Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Magnetic Resonance Imaging Radiomics - Towards the Personalisation of Radiation Therapy for Prostate Cancer

A THESIS SUBMITTED TO

THE UNIVERSITY OF DUBLIN, TRINITY COLLEGE

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY (PhD)

2020

Michelle Marie Leech

This thesis is dedicated to my late parents,

Michael and Margaret Leech.

Life is strange with its twists and turns
As every one of us sometimes learns
And many a failure comes about
When he might have won had he stuck it out;
Don't give up though the pace seems slow—
You may succeed with another blow.
Success is failure turned inside out—
The silver tint of the clouds of doubt,
And you never can tell just how close you are,
It may be near when it seems so far;
So stick to the fight when you're hardest hit—
It's when things seem worst that you must not quit.

DECLARATION

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

I consent to the examiner retaining a copy of the thesis beyond the examining period, should they so wish (EU GDPR May 2018).

Michelle Leech

March 2020

PLAGIARISM DECLARATION

I declare that the work in this thesis is my own.

The patient images utilised in this thesis were from the FUNCTPROST study and acquired via a data sharing agreement between Trinity College Dublin and Oslo University Hospital, Radiumhospitalet, Oslo, Norway, signed on 25th May 2018.

The FUNCTPROST study was approved by the institutional review board and the Regional Committee for Medical and Health Research Ethics, South East Norway (2010/1656). All patients provided written informed consent.

The construction of the phantom in this study was in collaboration with St. Luke's Radiation Oncology Network. The design of the phantom is my own.

Michelle Leech

March 2020

LIST OF ABBREVIATIONS

3DCRT: 3D Conformal Radiation Therapy

AAPM: American Association of Physicists in Medicine

ABS: Acrylonitrile Butadiene Styrene

ADC: Apparent Diffusion Coefficient

ADT: Androgen Deprivation Therapy

AI: Artificial Intelligence

ANOVA: Analysis of Variance

ANN: Artificial Neuron Network

ART: Adaptive Radiation Therapy

AUC: Area under the Curve

BOLD-MRI: Blood Oxygen Level Dependent Magnetic Resonance Imaging

BPH: Benign Prostatic Hyperplasia

CAIX: Carbonic Anhydrase IX

CBCT: Cone Beam CT

CCC: Concordance Correlation Coefficient

CG: Central Gland

CHHIP: Conventional versus Hypofractionated High Dose IMRT in Prostate Cancer

COV: Coefficient of Variation

COX-2: Cyclo-oxygenase-2

CT: Computed Tomography

CTCAE: Common Toxicity Criteria for Adverse Events

CTV: Clinical Target Volume

DAHANCA: The Danish Head and Neck Group

DCE-MRI: Dynamic Contrast Enhanced Magnetic Resonance Imaging

DICOM: Digital Imaging and Communications in Medicine

DVH: Dose Volume Histogram

ECE: Extra-Capsular Extension

EPO: Erythropoietin

EQD2: Equivalent Dose at 2 Gy per Fraction

ESUR: European Society of Urogenital Radiology

FAZA: 5 Fluoro-Azomycin-Arabinoside

FDM: Fused Deposition Modeling

FLAME: Focal Lesion Ablative Microboost in Prostate Cancer

FMISO: Fluoro-Misonidazole

FOV: Field of View

GTV: Gross Tumour Volume

GLCM: Grey Level Concordance Matrix

GLDZM: Grey Level Distance Zone Matrix

GLRLM: Grey Level Rung Length Matrix

GLSZM: Grey Level Size Zone Matrix

GLUT-1: Glucose Transporter 1

HIF-1 alpha: Hypoxia Inducible Factor 1 alpha

HIF-2 alpha: Hypoxia Inducible Factor 2 alpha

HYPRO: Hypofractionated versus Conventional Fractionated Radiation Therapy for Patients with Intermediate of High Risk Localised Prostate Cancer

ICC: Intra Class Correlation

IGRT: Image-Guided Radiation Therapy

IH: Intensity Histogram

IMRT: Intensity Modulated Radiation Therapy

iNOS: Inducible Nitric Oxide Synthase

ISUP: International Society of Urological Pathology

MAPS: Morphology, Asymmetry, Physiology, Size

MRI: Magnetic Resonance Imaging

NA: Number of Acquisitions

NCCN: National Comprehensive Cancer Network

NGLDM: Neighbouring Grey Level Dependence matrix

NGTDM: Neighbouring Grey Tone Difference Matrix

NSCLC: Non-Small Cell Lung Cancer

OAR: Organ at Risk

OCOG: Ontario Clinical Oncology Group

PET: Positron Emission Tomography

PHD2: Prolyl Hydroxylase 2

PIMO: Pimonidazole

PI-RADS: The Prostate Imaging Reporting and Data System

PCA: Principal Components Analysis

PMMA: Polymethylmethacrylate

PRECISION: Prostate Evaluation for Clinically Important Disease: Sampling using Image Guidance or Not.

PROMIS: Prostate MRI Study

PROPELLER: Periodically Rotated Overlapping Parallel Lines with Enhanced Reconstruction

PROTECT: Prostate Testing and Cancer Treatment

PSA: Prostate Specific Antigen

PVA-C: Polyvinyl Alcohol Cryogel

PVP: Polyvinylpyrrolidone

PZ: Peripheral Zone

RADAR: Randomised Androgen Deprivation and Radiotherapy

RIDER: Reference Image Database to Evaluate Therapy Response

ROI: Region of Interest

ROC: Receiver Operating Characteristic

RP: Radical Prostatectomy

RQS: Radiomics Quality Score

RTOG: Radiation Therapy Oncology Group

SABR: Stereotactic Ablative Body Radiation Therapy

SBW: Sampling Bandwidth

SI: Degree of signal attenuation

SMOTE: Synthetic Minority Oversampling Technique

STATS: Statistics

STL: Stereolithography

SUV: Standard Uptake Value

SVM: Support Vector Machine

TCIA: The Cancer Imaging Archive

TCP: Tumour Control Probability

TGF Beta: tumour Growth Factor Beta

TPU: Thermoplastic Polyurethane

TROG: Trans-Tasman Oncology Group

TRUS: Trans Rectal Ultrasound

TZ: Transition Zone

VEGF-A: Vascular Endothelial Growth Factor A

VMAT: Volumetric Modulated Arc Therapy

ABSTRACT

The primary function of a predictive biomarker is to accurately determine the outcome of a specified treatment. Radiomics is a developing field that involves advanced image analysis and high throughput extraction of mineable precise quantitative imaging descriptors or features that serve as non-invasive prognostic or predictive biomarkers.

Difficulty in attaining reproducibility of radiomics features has hindered their widespread use in the clinic to date. In Study 1 of this thesis, the impact of image acquisition parameters on feature reproducibility in T2-weighted MRI is analysed using a custom-designed 3D printed prostate phantom. This study concludes that the parameter with the greatest influence on feature reproducibility is slice thickness and that when a reasonable level of spatial resolution has been achieved, matrix size and field of view have limited impact on feature reproducibility.

In Study 2, the impact of image pre-processing on reproducibility of radiomics features in multiparametric MRI for prostate cancer is explored. From this work, it can be concluded that image normalisation and voxel resampling are critical to reproducible feature extraction and that linear interpolation algorithm appears favourable to nearest neighbour, particularly in the reproducibility of textural features.

Radiation therapy, using a variety of delivery methods, is a definitive management modality in treating prostate cancer. One of the main challenges in the management of prostate cancer is the stratification of those with clinically significant disease from those with indolent disease, as the treatment trajectories of such patients differ. Another issue is the potential to identify and dose escalate the hypoxic sub-volume, a known cause of radiation therapy treatment failure. Studies 3 and 4 of this thesis explore if radiomics features extracted from multiparametric MRI images can predict for pimonidazole (PIMO) scoring as a surrogate for hypoxia as well as some of the current clinical endpoints used in the stratification of prostate cancer (Gleason score at biopsy, clinical T stage and D'amico risk stratification score). Study 3, based on T2-weighted MRI, reports statistically significant models for prediction of PIMO score, Gleason score at biopsy, clinical T stage and D'amico risk classification

score. ROC analysis of these models identifies that their predictive potentials range from equivocal for clinical T stage (AUC=0.501) to moderate for D'amico risk classification score (AUC=0.753).

Study 4, based on DW-MRI and ADC maps, reports that the first order feature Stats_Var and the second order feature GLCM_InverseVar are present in all models identified predicting for PIMO score, Gleason score at biopsy, clinical T stage and D'amico risk stratification score. This indicates that they may have potential in the stratification of clinically significant from indolent prostate cancer. The textural feature GLRLM_RLV is identified in 4 models predicting for PIMO score, with the DW-MRI radiomics-only model yielding the highest predictive potential, with an AUC of 0.582. D'amico risk stratification score is predicted by models that included the morphological feature Shape_AreaDensityBB, with AUCs of 0.519 and 0.638 for DW-MRI and ADC maps, respectively.

This thesis reports reproducibility considerations for radiomics features in multiparametric MRI for prostate cancer. It also reports models with low to moderate prediction for PIMO score in prostate cancer, as well as clinical endpoints currently used in the stratification of prostate cancer patients. These models require external validation.

