum standardized uptake value (SUV$_{\text{max}}$) in the DIL, with a 5.15 mm margin [22]. The median [IQR] Dice-similarity coefficient of the segmentations with respect to the DIL on each histologic plane was 0.77 [0.62–0.93] suggesting moderate to substantial overlap. The median [IQR] signed volume difference was -18.7 [-74.7–37.2] mm$^3$, meaning segmentations were found to encompass less prostate tissue than the DIL defined on histology. This approach was found to have 100% sensitivity in targeting DILs within the prostate on a per lesion basis. More information on the qualitative performance of the DIL targeting scheme can be found in Figures 2 and 3 in Alfano et al. [22]. All PSMA-PET segmentations were masked to the volume of the prostate which was defined by a graduate student on the MRI for each patient. As our goal was to model dose escalation using clinically deliverable plans, we did not modify the original catheter distribution or add additional brachytherapy catheters to aid in the dose escalation; rather, we restricted our optimization to the existing catheter distribution.
Organ Dose metric | Standard planning goal [% prescription dose] | PSMA-targeted planning goal [% prescription dose]
---|---|---
Prostate | $V_{90}$ | $\geq 95$ | $\geq 95$
Prostate | $V_{100}$ | $\geq 90$ | $\geq 90$
Prostate | $V_{150}$ | $\leq 35$ | $\leq 38$
Prostate | $V_{200}$ | $\leq 11$ | $\leq 14$
Urethra | $D_{10}$ | $\leq 118$ | $\leq 120$
Rectum | $V_{0.5cc}$ | $\leq 80$ | $\leq 80$
PSMA-DIL | $V_{98}$ | N/A | 135

Table 4.1: Dose constraints followed during the standard and PSMA targeted HDR-BT treatment planning. $V_{xx\%}$ is the volume percentage of the structure that received $xx\%$ dose of the prescription dose, $D_{xx\%}$ is the prescription dose percentage that $xx\%$ of the structure received, and $D_{xxcc}$ is the maximum prescription dose percentage permitted to $xx$ cubic centimeters of volume in that structure.

### 4.2.7 Dose Analysis

To determine if there was any effect of targeted focal dose escalation using PSMA-PET, we calculated the dose the annotated digital histology cancer would have received within the standard HDR-BT treatment plans, and the dose to the corresponding PSMA-PET targeted treatment plans (Figure 1, output of columns 3 and 4). Specifically, we calculated the minimum dose being delivered to 98% and 90% of the cancer, as well as the mean dose that would have been delivered to each histology annotation.

### 4.2.8 Statistics

Prism 5.04 (GraphPad Software, San Diego, CA) was used to perform all statistical analysis. A threshold of $p<0.05$ was used to deem results significant. Data sets were all subject to the Shapiro-Wilk test to determine distribution normality. If the data sets were deemed to be normally distributed, a paired t-test was used to compare means, while if one of the data sets that were being compared were not normally distributed, the medians were compared using a Wilcoxon matched-pairs signed rank test.
Figure 4.1: Flow diagram illustrating the PSMA-PET and corresponding histology lesions being registered into the simulated HDR-BT treatment plans and corresponding PSMA-PET targeted treatment plans. The targeted plans were designed to target the PSMA-PET lesions, with the optimizer blinded to the location(s) of the histologic lesion(s). The dose to the histology was calculated for the standard plans and targeted plans.

### 4.3 Results

The histologically defined cancer would have received a significantly higher \( D_{98} \) within PSMA-PET targeted treatment plans in comparison to the plans optimized for standard whole-gland treatment, with a median [IQR] \( D_{98} \) of 16.5 [15.0–19.0] Gy compared to 15.2 [13.8–16.4] Gy in the standard plans (\( p \ll 0.05 \)). The median \( D_{90} \) to the histologic cancer within the PSMA-PET targeted treatment plans was 19.0 [16.5-21.0] Gy compared to 16.4 [15.0-17.4] Gy for the standard plans (\( p \ll 0.05 \)).

In terms of mean dose, the histologically defined cancer received a significantly higher dose within the PSMA-PET targeted treatment plans. The PSMA-PET targeted plans would have
delivered a median [IQR] mean dose of 25.6 [23.7–32.9] Gy, and the standard plans delivered a median mean dose of 21.4 [20.2–22.6] Gy (p ≪ 0.05). The percentage of focally targeted plans that deliver a given D$_{98}$ dose to the histology is presented in Figure 3. For example, Figure 3 demonstrated that the percent of plans that delivered a D$_{98}$ of 20 Gy was 0% for standard plans and 20% for PET-targeted plans. Additionally, it was found the percentage curves were significantly different with a log-rank test (p ≪ 0.05). In terms of dose to the surrounding organs at risk, we did not exceed dose constraints when targeting the DILs. Specifically, we achieved a median prostate V150 of 31.4%, a median prostate V200 of 11.7% , and a urethra D10 of 116.2% . Our constraints for the prostate V150 was 38%, 14% for the prostate V200, and 120% for the urethra D10.