LAY ABSTRACT

Radiation therapy is one of the main modalities for treating clinically significant prostate cancer. Stratification of patients into those who have significant and less significant disease is important to ensure that those patients who require intensified treatment receive it and to prevent 'over-treatment' of those patients who could be maintained on active monitoring, a process known as 'active surveillance'. It is also known in radiation therapy practice that tumours that do not have sufficient oxygen (known as hypoxia) will likely not be cured.

Current stratification measures are based on tests, some of which are invasive and uncomfortable for the patient. Staging of and planning for any cancer treatment involves acquiring numerous diagnostic imaging scans and Magnetic Resonance Imaging (MRI) scans in particular for prostate cancer. These are typically reviewed by radiologists.

This thesis explores if the information contained in MRI scans can be further exploited using a method known as 'radiomics' and if the information from this process can be utilised in predictive models to stratify patients for treatment and assess which tumours are hypoxic. It also explores how images need to be acquired and the processing that needs to be conducted on the images before radiomics analysis to ensure generalisability across different clinical departments.

VALUE OF THE RESEARCH

This research contributes to the field of radiomics in radiation oncology. First, to our knowledge, there are no reported phantom studies specifically on MRI radiomics analysis in the literature, so Study 1 of this thesis is unique in this regard.

As there are no defined pre-processing measures specifically recommended for MRI radiomics analysis to date, Study 2 provides valuable results in recommending pre-processing guidelines for MRI radiomics research.

Radiomics analysis of MRI sequences to predict pimonidazole score as a surrogate for hypoxia has not been reported in the literature to date. Studies 3

and 4 of this thesis explore this relationship and discuss its potential in radiation therapy workflow.

RESEARCH OUTPUTS

Manuscript published:

Leech M, Gaffney J and Marignol L. Improving non-invasive detection of prostate cancer using diffusion-weighted MRI. Advances in Modern Oncology Research. 2016. 2 (6): doi: 10.18282/amor.v2.i6.152

Manuscripts under preparation:

- A review of Radiomics in the personalisation of treatment in Prostate Cancer.
- The identification of hypoxia in Prostate Cancer through MRI Radiomics and Pimonidazole Scoring.
- The impact of image acquisition parameters on reproducibility of T2 - weighted MRI Radiomics - A 3D printed phantom analysis.
- The impact of image pre-processing on the reproducibility of Radiomics features for multiparametric MRI in Prostate Cancer.

Presentation:

Invited Presentation: Image feature extraction using diffusion-weighted MRI in the development of personalised radiation therapy for intermediate and high risk prostate cancer. MAASTRO Clinic, The Netherlands. September 2016.

ACKNOWLEDGEMENTS

Thank you to my supervisor, Dr. Laure Marignol for your guidance and extreme patience! I also want to thank you sincerely for managing the department in my absence these past months. Thank you to Prof. Michael Gill, Head of School of Medicine, for facilitating my leave of absence to complete the writing of this thesis.

Many thanks to Prof. Heidi Lyng and Dr. Tord Hompland from Oslo University Hospital, Norway, for sharing their images from the FUNCTPROST study for radiomics analysis in this thesis.

Thank you to Dr. Ralph Leijenaar and Mr. Edward Scrivener from Oncoradiomics for answering my numerous queries on use of the software.

Sincere thanks to Mr. Ciaran Malone and Mr. Richard Collier from St. Luke's Radiation Oncology Network for their immense patience as we went through revision after revision about how best to create the phantom. I am forever indebted to you both and to Mr. John McGivney and Mr. Stephen Broderick for introducing me to you.

Thank you to Dr. Suzanne Dennan and the MR radiographers in St. James's Hospital for facilitating the scanning of the phantom.

Thank you to Prof. Earle Waghorne for his assistance on attempt one of the phantom building. It was wonderful to have the input of a chemistry expert (and friend!) to assist with the solution calculations.

Thank you to all of my colleagues in the Discipline of Radiation Therapy for your encouragement and support.

Thank you to my ESTRO friends all around the world. Working with you is one of the most rewarding aspects of my professional career and you always help me to 'raise my game'.

To Prof. Mary Coffey, what can I say?! I made it and it's all thanks to you. Thank you for every opportunity you have given me over the years and for inspiring me to do the best I can for this profession every day. Looking forward

to many more of our ‘great ideas’ (!) days in the future and several glasses of Prosecco to celebrate, somewhere in Europe.

Thanks to my wonderful ‘elder lemons’- my sisters Anne, Betty (Liz), Colette and Geraldine and your families. You are forever looking out for me and encouraging me to keep going (the luck of being the ‘baby’ of the family). Thanks also for the practical help (child minding, thank you Motley!). They say you can ‘choose your friends and not your family’, but I’d choose you as my friends and my family any day. Xx

Thank you to my friends for their encouragement and their flexibility in accommodating weekends away etc. for ‘when Michelle is in the country!’. Thank you Mary Claire, Breda, Maria, Paula, Anne Marie, Claire, Rachel, Lesley and Caitriona for listening to my giving out at the various stages of this process. You’ll be glad to hear that it won’t be long until I find something else to give out about instead!

Thank you to my darling children Elliott and Vivienne. You are the light and joy of my life.

Finally, to my husband David Cass, thanks for being my harshest critic and my greatest supporter all rolled into one (and they are equally important!). Thanks for proof-reading, it’s always handy to have an editor in the house - I hope my grammar and syntax has met your exacting standards! Love you xx

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES.

CHAPTER 1 INTRODUCTION

1.1 PROSTATE CANCER

The recognised changes in demography (aging) and lifestyle (obesity) of the male population are predicted to give rise to an increase in the incidence of prostate cancer [1]. Patients with low-volume, low-grade disease, without extraprostatic extension, lymphovascular invasion or rare histology, such as small cell or sarcoma, can be offered conservative management strategies such as active surveillance, while those with a high risk of disease progression require definitive treatment. Existing guidelines for active surveillance in prostate cancer vary widely but the majority state that the most suitable patients are those with pre-treatment clinical stage T1c or T2 tumours, serum prostate specific antigen (PSA) levels <10 ng/ml, Gleason score at biopsy of 6 or less, a maximum of one or two tumour-positive biopsy core samples and/or a maximum of 50% of cancer per core sample [2]. The aim of active surveillance is to minimise treatment-related toxicity without compromising survival. This can be a difficult balance, as seen in the recently published 10-year follow up of the Prostate Testing and Cancer Treatment (ProtecT) trial, where men were randomised to definitive treatment (surgery or radiation therapy) or to active surveillance. Two cohorts included 1643 men who agreed to be randomised and 997 who declined randomisation and chose treatment. There was stronger evidence of greater prostate cancer mortality and lower rates of patient-reported harm in the active surveillance group, relative to the treatment groups [3].

Prostate cancer risk stratification is currently limited to clinical examination, serum prostate specific antigen and transrectal prostate biopsy [4]. Clinically significant prostate cancer is diagnosed when the Gleason score is ≥ 7 (including Gleason 3 + 4 with prominent but not predominant Gleason 4), and/or a volume ≥ 0.5 cc and/or extraprostatic extension, with 70-75% of all prostate cancers arising in the peripheral zone [5, 6]. Some patients with negative biopsy can have continued suspicious clinical features of prostate cancer, such as PSA > 4.0 ng/mL, suspicious digital rectal examination and abnormal prostate specific antigen velocity [7]. The current staging assessment underperforms, very likely due to the confounding issue of intratumoural heterogeneity [8]. Decisions on how to treat prostate cancer with radiation therapy are based on guidelines but such

guidelines have been developed for groups of patients and invariably will lead to overly aggressive treatment in some patients and insufficient treatment in others [9].. Cancer-specific biomarkers, including molecular markers Bax, Bcl-2, cyclo oxygenase, E-cadherin and Ki67 are likely to become part of future staging systems [10]. The expansion of medical imaging into the field of radiation oncology over past years means that the potential to develop prognostic image biomarkers is now a reality [9]. Multiparametric Magnetic Resonance Imaging (MRI) is one such imaging modality that is routinely utilised in prostate cancer and can be considered not only in its diagnosis but in the extraction of imaging biomarkers.

1.2. CURRENT MANAGEMENT OF PROSTATE CANCER WITH RADIATION THERAPY

Radiation therapy, using a variety of delivery methods, is one of the main definitive management modalities of prostate cancer, with approximately 30% of prostate cancer patients in the United States being currently treated this way [11].

Today the selection of radiation therapy for a particular prostate cancer patient is driven by clinical and pathological features defining spectrums of local recurrence risk. The assessment of the risk of recurrence and subsequent prostate cancer-specific mortality is based on a measure of the PSA level in a blood sample, the scoring of biopsy specimens by a pathologist (Gleason scores) and the estimation of tumour extent within (T) and outside the prostate (regional lymph nodes) [12-16]. Patients with PSA >20 ng/ml or biopsy Gleason score 8–10 or T2-3N0M0 localised prostate carcinoma are, for instance, recognised as high risk [15].

For these patients, optimal management remains unclear. A meta-analysis of trials with long-term follow-up has indicated that high-dose radiation therapy is superior to standard dose 3D conformal radiation therapy (3DCRT) in terms of preventing biochemical failure in localised prostate cancer patients, but that this did not translate into an improvement in either overall survival or prostate cancer specific survival [17]. Intensity modulated radiation therapy (IMRT) is

recommended over 3DCRT for the treatment of localised prostate cancer where a dose of >70 Gy is required [18, 19].

Data from randomised controlled trials recommend the combination of external beam radiation therapy with androgen deprivation therapy (ADT) to improve overall survival [16, 20], but the timing and duration of ADT has been an issue for some time. The Trans-Tasman Radiation Oncology Group (TROG) Randomised Androgen Deprivation and Radiotherapy (RADAR) trial assessed the addition of 12 months of adjuvant androgen suppression, 18 months of zoledronic acid, or both, in men with locally advanced prostate cancer who already received 6 months of androgen suppression and radiation therapy to the prostate. At the recent 10-year follow up, it was reported that 18 months of androgen suppression plus radiation therapy is more effective for locally advanced prostate cancer than 6 months of androgen suppression plus radiation therapy but that the addition of zoledronic acid is not beneficial [21].

Radiobiological models have indicated for some time that the α/β ratio of prostate cancer cells is 1.5 Gy and that they are therefore more sensitive to larger doses per fraction than conventional daily fractionation [10]. This has led to increased use of moderate hypo-fractionated radiation therapy based on the Radiation Therapy Oncology Group 0415 (RTOG 0415), the Ontario Clinical Oncology Group (OCOG), the Conventional versus Hypofractionated high dose IMRT in Prostate Cancer (CHHIP) and the Hypofractionated versus Conventional fractionated Radiation Therapy for patients with intermediate or high risk localised Prostate Cancer (HYPRO) trials [22-25]. RTOG 0415 included over 1000 patients. The trial design was non inferiority and the primary endpoint was disease free survival. No difference was found between moderate hypofractionated (70 Gy in 28 fractions over 5.6 weeks) and conventional radiation therapy (73.8 Gy in 41 fractions over 8.2 weeks) cohorts. There was no significant difference in acute toxicity but late urinary and Grade 2 gastrointestinal toxicity were significantly worse in the hypofractionated arm [22] In the OCOG trial of 1204 intermediate risk patients, no difference in biochemical free survival between conventionally fractionated (78 Gy in 39 fractions over 8 weeks) and hypofractionated (60 Gy in 20 fractions over 4 weeks) arms was observed. There was an increase in acute gastrointestinal toxicity, but not in late toxicity [23]. The CHHIP trial included over 3000 patients of all risk groups but

mainly low and intermediate risk. No significant difference between conventional fractionation of 74 Gy in 37 fractions over 7.4 weeks versus 60 Gy in 20 fractions over 4 weeks hypofractionation was observed, but those treated to 57 Gy in 19 fractions over 3.8 weeks had inferior biochemical free failure. Toxicities were mainly grade 1 and 2 and it was observed that for the hypofractionated arms, acute toxicity started early in the treatment course and finished earlier than in the conventional arm. There was no difference in acute urinary side effects and acute gastrointestinal effects were slightly worse. No difference in late toxicity was observed [24]. The HYPRO trial included 820 intermediate and high-risk patients who were treated over 6 weeks. Superiority could not be proven for the primary endpoint, which was 5 year relapse-free survival; neither was non-inferiority of acute late effects proven. Late effects were the same [25]. All data from the above trials were pooled in a meta-analysis in 2019 and no difference in overall survival and an increase in biochemical free survival with little difference in toxicity (albeit with some outliers in gastro intestinal toxicity) were reported [26]. Therefore moderate hypofractionation, typically the CHHIP protocol of 60 Gy in 20 fractions over 4 weeks, is now considered a standard of care for low and intermediate risk prostate cancer. However, recent evidence appears that for ultra-hypo fractionated radiation therapy for prostate cancer, the effective doses are lower than anticipated, and that biochemical control may be saturated at 80 Gy equivalent dose at 2 Gy per fraction (EQD2) [27]. However, this is based on whole gland irradiation.

Furthermore, the preliminary results of the RTOG 0521 trial [28] suggest a small but significant increase in overall survival from 89% to 93% at 4 years with the addition of adjuvant docetaxel and prednisone to long-term (24 months) androgen suppression and radiation therapy in patients with high-risk localised prostate cancer. For patients with Gleason 9-10 prostate cancer, multi-institutional analysis of patient outcomes identified extreme dose-escalated radiation therapy and short-term androgen suppression as associated with the least risk of developing metastatic disease [29]. Dose escalation to the index lesion as determined by multiparametric MRI has been deemed safe by a single blind randomised Phase III trial to investigate the benefit of a focal lesion ablative microboost in prostate cancer (FLAME) trial, comparing standard fractionation external beam radiation therapy of 77 Gy to the whole prostate to standard fractionation and an additional integrated boost of up to 95 Gy to the tumour as

determined by multiparametric MRI [30]. MRI was scheduled prior to ADT. Patients were treated with IMRT or volumetric modulated arc therapy (VMAT) according to local dose volume constraints for normal tissue for both arms, together with additional constraints of 77 Gy to 1 cc of the rectum and 80 Gy to 1 cc of the bladder. The primary endpoint of this trial will be 5-year biochemical progression-free survival. Two year treatment-related side effects have been reported using the Common Toxicity Criteria for Adverse Events (CTCAE) version 3.0 at baseline and during follow-up. Of 571 patients enrolled, 287 patients received standard treatment and 284 patients received dose-escalated treatment, with the majority of both arms classified as high-risk patients. The overall toxicity profile between the two arms for both genitourinary and gastrointestinal toxicity at two years was similar.

More rigorous treatments include focal radiation therapy boosts to targeted prostatic lesions using brachytherapy or stereotactic ablative body radiation therapy (SABR), as well as the use of ADT or systemic therapies [31]. A meta-analysis of over 6000 patients from 38 prospective SABR prostate cancer studies reported 5 and 7 year biochemical recurrence-free survival rates of rates of 95.3% and 93.7% respectively. Estimated late grade ≥3 genitourinary and gastrointestinal toxicity rates were 2.0% and 1.1% respectively [32]. This indicates the significant potential of SABR in the management of localised prostate cancer in the future. With the advent of the hybrid MR-linear accelerator, the ability to utilise tight margins in the delivery of SABR in prostate cancer is very promising [33, 34].

Personalising of prostate cancer radiation therapy in the future will include consideration of individual patient biological factors, such as tumour hypoxia, radiation-induced damage repair and intrinsic cellular radiosensitivity, as well as the extraction of imaging biomarkers using computed tomography (CT), MRI and PET. This thesis proposes radiomics, the high-throughput extraction and mining of large amounts of image features from standard of care MRI images, as a potential tool to assist in the personalisation process.

1.3. INTRODUCTION TO RADIOMICS

The primary function of a predictive biomarker is to accurately determine the outcome of a specified treatment [35]. Radiomics is a developing field that involves advanced image analysis and high throughput extraction of mineable precise quantitative imaging descriptors or features that serve as non-invasive prognostic or predictive biomarkers [36]. It is based on the hypothesis that tumour characteristics at cellular and genetic levels are captured in the phenotypic patterns seen in medical images [37]. These features are often extracted from medical images using advanced mathematical algorithms [38]. Currently, imaging features are generally only visually assessed by radiologists or nuclear medicine physicians and described as qualitative biomarkers [37].

The field of radiogenomics tries to correlate imaging characteristics with gene expression patterns, gene mutations and other genome-related characteristics. It aims to facilitate a deeper understanding of the intrinsic heterogeneity in tumours, with the overall goal of developing imaging biomarkers for outcomes that incorporate both phenotypic and genotypic metrics [39]. Meanwhile, radiomics is the methodology behind the conversion of medical images with data of interest, including patient characteristics, outcomes and omics data to support decision-making, including stratification of patients for therapy [40].

1.4. MULTIPARAMETRIC MRI IN THE DIAGNOSIS OF PROSTATE CANCER

1.4.1 T2-WEIGHTED MAGNETIC RESONANCE IMAGING.

Multiparametric MRI provides excellent contrast of anatomic structures on T2-weighted imaging as well as diffusion characteristics on Diffusion Weighted MRI (DW-MRI) and associated Apparent Diffusion Coefficient (ADC) maps [41]. Multiparametric MRI has a high sensitivity and high negative predictive value for aggressive prostate cancer [42] but its sensitivity is significantly higher for tumours > 5 mm in diameter [43] and for tumours with Gleason scores >7 [44]. T2-weighted imaging is an MR imaging modality where the sensitivity of the tissue is characterised by measuring the relaxation time (spin-spin) of the applied magnetic field [45]. T2-weighted imaging yields relatively high-resolution images

of the prostate and surrounding anatomy, allowing for appreciation of structural differences [46].