![Box-and-whisker plot of the (left) D$_{98}$, (middle) D$_{90}$, and (right) mean doses delivered to the digital histology annotations within PSMA-PET targeted treatment plans and standard treatment plans. The whiskers represent the 10-90th percentiles.](image)

**4.4 Discussion**

We found that by focally escalating the dose to the PSMA-PET segmentations, we were able to achieve a significantly higher D$_{98}$, D$_{90}$ and mean dose to the corresponding histologically-defined cancer in comparison to the standard whole-gland designed treatment plans. Additionally, we found that by targeting the PET lesions we were able to significantly increase the percentage of plans that would have received a higher D$_{98}$ dose to histologic disease (p ≪ 0.05).
Figure 4.3: Line graph displaying the percentage of treatment plans that delivered a given $D_{98}$ to the histology cancer. PET-targeted treatment plans are colored in blue and standard treatment plans are colored in magenta.

In the context of focal lesion dose escalation, much of the previous literature has focused on the use of mp-MRI. To the best of our knowledge, there have not been any previous studies investigating PSMA-PET as the imaging modality for lesion-directed dose escalation within the prostate, validated with histology. Additionally, outside of a study by Smith et al. (which used mp-MRI as the lesion imaging modality) [11], we are not aware of other studies that
have used dosimetry of prostatic digital histology to confirm the effectiveness of lesion dose escalation.

In 2016 Gomez et al. [9] performed a phase II clinical study that dose-escalated mp-MRI DILs using HDR-BT. Using a rigid registration method from mp-MRI to TRUS, Gomez successfully escalated dose to targeted lesions to a $D_{98}$ of 18.7 Gy with acceptable tolerance and toxicity profiles. More recently, Hrinivich et al. demonstrated that replanning clinical stereotactic ablative radiotherapy plans by escalating dose towards suspicious PSMA-PET hot spots for oligometastatic PCa resulted in overall increases to the maximum dose to the PTV and decreases to the maximum dose to the OAR [27]. While we used original previously achieved catheter distributions from past treatment plans to target the lesion, additional catheters could be inserted to attempt to further dose escalate the PSMA-PET DILs. In 2019, Wang performed a study that investigated the effect of adding one additional needle on the dose to mp-MRI lesions. In that study, they found that mp-MRI DILs would receive a V150 dose of $46.4\pm28.6\%$ through standard whole-gland planning, while DIL-targeting treatment with the original catheters would achieve a mp-MRI V150 dose of $87.0\pm15.5\%$, and targeted plans with an additional targeted needle added would achieve a V150 of $93.0\pm14.1\%$. The work performed by Wang demonstrated that the majority of targeted dose can be achieved with the original catheters and that the additional catheter only caused a 6% V150 increase to the mp-MRI DIL [10].

These papers have all demonstrated the feasibility of lesion dose escalation, and the localization capabilities of PSMA-PET. Our study adds to these results by not only demonstrating the feasibility of PSMA-PET lesion-directed focal dose escalation, but also confirming the effectiveness of doing so by calculating the dose that would have been delivered to the ground truth histologically defined cancer. Our PSMA-PET lesions were created using a thresholding guideline of $81\%$ SUV$_{\text{max}}$ and a 5.15 mm margin without any user interpretation, and histologic clinically significant cancer was defined in concordance with Kasivivanthan et al. 2018, as all cancer with a Gleason Score of 7 (3+4) or greater [24].
Since patients undergo salvage prostatectomy after HDR-BT very infrequently, it is rare to obtain post-prostatectomy histology from HDR-BT patients. In the rare cases where salvage prostatectomy is done after HDR-BT, a substantial amount of time lapses between HDR-BT and prostatectomy. Therefore, it is impractical to obtain a PSMA PET/MRI scan, perform HDR-BT, perform prostatectomy, and obtain prostatectomy histology all on the same patient, with a short time frame between the PET/MRI scan and obtaining histology. This short time frame is important since the histology must serve as a ground truth for the PSMA PET signal. Therefore, to conduct this study with actual HDR-BT catheter implants that capture real spatial inaccuracies and catheter deflections that occur in the clinical setting, we registered prostates on PET/MRI scans taken shortly before surgery from a prostatectomy cohort, to prostates on TRUS containing real HDR-BT catheter implants in a brachytherapy cohort. This approach allowed for testing of PSMA-PET-based lesion targeting on realistic HDR-BT catheter distributions with validation against whole-mount prostatectomy histology, which has not been previously reported in the literature.

Several other PET DIL contouring methods have been reported in the literature. Spohn et al. reported on a study aiming to segment the DIL with the [18F]PSMA-1007 PET probe using semi-automatic methods and comparing it to manual user delineation [28]. The authors found that a threshold of 20% of SUV$_{\text{max}}$ achieved a median sensitivity of and specificity of 93% and 96%, respectively. The approach the study took is concordant with the approach taken in Alfano et al. with two major differences. First, the registration protocol is based on a manual alignment of in-vivo pre-treatment CT and multi-parametric MRI, which may induce some inherent registration error in mapping histology to PET. The study reported by Alfano et al. used co-registration of histology to in-vivo PET/MRI, eliminating the intermediate manual registration component. Second, the authors in this paper interrogated the more recently-developed [18F]PSMA-1007 radiotracer, compared to [18F]DCF-PyL radiotracer investigated in Alfano et al. [18F]PSMA-1007 has many advantages, including higher signal-to-noise ratio and a lack of urinary excretion of the tracer, meaning that there is no buildup in the bladder. This was one
of the limitations of the [18F]DCF-PyL radiotracer used in this study; signal from the bladder may leak into the prostate and may confound uptake in tumour and/or healthy tissue close to the base. Thus, we hypothesize that a focal boosting study against PET-DILs defined from PSMA-1007 would allow for more accurate segmentations to be generated, potentially further improving dose delivery to underlying disease. Fassbender et al. examined a gastrin-releasing peptide receptor-based radiotracer (GRPR) in comparison to a PSMA based radiotracer validated on co-registered histology [29]. This is different than the study investigated in Alfano et al. and our study which was only concerned with examining the expression of radio labeled PSMA. Overall, the study indicates that certain tumours have greater expression of GRPR compared to PSMA and vice-versa. Thus, primary PCa may be able to be more accurately defined if information regarding the expression of GRPR and PSMA was provided at the time of imaging. This may be a precursor step in creating DIL contours for focal dose escalation by informing physicians on the type of radiotracer that should be administered and may provide added benefit in terms of increased dose to underlying disease as a result.