The prostate gland is divided into three anatomical areas: the peripheral zone, the transition zone and the central zone. Prostate cancer in the central zone is rare [41]. On T2-weighted imaging a 'bulging' effect in the peripheral zone is highly suspicious for prostate cancer, as is a lenticular shape in the transition zone [46]. However, T2-weighted imaging is known to be associated with intensity drift artefacts, which cause the tissue-specific signal intensities to vary across scans, even with those obtained for the same patient on the same day and on the same scanner [31]. Additional issues with identifying prostate lesions on T2-weighted imaging include mixed signal intensities caused by benign prostatic hyperplasia (BPH) nodules that are appreciated as hypointense on T2-weighted imaging and prostatitis, which can also cause decreased signal intensity [47]. Highly differentiated prostate adenocarcinoma can have a high signal intensity on T2-weighted imaging, which can give false negative results [48]. Therefore, we can conclude that T2-weighted imaging is sensitive to but not specific to prostate cancer detection and must be combined with functional MRI techniques [49].

A further consideration with multiparametric MRI is post-biopsy haemorrhage [50]. Haemorrhage is seen as a hypointense lesion on T2 weighted imaging, which may be mistaken for cancer. On DW-MRI it may appear as a lesion with low ADC values, similar to cancer. Multiparametric MRI is therefore not recommended until 3 weeks post-biopsy [51]. However, haemorrhage may persist for more than 6 weeks [52-54]. Ideally, MRI pre-biopsy would circumvent this issue. Pre-biopsy MRI can provide information on cancer location and thus allow for more accurately directed biopsy.

The Prostate MRI Imaging Study (PROMIS) trial [55] has indicated that multiparametric MRI can reduce over-diagnosis of clinically insignificant prostate cancer and improve detection of clinically significant cancer. However, multiparametric MRI missed clinically significant cancer 11% of the time in this trial when clinically significant prostate cancer was defined as Gleason 4+3 or higher and 24% of the time if defined as Gleason 3+4 or higher. 'Missed' significant prostate cancer cases are therefore the main concern with multiparametric MRI [56].

1.4.2 DIFFUSION WEIGHTED MOLECULAR RESONANCE IMAGING.

Molecular diffusion of water is a process where water molecules move in random thermal motion. In biological tissues, the signal intensity of DW-MRI depends on the distance between restrictive boundaries such as cell membranes, fibres and macromolecules as well as the permeability of these boundaries. Random diffusion of water in biological tissues occurs in the intracellular spaces, extracellular spaces and across the cell membrane [57]. When a strong magnetic field gradient is in proximity to a tissue, the random motion results in an irreversible dephasing of the MR signal and this diffusion thereby becomes a source of contrast in the resulting MR image.

In the most commonly used method for DW-MRI acquisition, two symmetric motion-probing gradient pulses are incorporated into a single-shot-spin-echo (SE) T2-weighted sequence, one on either side of the 180° refocusing plate. The diffusion gradient causes the phase shift to change with position in malignant tissue, where diffusion is restricted, meaning that all spins that are at the same location along the gradient return to their initial position. This differs from the free molecule diffusion observed in benign tissue, where spins that have moved are subject to changed field strength during the second pulse after a 50 msec delay and undergo a total phase shift, leading to a decreased intensity of the measured MR signal [57]. The degree of signal attenuation *SI*, depends on the signal intensity of the T2-weighted image with no diffusion gradient applied SI_0, the degree of diffusion weighting known by the *b value* and the diffusion coefficient *D*, and is given by Formula 1.1. The *b* value is determined by the gradient properties applied and increasing the amplitude and temporal spacing of the two motion-probing gradients can gradually increase the sensitivity of diffusion-weighted imaging to diffusion [58]. Typical *b* values for prostate imaging are in the range of 0-1500 s/mm^2. In a high *b* value MRI image, areas with impeded diffusion, such as a tumour, will often appear to have higher signal intensity than that of surrounding tissue [59]. The random motion in the field gradients results in signal attenuation [60]. Other changes detected by DW-MRI include shifts of water from extracellular to intracellular spaces, restriction of cellular membrane permeability and disruption of cellular membrane depolarisation [61]. Intravascular water molecules may also add to signal intensity in DW-MRI but are negated when heavily diffused weighting is utilised

[57]. One of the limitations of DW-MRI images is that they can be of low spatial resolution [38], but this is overcome when combined with T2-weighted imaging.

$$SI = SI_0 \text{ X } \exp(-b \text{ X } D)$$

Formula 1.1. The degree of signal attenuation in DW-MRI.

Performing DW-MRI with two or more *b* values allows for the calculation of the ADC, which quantifies the slope of the line that describes the logarithm of the measured signal intensity and the *b* value. DW-MRI induces signal decay through diffusion, which results in dark areas where water can diffuse and, as ADC values measure the rate of diffusion of tissue, areas of limited diffusion appear dark [46].

Endorectal coils provide superior signal-to-noise ratio compared with a pelvic phased array coil in prostate cancer but cause reduced patient compliance and increase susceptibility artefacts [59]. Standardisation of DW-MRI techniques is difficult and a large variety of imaging parameters exist for DW-MRI, including number and size of *b* values, diagnostic threshold and type of coil [62].

1.5. PROSTATE IMAGING REPORTING AND DATA SYSTEM (PI-RADS)

The Prostate Imaging Reporting and Data System (PI-RADS) is a structured method for interpreting and reporting results from multiparametric MRI in the diagnosis of prostate cancer [63] On its update in 2019, it states that multiparametric MRI is still preferable in men where the clinical priority is not missing any significant cancers compared with overdiagnosis [64] Meta-analysis indicates that PI-RADS can detect tumours with a sensitivity of 0.70 to 0.84 and with a specificity ranging from 0.68 to 0.86 [65]. Furthermore, up to 25% of prostate cancers in the transition zone may be undetected on MRI, mainly due to the difficulties with additional pathology such as BPH [41]. Moreover, it has been cited that radiologists can miss up to 40% of clinically significant prostate cancers [44]. The assigned PI-RADS score is attributable only to multiparametric MRI findings and does not consider any clinical factors, such as PSA, history or previous biopsy. PI-RADS recommends that lesions with a PI-RADS score of 3, 4,or 5 should be reported, up to a maximum of four lesions. This scoring system gives information on lesion size and whether there is invasive behaviour or definitive extracapsular extension but until recently it failed to provide detailed information on the heterogeneity within the tumour [63]. However, the revised PI-RADS 2 guidelines now advocate for the use of tumour signal heterogeneity on multiparametric MRI to assist in grading prostate cancer [63].

A fundamental objective in the management of prostate cancer is to differentiate indolent from aggressive disease and subsequently provide risk-adapted, personalised care. Risk stratification at an earlier time point would result in more rigorous treatment strategies being undertaken at the outset for aggressive disease. Risk stratification is dependent on accuracy at biopsy and the use of MRI-targeted biopsy has been demonstrated by the Prostate Evaluation for Clinically Important Disease: Sampling Using Image Guidance or Not (PRECISION) multicentre randomised noninferiority trial as improving the detection of clinically significant prostate cancer relative to standard biopsy [66]. Men who had a positive result on multiparametric MRI underwent MRI targeted biopsy with real-time ultrasound guidance (51 with PI-RADS v.2 score of 3, 70 with score of 4 and 54 with a score of 5). A maximum of 12 biopsy cores per patient were obtained. Men who had a negative result on multiparametric MRI (71 with PI-

RADS v.2 score of 1 or 2) were not offered this protocol. Clinically significant cancer was detected in 95 (38%) of the MRI-targeted biopsy group compared with 64 (26%) in the standard-biopsy group. A higher percentage of cores was positive for cancer in the MRI-targeted biopsy group (422 of 967 cores, 44%) than in the standard-biopsy group (515 of 2788, 18%). As there was a lower percentage of men undergoing biopsy with fewer cores in the MRI-targeted biopsy group than the standard-biopsy group, patient-reported complications at 30 days were fewer in the former. Haematuria was reported as 30% in the experimental group compared with 63% in the control group, pain at the site of procedure was reported as 13% versus 23%, rectal bleeding was 14% compared with 22%, erectile dysfunction was reported as 11% versus 16% and haematospermia was reported as 32% in the MRI-targeted group and 60% in the standard-biopsy group.

1.6 HYPOXIA IN PROSTATE CANCER

High levels of hypoxia are associated with a poorer prognosis in prostate cancer [67]. Due to the inefficient oxygenation of tumour vasculature, tumours have been reported to have high regions of hypoxia, relative to the surrounding normal tissue and both acute and chronic hypoxia exist within tumours. Regions of a tumour with a high oxygen concentration are believed to be up to three times more amenable to radiation therapy than are hypoxic regions [68]. Malignant prostate cells secrete proangiogenic molecules, including, but not limited to, vascular endothelial growth factor A (VEGF-A), Tumour growth factor β (TGFβ), cyclo-oxygenase-2 (COX-2), prolyl hydroxylase (PHD2), inducible nitric oxide synthase (iNOS), carbonic anhydrase IX (CAIX), hypoxia inducible factor 1α (HIF-1α), hypoxia inducible factor 2α (HIF-2α), erythropoietin (EPO), glucose transporter 1 (GLUT-1), E-cadherin and fibroblast growth factor 2 [69, 70]. These molecules can also serve as intrinsic biomarkers in the identification of hypoxia, but to date, none has successfully been implemented in routine clinical practice for the assessment of hypoxia [71]. Other methods for identifying hypoxia in tumours include the invasive Eppendorf electrode method, extrinsic markers; such as pimonidazole (PIMO); and imaging techniques such as dynamic contrast enhanced MRI (DCE-MR), blood oxygen level dependent MRI (BOLD-MRI) sequences; or positron emission tomography (PET) using hypoxia-seeking ligands such as fluoro-misonidazole (FMISO), 5 fluoro-azomycin-arabinoside (FAZA) and

[^{18}F]HX4 [72, 73]. However, these imaging techniques are not able to fully sample the tumour microenvironment and both acute and chronic hypoxia are believed to be micro-regional [74]. PIMO is a 2-nitroimidazole compound which forms covalent bonds with cellular macromolecules at oxygen levels below 1.3% and allows for the visualisation of poorly oxygenated regions in histological specimens [75]. PIMO was first developed as a radiosensitiser but failed to demonstrate efficiency and is now used exclusively as an exogeneous marker, likely to reflect both acute and chronic hypoxia in tumours [76]. In head and neck cancer, it has been reported that patients with high PIMO scores have worse locoregional control than those with low scores [77].

There is evidence that intratumoural hypoxia is associated with cancer progression through HIF-mediated regulation of molecules that affect angiogenesis and metastasis, including VEGF-A and CAIX [70]. As the mainstay of radiation cell damage is through double strand DNA breaks directly or indirectly, through accumulation of reactive oxygen species, it is widely recognised that hypoxia in tumours affects response to radiation therapy and that a hypoxic tumour microenvironment may even be a means for tumour progression and metastasis [78]. The genetic traits that hypoxic tumour cells acquire during malignant progression also contribute to radioresistance. These cells still undergo proliferation, despite their damaged DNA and loss of apoptotic potential [71]. When a tumour grows to > 200μm, angiogenesis is induced by creating capillaries from already present micro vessels, increasing microvasculature [79]. High cell density areas with limited perfusion within the tumour are problematic as they can represent cell populations that have adapted to live under such conditions and, although hypoxic, are resistant to apoptosis and therefore become resistant to radiation therapy [72]. One of the main issue with hypoxia in prostate cancer is that although it can be identified using one of the imaging techniques described with excellent precision, the prostate moves during external beam radiation therapy delivery, prompting some to advocate for brachytherapy as the radiation therapy modality of choice to boost the hypoxic region [72]. The development of a non-invasive strategy, such as radiomics, to identify hypoxic regions within tumours is of much interest as identification of such features could then become standard practice [80] and dose escalation to hypoxic regions could lead to improved treatment outcomes. An example of an hypoxia-modified protocol is the now standard inclusion of nimorazole into the

treatment protocol of patients with highly hypoxic head and neck tumours in the Danish Head and Neck Group (DAHANCA) trials [81].

1.7. RADIOMICS PROCESS

Radiomics offers the potential to analyse whole tumours in 3D, as well as sub-regions or 'habitats' within tumours [82]. Heterogeneity within tumours and in metastatic sites, even within the same patient, is believed to be a major cause of treatment failure [83]. Combining quantitative information from imaging with pathology, demographic and other biomarkers will pave the way for personalised treatment selection and monitoring [39]. Intra- and inter-tumour heterogeneity exist within and between patients, which means malignant tumours can show features that are associated with poor treatment outcome and negative tumour behaviour. The ability to ascertain such variances in tumour phenotype through a non-invasive imaging method is a key feature of radiomics and assists in the planning and personalisation of treatment to individual patients; this is a move away from current population-based treatment protocols, as previously discussed. Radiomics biomarkers are more commonly referred to as radiomics 'features'. These features provide readily available, cost-effective, non-invasive tools for screening, detecting tumours and serial monitoring of patients, including assessments of response to therapy and identification of therapeutic complications. Radiomics features can enable tracking of a particular tumour repeatedly over time (delta-radiomics) [84], map spatial heterogeneity within tumours and evaluate multiple different lesions independently within an individual [85].

The workflow of radiomics is given in Figure 1.1 and consists of image acquisition and processing, region of interest (ROI) definition, feature extraction, data integration and clinical application. All these elements are associated with inherent difficulties and each requires its own series of steps. The Image Biomarker Standardisation Initiative [86] is an independent international collaboration which is working on the standardisation of the steps in the radiomics process and whose recommendations are adhered to in this thesis.

MR images may require de-noising prior to ROI determination. Noise that may be captured during the MR process includes eddy-current distortions, physiological motion and instabilities within the scanning hardware itself [87]. In many cases,

such noise reduction is part of post-processing offered by vendors during image reconstruction. As MR images do not reflect physical parameters like in CT, their use in radiomics analysis is more complicated. MR images vary depending on magnetic field strength, voxel size, pulse sequences, vendor and reconstruction algorithms, which makes repeatability and reproducibility in MR radiomics quite a challenge [88] and therefore a focus of this thesis. For T1 and T2-weighted MRI, the intensity signal is non-quantitative, which means that, technically, their comparison across patients is not possible without image pre-processing. For quantitative sequences, such as ADC maps, reproducibility and repeatability concerns are valid due to the large variation in acquisition parameters [89].

In radiomics, the image features extracted are based on the defined ROI. This is typically the demonstrable tumour, gross tumour volume (GTV), but radiomics analysis can also take place in critical normal structures called organs at risk (OAR). The ROI can be manually delineated by a clinician utilising knowledge of the natural history of the cancer, as well as accompanying clinical information, such as histopathological and radiology reports. Such ROIs are termed 'observer-defined ROIs'. These ROIs can be considered subjective, as they depend on the information available to the clinician to perform the delineation. Much research has focused on the impact of inter-observer and intra-observer variability in ROI delineation in the field of radiation oncology and is summarised in a systematic review by Vinod et al [90]. To overcome such subjectivity, auto-segmentation or auto-delineation software has been developed in recent years. However, auto-segmentation multi-atlas and hybrid-based algorithms also have to overcome image noise, differences in voxel intensity over a ROI and partial volume effects due to pre-defined voxel sizes [91]. Therefore there is increased interest in the use of artificial intelligence, particularly machine learning algorithms, to find segmentation solutions [92]. The machine learning methods that have been specifically used in auto-segmentation are artificial neural networks (ANN), support vector machine (SVM) and deep learning [93]. ANN is a machine learning approach modelled on human neural pathways. Each 'neuron' receives an input and can assign importance to a connection of the resultant network of outputs until a correlation is detected (Figure 1.2). Whatever the method of delineation, definition of the ROI is paramount as it influences the resultant downstream process and is known to be a weak link both in radiomics workflow and radiation

therapy workflow, in general. This ROI is then mined using automated extraction using data-characterisation algorithms [94, 95].

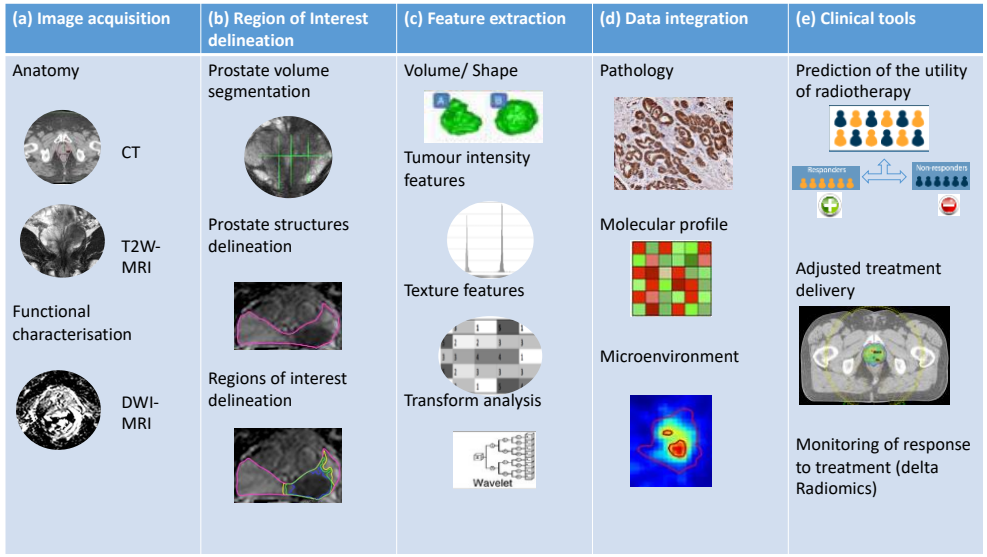


Figure 1.1. Schematic representation of the proposed integration of the radiomics workflow within a personalised approach for prostate radiation oncology. *a) The prostate can be anatomically defined using computational tomography (CT) and T2-weighted magnetic resonance imaging (MRI). Further non-invasive characterisation can be achieved using advanced molecular imaging techniques such as diffusion weighted imaging (DWI) MRI. b) Following systematic segmentation of the prostate, the prostate structures and tumour ROI can be delineated. c) Quantitative radiomics imaging features such as shape and textural can be extracted. d) These radiomics data can be combined with pathological assessment of prostate tissue, advanced molecular profiling (radiogenomics) and assessment of the tumour microenvironment. e) This integrated information can inform the utility of radiation therapy for the management of each patients tumour, the design of adjusted radiotherapy treatment delivery and facilitate monitoring of response to treatment.*

1.7.1. INTRODUCTION TO RADIOMICS FEATURES

Shape-based radiomics features, including geometrical shape and topological characteristics, are used to describe tumour growth directionality. Total volume, total surface area and surface to volume ratios and sphericity can be calculated [38]. Textural features characterise the spatial variations of signal intensity in an image [96]. First order statistics describe, for example, the mean, maximum, minimum, skewness, energy and entropy of the histogram. Second and higher order textural features can be extracted, for example, from grey level co-occurrence matrices (GLCM) and grey run matrices (GRLM) [38]. Features can run to hundreds or thousands. The feature families that will be explored in this thesis are detailed in Chapter 2.