Figure 4 illustrates an ideal targeting candidate with a localized unifocal lesion within the prostate that is distal to the urethra, allowing for optimization to occur under the allocated dose constraints. This was not the case for all patients; some exhibited extensive disease that encapsulated nearly half the prostate gland while being proximal to the urethra. These patients, although included in our data for analysis, may not be optimal focal targeting candidates given the extent of their disease and may be better candidates for radical whole gland therapy such as prostatectomy.

The results presented in this paper need to be interpreted in context of the limitations of our study. First, the TRUS and MRI/PET registration error was not evaluated. The purpose of the registration was to place both the PSMA-PET DILs and corresponding pathologically defined cancer within HDR-BT treatment plans, while maintaining the spatial relationship between the DIL and pathologically confirm disease. Previous studies have published the error between TRUS and MRI to between 2.1 mm and 3.5 mm [30, 31]. It should be noted that those errors
were for intra-patient registration. However, we performed inter-patient registration. We performed inter-patient registration to simulate realistic spatial distribution of HDR-BT catheters surrounding the PSMA-PET defined DILs. The critical registration error of our work was the error between the histology and MRI, which was meticulously measured and reported by our group [13] to be less than 2 mm. Also due to our MRI-histology registration algorithm’s optimization for registration of transverse, mid-gland histology sections, our data set excluded patients having a DIL in sagittally-sectioned apex and base regions. Lesions in the prostate located near the base may be confounded by the high tracer uptake in the bladder. Newer PET imaging agents, such as [18F]PSMA-1007, are not excreted through the urine and might mitigate this interference [32]. Our experimentation was also limited to one segmentation algorithm designed to encompass little to no healthy tissue while maximizing the amount of cancerous tissue; future work in this area may revolve around optimizing the thresholding and margin expansion criteria to create contours that cover more histologic disease and achieve higher targeted doses while maintaining low dose to organs at risk. Fourth, this study was a single-institution retrospective analysis on a small cohort of men with lesions predominantly distal to the urethra; our findings need to be further validated on a larger, multi-institutional data cohort to account for patient and catheter implant variability.

4.5 Conclusions

This study is the first to confirm the potential effectiveness of computer generated PSMA-PET lesion directed focal dose escalation using HDR-BT for PCa through the utilization of co-registered digital histology. Our findings suggest that focally escalating dose to PSMA-PET defined mid-gland lesions will lead to significantly higher dose to histologically defined disease.
References


nant intraprostatic lesion using multiparametric magnetic resonance imaging including spectroscopy: Results of a prospective study,” *Brachytherapy*, vol. 15, no. 6, pp. 746–751, 2016.


Chapter 5

Conclusions and future work

5.1 Project summary and conclusions

The work presented in this thesis aims to make strides towards augmenting mp-MRI with the goal of improving cancer classification through the development of a radiomics machine learning model and advances in knowledge related to DIL segmentation recommendations on PSMA-PET. By leveraging an in-house, high-fidelity registration between post-prostatectomy histology and in-vivo mp-MRI [1], we investigated two primary solutions for improving mp-MRI interpretability: (1) the application of artificial intelligence methods and (2) the use of additional imaging in the form of PSMA-PET. Artificial intelligence methods were developed in the form of a radiomics based machine learning model for classifying lesions on mp-MRI, which was validated based on imaging from 60 patients covering all Gleason grade groups. A subset of these patients were imaged on a hybrid PET/MRI machine and thus with accurate fusion of histology to PET imaging, we investigated clinical threshold and margin parameters as suggestions for delineating the DIL. For this experiment, we optimized thresholds and margins on a subset of 12 patients for two different clinical applications: (1) focal therapy (highly sensitive procedure) and (2) guided biopsy (highly specific procedure). Furthermore, we examined the effects of applying the computer generated DIL contours on PSMA-PET for intervention
planning, namely focal dose escalation in HDR-BT treatment plans. The objective was to measure any changes in focal boost plans compared to the current clinical standard whole-gland plans. Thus, this thesis contributes to new technology and clinical knowledge that is validated on a state-of-the-art high-fidelity reference standard of transverse histology to in-vivo imaging that is unique to our group. The accomplishment of all the objectives (Section 1.8.1) led to the following answers to the central hypotheses (Section 1.8.2).

5.1.1 Achievement of research objectives and testing hypotheses

Objective 1: To develop and test a machine learning model on radiomic features derived from an un-biased sampling of clinically significant malignant and benign lesions on mp-MRI validated with co-registered histology.