As the number of features that may be extracted are numerous, appropriate feature reduction methods should be used in radiomics analysis. Feature selection assists in curtailing the 'curse of dimensionality', where a high number of features in limited cohorts can negatively affect the prognostic power of a model. As the number of features increases, a higher sample size is required [97]. To build reliable models, variable features must be tested for stability and reproducibility [98]. In addition, features must be explored for co-variance and features with high correlation identified as redundant and combined into a single descriptor [99].

Second ordered ranking creates a model with two or three highest priority features for each family [36]. This enables the correlation of prominent features with clinical outcomes such as overall survival or biological endpoints such as gene expression [100], or in this thesis PIMO score. Reducing features will not only make it easier to build models but also help avoid model overfitting and increase the generalisability of radiomics-based predictive models [101].

This mineable database of predictive and prognostic relationships between non-invasive imaging and clinical outcomes [38] provides a basis for the incorporation of imaging features into predictive models for treatment outcome and the evaluation of their efficacy [39]. Features extracted from macroscopic images can reveal different genomic and proteomic patterns, which are useful in quantifying and monitoring phenotypic evolution during disease or treatment progression. Selection of task-specific features is an important step

and machine learning will be useful for this in the future. It is also possible to use an arbitrary 'cut off' of a dynamic range but it must be acknowledged that features with a lower range may contain useful information and may be missed using this method [38]. A validation dataset from an external institute is required to confirm the prognostic value of radiomic features, although it is acknowledged that acquisition of such datasets can be difficult, due to confidentiality issues, difference in data collection and high cost [102].

1.8 RADIOMICS IN OTHER TUMOUR SITES – WHAT CAN WE LEARN FOR PROSTATE CANCER?

The transformative potential of radiomics in prostate cancer radiation oncology is in its infancy but evidence from other sites points towards a tremendous potential to advance our ability to determine the utility of radiation therapy and deliver true personalised treatment. The seminal work in radiomics research was performed in seven independent cohorts of 1019 lung and head-and-neck cancer patients [36]. Analysis was performed of 440 CT radiomic features and the phenotypic differences of the tumours were quantified based on tumour image, intensity, shape, texture and multiscale wavelets. From this, a four-feature prognostic radiomic signature was built. Results indicated that the signature could identify both lung and head and neck cancer, suggesting a general prognostic tumour phenotype, thus having a high prognostic power. This work also demonstrated that routine imaging can provide diagnostic, prognostic and biological information, which is useful in the pursuit of personalised radiation therapy.

In the treatment of non-small cell lung cancer with SABR, the radiomics feature *Wavelet LLH stats range,* which measures the range of voxel intensity values, was prognostic for the development of distant metastases and while requiring validation in a larger cohort, could potentially serve as an imaging biomarker for stratification of such patients in the future [103]. Similarly, it has been demonstrated that radiomic assessment has superior ability to detect recurrences in early stage non-small cell lung cancer (NSCLC) within 6 months post stereotactic treatment, compared with standard physician detection, although validation of the histologic patterns of failure with radiomics is still needed [104]. In a CT-based radiomics analysis [105], 635 features were extracted from images of 182 patients who had undergone chemoradiation for

locally advanced lung adenocarcinoma. These features were extracted and evaluated against basic metrics for their ability to predict distant metastases or overall survival. Results from the study showed a significant link between features extracted from radiomics and distant metastases for patients with locally advanced lung adenocarcinoma. An independently validated radiomics signature was also created for distant metastases. In hepatocellular carcinoma, a poorly defined tumour margin on CT was predictive of a 61-gene doxorubicin resistance radiomics signature [106].

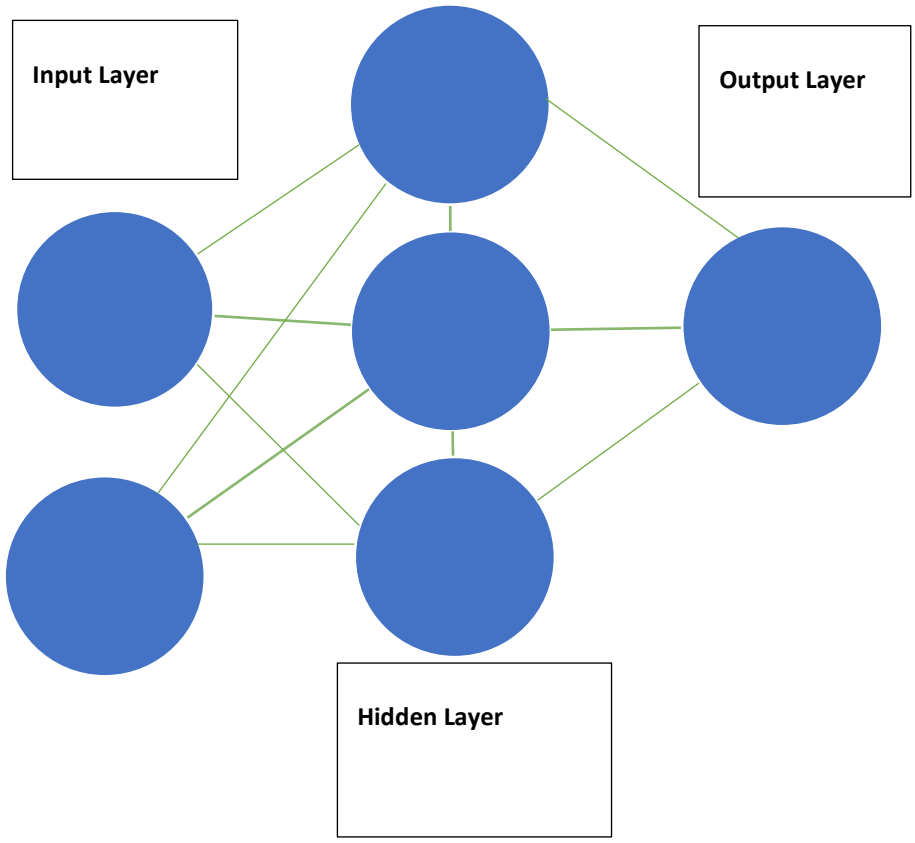


Figure 1.2. Schema of Artificial Neuron Network (ANN). *In an ANN, the input layer is based on data known to and inputted by the user. The hidden layer is where correlations between individual inputs are performed by the 'neurons' and weights assigned to the importance of each correlation. The results of this process is then communicated to the user in the output layer.*

1.9. RADIOMICS IN PROSTATE CANCER- WHAT DO WE ALREADY KNOW?

Most of the early work in prostate cancer radiomics focused on detecting prostate cancer by stratifying malignant from normal prostate cancer tissue. Ginsburg et al [41], when using a zone-specific radiomics classifier, reported a significant improvement in the accuracy of cancer detection in the peripheral zone. Eighty evaluable patients from three institutes were included in the study and 224 radiomic features, including edge descriptors, first-order statistical, co-occurrence and wavelet features were extracted from T2-weighted imaging and ADC maps. Gabor wavelet features, co-occurrence textural features and edge descriptors were important extracted features in distinguishing malignant from benign tissues in both the transition (TZ) and peripheral zones (PZ). Gabor wavelets appear to be particularly useful in this distinction in the TZ.

Another radiomics model (MAPS-Morphology, Asymmetry, Physiology, Size) for prostate cancer detection included the initial identification of tumour regions within the prostate using multiparametric MRI and regional morphology, followed by radiomics feature extraction. Extracted features were then split into high-level groups based on the MAPS features they described. The MAPS model achieved an accuracy of 87%, although it is noted that validation in a larger dataset remains necessary [46]. Wibmer et al [102] reported that textural features (energy, entropy, correlation, homogeneity and inertia) from T2-weighted imaging and ADC maps have the potential to differentiate between benign and malignant prostate tissue.

There is no doubt molecular imaging techniques are advancing our ability to assess the extent and aggressiveness of prostate cancer [107]. However, differential analyses of extensive genetic profiles of specimens, using advanced statistical classification methods, such as cluster analysis, are also progressing the use of genetic signatures from tumour tissue in providing additional prognostic information [108, 109]. For instance, molecular profiling of archived formalin fixed paraffin embedded tumour specimens has been used to try to predict radiation therapy outcome for high-risk prostate cancer patients through the analysis of genomic instability [110] and TMPRSS2-ERG gene fusion [109]. Co-expression of the ALDH1A1 beta-catenin in prostate tumor specimens was related to radiation therapy outcomes [111], and a shorter number of CAG

repeats on the androgen gene was associated with a benefit in the addition of androgen deprivation therapy to radiation therapy [112].

While not investigated in this thesis, radiomics offer a unique means to correlate genetic changes and image features [113]. The first evidence was provided by Stoyanova et al [82, 114] in a series of prostate cancer patients. The authors identified spatially distinct imaging areas using MRI-US fused data. Tissue from these identified targets was collected for pathology and gene expression analysis. The radiomics features were extracted from these biopsied areas and from nearby regions appearing 'normal'. The association between radiomics features and genomic classifiers was examined across six patients. All had more than two positive biopsies and PSA values ranged from 4.2 ng/mL to 10 ng/mL. Five patients had T1c tumours and one patient a cT2b tumour. Classification of prostate cancer using three commercially available genetic signatures (Prolaris Cell cycle Progression, Decipher and Genomic Prostate Score) successfully discriminated these patients according to Gleason Score and risk category. Similarly, 49 quantitative radiomics features were extracted to enable the characterisation of the prostate and each biopsy-targeted ROI. When the analyses were combined, a strong correlation between radiomics features and the genes associated with adverse outcomes across three commercially available prostate cancer classifiers was identified (two-way hierarchical clustering of Pearson correlation distances). Subsequent analysis further related radiomic features to distinct gene ontology biological processes (e.g. immune-inflammatory and cell-stress responses). Although this method requires validation in a larger independent cohort, it illustrates that targeted biopsies, using radiogenomic data, together with clinical data have the potential to directly impact the management strategy of patients. Further evaluation of the potential of this approach to enable the personalisation of prostate radiation therapy is warranted.

There is limited evidence to date on the use of radiomics to detect recurrence in prostate cancer. One study [96] identified with regression analysis that T2-weighted Haralick textural features were significantly associated with biochemical recurrence in a cohort of 74 patients, indicating the potential of radiomics in this field, pending further research.

1.10 FEATURE STABILITY IN RADIOMICS ANALYSIS

Stability of features extracted from images at different time points is an important factor in the potential use of radiomics features as imaging biomarkers. To serve as a tumour phenotype biomarker to aid in cancer diagnosis, the value of the biomarker must remain unchanged or minimally changed when the feature is recomputed from a repeat scan after a time interval [115]. Repeatability refers to a comparison of features extracted under constant conditions and reproducibility under varying conditions [116], but the terms appear to be used interchangeably in the literature. The most stable features between images taken at different time periods have been reported as the more robust to inter-observer variability in tumour delineation in lung cancer patients [117]. Features with higher stability for test-retest and delineation inaccuracies have shown higher prognostic performance in NSCLC [36]. Zhao et al [115] have found that for all imaging settings, non-texture features such as tumour size, shape, surface shape and histogram-derived density statistics are highly reproducible. For PET imaging, using a fixed bin size in units of standard uptake value (SUV) appears to result in better repeatability [117, 118]. While the Image Biomarker Standardisation Initiative has set standards for radiomics feature extraction in PET and CT there is as yet no consensus for MRI [119]

1.11 CONCLUSION

Rapid progress is being made in prostate cancer imaging and these advances are beginning to be applied to optimise prostate cancer radiation therapy. With increased methods of treatment intensification, clinicians urgently require better risk stratification tools in order to discriminate between tumours which can be managed with active surveillance and those that require definitive treatment. In the near future in radiation oncology, it is likely that patients will be treated with fewer treatment fractions with increased imaging and adaption to changing tumour and normal tissue. While the evidence base for radiomics in prostate cancer is in its infancy, it will not be utilised in isolation but together with clinical data, pathological data and, increasingly, genomic data. In the interim, considerable analysis and phantom-based studies are still required on the reproducibility of MR-based radiomic features [49] and this will be the focus of two studies in this thesis.

1.12 AIMS OF THESIS

The aims of this thesis are:

1. To determine if the radiomics models identified can be considered reproducible under differing imaging and pre-processing conditions.
2. To determine if radiomics features from multiparametric MRI can predict for pre-defined clinical features in prostate cancer.

These aims will be met through the completion of four studies. In Study 1, a custom-built pelvic phantom will be created and scanned at 1.5T with T2-weighted MRI multiple times, changing one scanning parameter at each acquisition. The reproducibility of radiomics features under these changed parameters will subsequently be analysed. In Study 2, the reproducibility of radiomics features will be analysed when pre-processing parameters, such as normalisation, are altered. This will be conducted using multiparametric MRI (T2-weighted diffusion weighted and apparent diffusion coefficient maps) of a prostate cancer cohort. Using the same cohort, Studies 3 and 4 will examine if radiomics features derived from T2 weighted and diffusion weighted and apparent diffusion coefficient maps respectively, can predict for clinical features, including hypoxia, in prostate cancer.

CHAPTER 2. MATERIALS AND METHODS

2.1. PATIENT COHORT

The study used the irrevocably anonymised data from 114 intermediate or high-risk prostate cancer patients, as per the D'amico risk classification, recruited as part of the FUNCTPROST study in Norway, investigating Consumption and Supply based Hypoxia Imaging (CSH Imaging) [120] under a data sharing agreement signed between Trinity College Dublin and Oslo University Hospital, Radiumhospitalet, Oslo, Norway, on 25th May 2018.

The FUNCTPROST study was approved by the institutional review board and the Regional Committee for Medical and Health Research Ethics, South East Norway (2010/1656). All patients provided written informed consent. The materials and methods have been previously described but the main points relevant to the current radiomics studies in this thesis are given below.

All patients were treated with radical prostatectomy (RP). All patients received the hypoxia marker pimonidazole (PIMO) prior to RP at a dose of 500 mg per m^2 body surface. Forty-three patients received pimonidazole hydrochloride (Hypoxyprobe Inc.) intravenously (IV) in 100 mL 0.9% NaCl IV over 20 minutes and 60 received oral PIMO due to a cessation in production of the IV product by the manufacturer; 11 patients received neither. This was due to either anaphylactoid reaction to MR contrast agent or absence of study personnel, or correct quantification of pimonidazole staining was not possible due to postponing of surgery [120].

RP was performed using a three-armed robotic DaVinci system (Intuitive Surgical). Pelvic lymph nodes were considered negative if negative on pre-operative MRI based on nodal diameter and morphology on T2 imaging and if there was an undetectable prostate specific antigen (PSA) 6 weeks post-prostatectomy or by pathological evaluation. Pelvic lymph node dissection was performed in the majority of high-risk patients and in those whose nodes were suspicious on pre-operative MRI.

The location of the index lesion was determined by an experienced radiologist, based on multiparametric MRI. Based on this location, the prostate was cut horizontally in half. The prostate specimen was fixed in 10% buffered formalin for a minimum of 48 hours and embedded in paraffin blocks. The tumour was outlined by an uropathologist based on 5-μm-thick haematoxylin and eosin (H&E)-stained whole-mount sections of each block. Histopathologic staging and grading were performed according to the TNM system and Gleason scoring system. Tumour size was defined as the largest extent of the index lesion, based on two perpendicular diameters. Where multifocality existed, the index lesion was determined based on pathologic T stage, Gleason score and tumour size, in this order, as per International Society of Urological Pathology (ISUP) guidelines [121].

2.1.2 MRI ACQUISTION

The scans were acquired adhering to the European Society of Urogenital Radiology (ESUR) 2012 guidelines [63]. The sequences analysed in this thesis were morphologic T2 **P**eriodically **R**otated **O**verlapping **P**aral**lEL L**ines with **E**nhanced **R**econstruction (PROPELLER), diffusion weighted (DW) (*b* = 1500) and apparent diffusion coefficient (ADC) maps. The images were acquired on a 1.5 Tesla General Electric Discovery 450 magnet using a 32 channel phased array coil (GE Medical Systems). The index lesion was outlined on these sequences, based solely on the pathologic specimen. The T2 sequence covered both the lower abdomen and the pelvis and the DW the pelvis only. Patients were placed as close to the magnet isocentre as possible to minimise bias in the DW parameters caused by possible gradient nonlinearity.

2.1.3 PIMONIDAZOLE ADMINISTRATION AND SCORING

Patients received either intravenous (IV) or oral PIMO. Those who received IV pimonidazole hydrochloride (Hypoxyprobe Inc) (n=43) had it applied in 100 mL 0.9% NaCL IV over 20 minutes. Sixty patients received oral pimonidazole in tablet form. All pimonidazole was administered 13 to 24 hours pre-prostatectomy. Eleven patients did not receive pimonidazole in any format (9 IV and 2 oral) due to either postponement of surgery, absence of study personnel at the required timepoint or anaphylactic reaction to the MR contrast agent. These patients have

still been included in the study as data on their other clinical features were available for analysis. The staining pattern was evaluated by an experienced pathologist, who was blinded to MRI. Immunoscores from 0 to 5 were attributed to the prostate glands in fractions of nuclear and cytoplasmic staining and the average value of these scores was used as the PIMO hypoxic score. The scoring system is outlined in Table 2.1 below.