In Chapter 2, a radiomics based machine learning model was developed to classify PCa vs. non-PCa regions on mp-MRI validated using accurately co-registered histopathology [1]. Previous studies in this space had validated their machine learning models based on consensus mapping [2–5], TRUS biopsy [6–10], and MRI-guided biopsy [11, 12]. Each of these methods has its own inherent limitations. Consensus mapping approaches are biased towards the accurate localization of mutually identified tumours but not for confounding and/or healthy tissue. TRUS biopsy based approaches have previously shown widely varying detection rates (23%-53%) and thus are prone to missing cancers in the gland [13]. MRI-guided biopsy based approaches have been shown to identify significantly less low-grade disease compared to high-grade [14]; data sets with a reference standard in this space would be biased towards higher-grade tumours that are identifiable by radiologists on MR imaging. To overcome these biases, our lab used accurately co-registered, transverse radical prostatectomy histology and in-vivo mp-MRI with a measured target registration error, mapping pathology annotations of cancer to the in-vivo imaging space [1]. Using regions of cancer and non-cancer derived from pathology, we found that a 5-feature NAIVEB classifier was able to produce a LOPO cross-validation AUC of 0.80. Thus, we have confirmed the first hypothesis outlined in Chapter 1.8.1. Hypoth-
esis 1: Through the work presented in Chapter 2, a 5-feature NAIVEB classifier can classify PCa vs. non-PCa on mp-MRI validated on co-registered histology with an AUC \( \geq 0.80 \).

Objective 2: To investigate optimal combinations of threshold of SUV\(\text{max}\) and margin expansion for DIL segmentation on PSMA-PET imaging for focal therapy/boosting (sensitivity \( \geq 95\% \) with maximal specificity) and guided biopsy (specificity \( \geq 95\% \) with maximal sensitivity).

In Chapter 3, we investigated recommendations for the segmentation of the DIL on PSMA-PET/MRI validated using accurately co-registered histopathology. Recently, there has been plenty of evidence indicating that PET radiotracers targeting PSMA are able to identify primary PCa with good concordance to relevant histopathology [15–17]. However, recommendations for boundary delineations of the DIL on PSMA-PET imaging remains poorly described in the literature. Previous work attempted to create recommendations using both SUV\(\text{max}\) cutoffs [15] and thresholds [18] for defining the DIL. However, these studies were primarily limited by their reference standard being the correlation of underlying histology to CT imaging, a task that is difficult due to the lack of soft tissue contrast present for accurate registration. A subset of patients in our data set underwent pre-operative imaging on a PET/MRI scanner, thus we had a registration pipeline of accurately fused in-vivo PSMA-PET imaging to transverse radical prostatectomy specimens annotated by a pathologist [1]. To delineate the underlying DIL, we proposed segmentation recommendations consisting of a combination of threshold of SUV\(\text{max}\) and margin expansion. This allowed us to generate optimization functions to produce thresholds and margin expansions for different clinical applications, namely: (1) focal therapy and (2) guided biopsy. We proposed criteria for focal therapy delineation to maintain a sensitivity \( \geq 95\% \) while maximizing the specificity, and criteria for guided biopsy to maintain a specificity \( \geq 95\% \) while maximizing the sensitivity on a per-voxel basis. This allowed us to produce focal therapy targeted thresholds of SUV\(\text{max}\) \( \approx 70\% \) with an \( \approx 8\)mm margin and guided biopsy targeted thresholds of SUV \( \approx 80\% \) of SUV\(\text{max}\) with a \( \approx 5\)mm margin. Thus, we have confirmed the second hypothesis outlined in Chapter 1.8.1. Hypothesis 2: Through the work
presented in Chapter 3, DILs on PSMA-PET imaging can be segmented using a threshold of $SUV_{\text{max}}$ and radial margin expansion with maximal sensitivity and specificity for focal therapy and guided biopsy applications.

**Objective 3:** To investigate the change in dosimetry in HDR-BT to underlying PCa in focally boosted treatment plans using generated PSMA-PET segmentations compared to whole-gland treatment plans.

In Chapter 4, 3D contours of the DIL from Chapter 3 were used to perform focal boosting on HDR-BT plans to compare against the standard whole gland clinical plans, validated based on the underlying co-registered histology. Currently in the clinic, it is common that PCa patients undergoing HDR-BT are treated with a prescription dose to be delivered to the whole gland. However, it has been shown that the DIL is most commonly the site of recurrence [19, 20], suggesting preferential targeting of dose to these areas. HDR-BT has previously shown to be a feasible treatment modality for boosting doses towards regions within the prostate while maintaining dose constraints to other nearby organs at risk [21–23]. Many studies have previously investigated the use of mp-MRI for delineating DILs and boosting dose towards these contours [22, 24, 25]. However, the use of computer-generated PSMA-PET based contours for dose targeting, validated against co-registered histology, has not been explored in the literature. Using the same patient cohort as identified in Chapter 3, we proposed a simulation study using real clinical implants to boost seed dwell times towards computer-generated PMSA-PET defined contours and measured the dose delivered to the underlying histology compared to standard-of-care whole-gland treatment plans. The results of Chapter 4 concluded that targeting dose to these PSMA-PET contours resulted in a significant increase in dose to the underlying high-grade PCa. Thus, we have confirmed the third hypothesis outlined in Chapter 1.8.1.

**Hypothesis 3:** Through the work presented in Chapter 4, inverse-planning optimization of dwell times in HDR-BT targeted at segmentations of DILs on PSMA-PET resulted in a significantly increased $D_{98}$ to all underlying co-registered high-grade histology.
5.1.2 Advances in knowledge and technology

A radiomics based machine learning model to classify PCa on multi-parametric MRI validated on co-registered histopathology provided considerations for potential biases that may confound previously developed CAD systems

Chapter 2 presented a pipeline for developing an unbiased CAD model for classifying PCa on multi-parametric MRI. This study was the first to perform model validation against co-registered prostatectomy whole-slide images with a measured target registration error [1]. Furthermore, evidence in recent literature has indicated that radiomic texture features may be correlated to volume based characteristics of lesions on imaging [26, 27] which have not been considered in prior PCa classification models for a similar task [28–32]. We proposed an algorithm that rigourously sampled the pathology defined healthy tissue and ensured there was no differences in mean area between ROIs for both cancer and non-cancer. As mean area imbalance grows, we found there to be a trend of increasing AUC, suggesting that the underlying first and second order radiomic features may not be invariant to mean volume differences between classes. Prior to this observation, only feature changes when volume differences were applied had been studied, it was unknown how these effects translated to resulting machine learning performance. This suggests that users developing CAD models need to be aware of lesion size when performing similar experiments. By transforming healthy tissue regions generated by using cancer tissue masks into area equivalent regular circles, we found that our system was invariant to differences in shape between classes. Prior to this finding, it was unknown whether the shape of the annotated regions may influence a classifier to produce a positively biased result. The results from Chapter 2 also suggest that a machine learning classifier trained on handcrafted radiomic texture features perform better in the CG compared to the PZ, which was previously suggested by another study [33]. Furthermore, given that a subset of patients in our data set were imaged on a different MRI scanner, Chapter 2 also presented findings that showed a radiomics machine learning model performs worse when trained on data from one scanner and tested on another as opposed to being subjected to data from both devices. Therefore,
This thesis contributes to the technological advancement of a machine learning system for PCa classification on mp-MRI validated against co-registered mid-gland, prostatectomy histology. This thesis also contributes to the knowledge of the effects of underlying experimental biases, namely: selection, size, shape, location and scanner origin providing considerations for future research studies in this space.

A set of DIL segmentation recommendations for PSMA-PET imaging validated on co-registered histopathology was presented for different clinical applications in PCa management.

Chapter 3 presented a pipeline for investigating thresholds of $SUV_{max}$ and radial margin expansion for segmenting the DIL on PSMA-PET imaging. Our proposed experiment generated segmentations ranging from 1-100% $SUV_{max}$ in 1% intervals with margins from 0-20 mm in ≈1 mm steps at each interval, creating segmentations with measured sensitivity and specificity for each. We were the first group in this space to propose segmentation recommendations based on optimized voxel-wise sensitivity and specificity for DIL classification in different clinical problems. We also performed a cross-validation study to assess the generalizability of these parameters when exposed to unseen data and found similar resulting sensitivity and specificity when compared to results measured on the whole training cohort. Previously, no other study in this space assessed segmentation and/or margin expansion segmentation recommendations through cross-validation. Future studies should consider applying cross-validation techniques to their studies to make inference on generalizability to new, unseen data. Also, the work in Chapter 3 compared the voxel-wise sensitivity and specificity of our approaches to those made by a radiation oncology resident and found an improved per-lesion sensitivity, supporting previous findings that these methods could be used to improve lesion detection on PSMA-PET [18]. Furthermore, we assessed the generated segmentations made for guided-biopsy applications against co-regisered histopathology and found that 11/12 patients were able to have their primary lesion biopsied in ≤2 attempts if targeted using our methodology [34]. Practically, in clinical scenarios where patients present with rising PSA and negative TRUS biopsy and mp-
MRI results, PMSA-PET may then be used in combination with these parameters to target suspicious regions for biopsy. Therefore, this thesis contributes to the advancement in knowledge of DIL segmentation strategies on PSMA-PET imaging for focal therapy and guided biopsy applications. This thesis also provides evidence of clinical benefits such as primary lesion biopsy frequency when applying these segmentation techniques, suggesting that future studies put their contouring strategies in the context of current clinical problems.

**Boosting dose to DILs identified by computer-generated PSMA-PET segmentations for HDR-BT procedures results in an increased dose to underlying high-grade histology**

Chapter 4 presented a pipeline for analyzing the effect of boosting dose to segmentations derived in Chapter 3 versus contemporary whole-gland treatment plans based on dose delivered to underlying co-registered histopathology. We were the first group in this space to propose a framework of targeting computer-generated PSMA-PET segmentations (Chapter 3) versus standard whole-gland plans with dose evaluated on accurately co-registered histopathology [1]. Aside from one other study [25], we are also the first to evaluate dosimetry to prostatectomy, whole-slide images registered to the in-vivo imaging space [1]. Thus, this study contributes to the field twofold: (1) we demonstrated that boosting to computer generated PSMA-PET contours is feasible using realistic catheter implants and (2) this method was confirmed to be effective in boosting dose to underlying cancer. Furthermore, this boosting procedure was found to be feasible in terms of dose delivered to nearby critical structures, such as the urethra and rectum, by meeting all dose constraints. Dose delivered to underlying low-grade histology was found to be unchanged compared to whole-gland plans suggesting that this cancer subtype does not benefit nor suffer from focal boosting procedures. Therefore, this thesis contributes to the advancement in knowledge surrounding the effect of increased dose delivery to underlying high-grade histology as a result of boosting dose to computer-generated PSMA-PET segmentations versus whole-gland treatment plans.
5.2 Suggestions for future work

5.2.1 Investigating the histopathology of MR-invisible lesions

The work in this thesis from Chapter 2 aimed to leverage machine learning models for classifying disease validated from accurately co-registered ground truth histology. However, it has been found previously that some lesions are deemed “MR-invisible” with no visual distinction able to be made by radiologists, even after viewing in a second pass [35]. One interesting avenue of work, would be to group lesions that have low visibility on MR imaging and analyze the underlying morphological patterns of disease on histology that is present compared to MR-visible lesions. Figure 5.1 displays an example patient in our data set that presents with a low-grade Gleason score 6 lesion on the posterior side of the prostate with no visual cues of disease presence located on the corresponding apparent diffusion co-efficient map and T2W image. If there was a certain pattern or morphology identified that correlated to the level of visibility on MRI, it would be of value in the clinical setting for diagnosis; patients with a positive finding on biopsy could have their disease analyzed by pathology and the results of the study would determine if MR imaging is useful for the lesion in question. Previous work has analyzed the histopathological features of PCa and correlated them to mp-MRI findings [36–38] however these studies are limited by biases in their reference standard as previously outlined in section 5.1.2. A previous study assessed the correlation of histologic subtype validated against accurately co-registered histopathology; however this study was limited to only assessing T2W

![Figure 5.1: An apparent diffusion coefficient map, T2W image and corresponding pathology annotations made on histology for a patient with low-grade disease with limited visibility on imaging.](image-url)
signal intensity [39]. Currently there is a gap in knowledge regarding the correlation of these histopathologic subtypes to other pulse sequences in the prostate mp-MRI protocol (e.g. diffusion weighted imaging). One objective of this study could be:

1. To determine the relationship between pathology generated contours of different PCa subtypes on histology and corresponding regions on all in-vivo mp-MRI sequences.

Furthermore, the use of histopathologic subtypes to correlate with mp-MRI signal is limited by the inter-observer variability in histologic lesion interpretation by different pathologists. One interesting avenue of work would be leverage computer-vision techniques to classify pathology annotations of PCa on digitized histology as visible or invisible on accurately co-registered mp-MRI. This measure of visibility would consist of a binary value associated with a cancerous region on mp-MRI as determined by a radiologist. This would result in the following objective:

1. To develop a machine learning and/or deep learning model to classify pathology generated PCa contours on whole-mount digitized histology as visible or invisible on mp-MRI validated with an accurate registration framework between histology and in-vivo imaging.

Patients who have a lesion that is deemed to be “MR-invisible” may be suitable candidates for PSMA-PET imaging [40], where the work from Chapter 3 may be useful to assist physicians in making contours of the primary disease.

5.2.2 Determining optimal PSMA-PET DIL segmentation criteria for maximizing dose to underlying high-grade disease

The work from Chapter 4 proposed that boosting dose to computer-generated segmentations of DILs on PSMA-PET imaging produced increased dose to high-grade disease while maintaining dose constraints to organs at risk compared to whole-gland plans. Feasibility of PSMA-PET targeting has not been investigated in the literature previously, thus our experiment opted for a
highly specific segmentation identified in Chapter 3 similar to the criteria for guided biopsy. An interesting approach for future work would be to develop optimization criteria that produced the maximum amount of dose to underlying high-grade disease while maintaining all dose constraints. Furthermore, it would be of interest in determining how this boosting procedure compares to the previously developed methods of boosting based on contours defined in the mp-MRI space [22, 24, 25]. The objectives of this study could be broken down in the following form:

1. To determine a recommended threshold of $\text{SUV}_{\text{max}}$ and margin expansion that produces the highest dose to underlying high-grade disease while maintaining relevant dose constraints to nearby critical structures.

2. To compare the dose delivered to underlying high-grade disease from boosting doses towards segmentations defined in the previous objective to those defined on mp-MRI from the same patient.

The results of this study would provide physicians with recommendations for optimizing DIL segmentations on PSMA-PET for HDR-BT treatments. It would also inform whether there is any advantage to having PSMA-PET imaging performed for this procedure versus mp-MRI.
References


Appendix A

**In-vivo PET/MR Imaging**

In-vivo prostate PET/MRI was obtained within 6 weeks before surgery and at least 6 weeks after biopsy. Mean time from imaging to radical prostatectomy was $2.5 \pm 1.5$ weeks (median 2.5, interquartile range 0.5–4.5). Imaging was performed at 3T on a Siemens Biograph mMR (Erlangen, Germany) hybrid PET/MRI scanner using an abdominal phased array and saline-filled endorectal coil. Patients were asked to take 30mLs of milk of magnesia prior to each MRI examination. In addition to standard clinical mp-MRI, a 3D-T2 weighted sequence was constructed using Sampling Perfection with Application optimized Contrasts using different flip angle Evolution (T2-SPACE) protocol with a repetition time 1700 ms, echo time 101 ms, bandwidth 601 Hz/pixel, 2 averages, field of view 23 cm × 23 cm, slice thickness 1.0 mm, slice spacing 1.0 mm, 72 slices, and 135° flip angle, image matrix 320 × 320 pixels. MRI-based attenuation correction was performed using a 2-point Dixon acquisition with 3 segmented components: air, fat and water. PET imaging was performed using a molecular probe targeting the prostate-specific membrane antigen called $[18F]$ DCFPyL prepared and supplied by the Centre for Probe Development and Commercialization (CPDC) (Hamilton, Ontario) and was manufactured with Health Canada Quality (Chemistry and Manufacturing) Certification. The radiotracer was injected intravenously at a dose of 325 MBq and images were acquired approximately two hours post injection at St. Joseph’s Hospital, London, Canada. PET images were reconstructed using an ordered subset expectation maximization method with three iterations and 21 subsets, a 4 mm Gaussian filter, and attenuation and scatter corrections. Image
matrix size was $172 \times 172$ pixels with a zoom factor of two and the final reconstructed spatial resolution was $2.09 \times 2.09 \times 2.03$ mm$^3$. MRI acquisition spanned the full 45–60 minutes of the exam while PET imaging was acquired within the first 15 minutes. Voxels on PET images were quantified by standardized uptake value (SUV) and up-sampled to a $0.644 \times 0.644 \times 1$ mm resolution using nearest-neighbor interpolation to match the in-vivo 3D MRI.

**Histology**

After whole-mount paraffin embedding of the prostate specimen post-radical prostatectomy with apex and seminal vesicles removed, the mid-gland was sliced into 4.4 mm transverse sections. Using a microtome, 4 µm slices were cut from each section and stained with hematoxylin and eosin. All slides were digitized at 0.5 µm/pixel (24-bit color) on a Scan Scope GL (Leica Biosystems, Concord, CA) bright field scanning system. Histology sections were annotated (intra-prostatic lesions graded and segmented) by a pathology assistant trained in prostate cancer morphology, supervised by a genitourinary pathologist.

**Registration**

Histology slides were registered to the ex-vivo specimen using internal and external fiducials via an affine transformation. Ex-vivo to in-vivo MRI registration was performed using fiducials corresponding to user identified homologous anatomical landmarks (i.e. BPH nodules, calcifications etc.) via a thin plate spline deformable registration. The contemporaneous acquisition of PET and MRI provided us with intrinsic registration between the two modalities, thus producing a registration scheme between histology and in-vivo PET imaging. To account for error in histology to in-vivo PET registration, histologically defined cancer was radially dilated on the oblique plane by 2.5 mm to account for the target registration error (1.7 mm) and the PET reconstruction (2 mm) by addition in quadrature.
Appendix B

*In-vivo PET/MR Imaging*

MR imaging was performed at 3T on a Siemens Biograph mMR (Erlangen, Germany) hybrid PET/MRI scanner. Images were acquired using a saline-filled endorectal coil and an abdominal phased array. In addition to a standard mp-MRI protocol, a 3D T2-weighted sequence was acquired using an echo time of 101 ms, repetition time of 1700 ms, bandwidth of 601 Hz/pixel, 2 averages, field of view of 23 cm × 23 cm, in-plane resolution of 0.644 × 0.644 × 1 mm, image matrix 320 × 320 pixels, slice spacing of 1.0 mm, slice thickness of 1.0 mm, 72 slices, and a 135° flip angle. MRI-based attenuation correction was performed using a 2-point Dixon acquisition with 3 segmented components: air, fat and water. PET imaging was performed using the [18F]DCFPyL molecular probe targeting PSMA, manufactured with Health Canada Quality (Chemistry and Manufacturing) Certification and prepared and supplied by the Centre for Probe Development and Commercialization (CPDC; Hamilton, Ontario, Canada). [18F]DCFPyL was injected intravenously at a dose of 325 MBq and image acquisition was performed approximately two hours post injection. PET images were reconstructed using an ordered subset expectation maximization method (OSEM) with 21 subsets and three iterations with a zoom factor of two, a 4 mm Gaussian filter, and attenuation and scatter corrections. Final reconstructed spatial resolution was 2.09 × 2.09 × 2.03 mm³ with an image matrix size of 172 × 172 pixels. MRI acquisition spanned the full 45-60 minutes of the exam with the 3D T2-weighted acquisition acquired subsequent to the 2D-T2 weighted imaging, while PET imaging was acquired for the first 15 minutes simultaneously. Voxels on PET images were
quantified by SUV. PET volumes were up-sampled to a $0.644 \times 0.644 \times 1$ mm resolution using nearest-neighbor interpolation to match the in-vivo 3D MRI.

**Histology Examination**

The radical prostatectomy specimens had the apex, base and seminal vesicles removed before whole-mount paraaffin embedding; these regions were sectioned sagittally and co-registered to in vivo imaging. The remaining mid-gland tissue was subsequently sliced in 4.4 mm transverse sections. Using a microtome, we cut 4 μm slices from each section and stained them with hematoxylin and eosin. We digitized all slides at 0.5 μm/pixel (24-bit color) on a ScanScope GL (Leica Biosystems, Concord, CA) bright field scanning system.

**PET/MRI-TRUS Registration**

The registration between PET/MRI was performed in a two-step process. The first step involved an iterative closest point transformation to rigidly align the prostate defined on PET/MRI to the prostate defined on TRUS. The second step was to deformably register the surface of the prostate defined on PET/MRI to the surface of the prostate defined on TRUS. This was performed with a thin plate spline transformation. This resulted in a transformation for all points within the PET/MRI prostate to be registered to the TRUS prostate, and was used to map the DILs to the TRUS.
Permission for Reproduction of Published Materials
Dear Mr. Ryan Alfano,

We hereby grant you permission to reprint the material below at no charge in your thesis subject to the following conditions:

1. If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies.

2. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

“This article was published in Publication title, Vol number, Author(s), Title of article, Page Nos, Copyright Elsevier (or appropriate Society name) (Year).”

3. Your thesis may be submitted to your institution in either print or electronic form.

4. Reproduction of this material is confined to the purpose for which permission is hereby given.

5. This permission is granted for non-exclusive world English rights only. For other languages please reapply separately for each one required. Permission excludes use in an electronic form other than submission. Should you have a specific electronic project in mind please reapply for permission.

6. As long as the article is embedded in your thesis, you can post/share your thesis in the University repository.

7. Should your thesis be published commercially, please reapply for permission.

This includes permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission.

This includes permission for UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission.

8. Posting of the full article/chapter online is not permitted. You may post an abstract with a link to the Elsevier website www.elsevier.com, or to the article on ScienceDirect if it is available on that platform.

Regards,

Kaveri

ELSEVIER | Permissions Granting Team
From: Administrator  
Date: Tuesday, April 13, 2021 07:29 PM GMT

Dear,

Thank you for contacting the Permissions Granting Team.

We acknowledge the receipt of your request and we aim to respond within seven business days. Your unique reference number is 210413-026412.

Please avoid changing the subject line of this email when replying to avoid delay with your query.

Regards,
Permission Granting Team

From:  
Date: Tuesday, April 13, 2021 07:29 PM GMT

Submission ID: 1158386  
Date: 13 Apr 2021 8:28pm  
Name: Mr. Ryan Alfano  
Institute/company: Western University

Type of Publication: Journal  
Title: Radiotherapy & Oncology  
Authors: Ryan Alfano, Glenn S. Bauman, Wei Liu, Jonathan D. Thiessen, Irina Rachinsky, William Pavlosky, John Butler, Mena Gaed, Madeleine Moussa, Jose A. Gomez, Joseph L. Chin, Stephen Pautler, Aaron D. Ward  
Year: 2020  
From page: 34  
To page: 41  
ISSN: 0167-8140  
Volume: 152  
Article title: Histologic validation of auto-contoured dominant intraprostatic lesions on [18F] DCFPyL PSMA-PET imaging  
I would like to use: Full article / chapter
I am the author of the Elsevier material: Yes
Involvement: My name is Ryan Alfano and I am the first author of the manuscript and was involved in designing the experiment methodology, data curation, preparation, and analysis and manuscript preparation.

In what format will you use the material: Print and Electronic
Translation: No

Proposed use: Reuse in a thesis/dissertation

Material can be extracted: Yes

Additional Comments / Information: To whom it may concern, My name is Ryan Alfano and I am a 5th year PhD student in Medical Biophysics at Western University in London, Ontario, Canada. I am writing my PhD thesis in integrated article format which requires me to use this publication as a Chapter in the thesis.

This email is for use by the intended recipient and contains information that may be confidential. If you are not the intended recipient, please notify the sender by return email and delete this email from your inbox. Any unauthorized use or distribution of this email, in whole or in part, is strictly prohibited and may be unlawful. Any price quotes contained in this email are merely indicative and will not result in any legally binding or enforceable obligation. Unless explicitly designated as an intended e-contract, this email does not constitute a contract offer, a contract amendment, or an acceptance of a contract offer.
Curriculum Vitae

Name: Ryan Alfano

Post-Secondary Education and Degrees:
University of Windsor Ontario, Windsor, ON
B.Sc. Physics with co-op
09/2012 - 08/2016

Western University, London, ON
Ph.D. Medical Biophysics
09/2016 - 08/2021

Honours and Awards:
Western Graduate Research Scholarship
09/2016-08/2021

Trainee Travel Award
11/2017

Department of Science Dean’s List
09/2013-09/2016

Outstanding Scholars Scholarship
01/2013-01/2014

Entrance Scholarship
09/2012

Related Work Experience:
Theranostics Journal Reviewer
09/2020

London Regional Cancer Center QA-ship
09/2018-12/2018, 09/2019-12/2019
Publications:


Conference Proceedings:


SPIE Medical Imaging: 109501D, 8-pages, (16/02/19-21/02/19). International. San Diego, CA, USA. Podium Presentation.


Refereed Abstracts:


nosis Model for Prostate Cancer Classification on Multi-Parametric MRI” Cellular and Molecular Imaging Symposium (01/05/18). Regional. Poster Presentation.


Volunteer and Leadership Experience:

1. Medical Biophysics Graduate Mentorship Program - Mentor
   London, Ontario, CA
   08/2020-05/2021

2. Imaging Network Ontario - Session Chair and Presentation Evaluator
   London, Ontario, CA
   03/2020

3. University of Windsor Science Society - Organizational member
   Windsor, Ontario, CA
   09/2012-08/2016

4. University of Windsor Physics Club - Organizational member
   Windsor, Ontario, CA
   09/2012-08/2016