Pimonidazole Hypoxic Score	Percentage Hypoxia
0	0%
1	1%-10%
2	11%-50%
3	51%-90%
4	91%-99%
5	100%

Table 2.1. Pimonidazole Hypoxic Scoring System used in the FUNCTPROST study [120].

2.1.4 REGION OF INTEREST (ROI) DEFINITION

The prostatic index tumour was contoured on T2, DW (*b*=1500) and ADC images on a single axial slice by an experienced radiologist, based specifically on the histological specimens, rather than on the images. Therefore, for example, no threshold values were applied for delineation on ADC images. The ROIs were saved as DICOM RTSTRUCTs, which is a set of *(x,y,z)* points that define closed loops planar polygons [86].

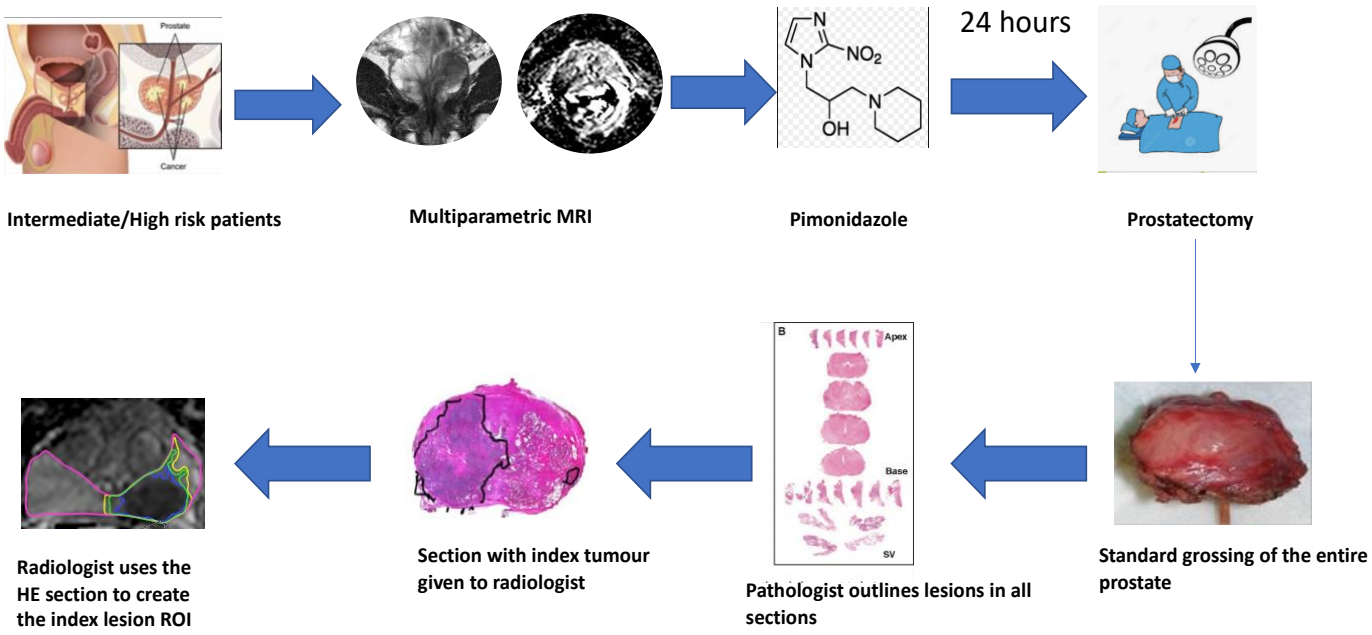


Figure 2.1. Steps in clinical procedure of the FUNCTPROST study [120].

2.2 RADIOMICS FAMILY FEATURE EXTRACTION

Radiomics family feature extraction on the defined ROI was performed using a special evaluation version of RadiomiX (RadiomiX Research Toolbox version 20180831(OncoRadiomics SA, Liège, Belgium)) for non-clinical use. In Studies 3 and 4,165 features were extracted based on the prostatic index lesion for each of T2, DW and ADC MR sequences. Before feature extraction, all features were normalised to zero mean, unit variance, resampling was performed using 3 mm isotropic voxels and the bin width utilised was 0.1. The interpolation algorithm utilised was linear. MR image intensity is relative and not directly comparable between images [122]. Normalising to zero mean, unit variance, meant that the image intensities were shifted and scaled. With this particular approach the resulting image intensities are expressed in standard deviations (σ). This resulted in a normalised image with intensity values roughly between -3 and 3 (σ), thereby reducing the residual noise across all ROIs. Most noise in MR imaging results from the Brownian motion of electrolytes in the patient's tissue [123]. Grey level normalization is also known to minimize the effect of brightness and contrast variations on textural feature analysis [124]. There were 165 features extracted for each MR sequence. The features were grouped in families, as described in the Image Biomarker Standardisation Initiative [86] in Table 2.2.

FEATURE FAMILY (NUMBER INCLUDED IN ANALYSIS)	FEATURES ANALYSED INCLUDED IN THIS FAMILY	FEATURE FAMILY MEANING
MORPHOLOGICAL/ GEOMETRIC FEATURES **(N=19)**	Shape_Asphericity Shape_Centroid distance Shape_Compactness 1 Shape_Compactness 2 Shape_Compactness 3 Maximum diameter (n =4) Shape_Major Axis length Shape_Minor axis length Shape_Elongation Shape_Spherical disproportion Shape_Sphericity Shape_Surface area Shape_Surface to volume ratio Shape_Volume Shape_Volume density bounding box Shape_Volume density ellipsoid	Morphological features are based on the voxels within the region of interest. They describe the geometric aspects of the region of interest.
LOCAL INTENSITY FEATURES **(N=2)**	Local intensity peak Global intensity peak	The local intensity peak is the mean intensity in a 1 cm^3 spherical volume, which is centred on the voxel with the

		maximum intensity level in the region of interest. The global intensity peak is the mean intensity calculated within a $1cm^3$ neighbourhood for every voxel in the region of interest and selecting the highest intensity peak value.
INTENSITY-BASED STATISTICAL FEATURES/ FIRST ORDER STATISTICS **(N= 19)**	Stats_Energy Stats_Entropy Stats_Kurtosis Stats_Maximum grey level Stats_Mean grey level Stats_Mean absolute deviation Stats_Median grey level Stats_Minimum grey level Stats_Range Stats_Root mean square Stats_Skewness Stats_Robust mean absolute deviation Stats_10^{th} percentile Stats_90^{th} percentile Stats_Interquartile range Stats_Uniformity Stats_Variance	First order or intensity-based statistical features describe the distribution of grey levels within the region of interest and are calculated from the volume.

	Stats_Coefficient of variation Stats_median absolute deviation	
INTENSITY HISTOGRAM FEATURES **IH (N= 26)**	IH_Coefficient of variation IH_Energy IH_Entropy IH_Interquartile range IH_Kurtosis IH_Maximum grey level IH_Maximum histogram gradient IH_Maximun histogram gradient grey level IH_Mean IH_Mean absolute deviation IH_Median IH_Median absolute deviation IH_Minimum grey level value IH_Minimum histogram gradient IH_Minimum histogram gradient grey level IH_Mode IH_10^{th} percentile IH_90^{th} percentile IH_Quartile coefficient of dispersion	The intensity histogram is created by discretising the original set of grey levels into grey level bins and is calculated using the region of interest volume.

	IH_Robust mean absolute deviation IH_Range IH_Root mean square IH_Skewness IH_Standard deviation IH_Uniformity IH_Variance	
GREY LEVEL CO-OCCURRENCE MATRIX **GLCM (N= 26)**	GLCM_autocorrelation GLCM_average GLCM_cluster prominence GLCM_cluster shade GLCM_cluster tendency GLCM_contrast GLCM_correlation GLCM_difference average GLCM_difference entropy GLCM_difference variance GLCM_dissimilarity GLCM_energy GLCM_entropy GLCM_homogeneity1 GLCM_homogeneity2 GLCM_informational measure of correlation 1	GLCM is a matrix that expresses how combinations of discretised grey levels of neighbouring voxels are distributed along one of the image directions.

	GLCM_informational measure of correlation 2 GLCM_normalised inverse difference moment GLCM_normalised inverse difference GLCM_Inverse variance GLCM_Maximun Correlation GLCM_Maximum Probability GLCM_Sum average GLCM_Sum entropy GLCM_Sum of squares GLCM_Sum variance	
GREY LEVEL RUN LENGTH MATRIX **GLRLM (N=16)**	GLRLM_Grey level non-uniformity GLRLM_Normalised grey level non-uniformity GLRLM_Grey level variance GLRLM_High grey level run emphasis GLRLM_Low grey level run emphasis GLRLM_Long runs emphasis GLRLM_Long run low grey level emphasis GLRLM_Long run high grey level emphasis GLRLM_Run entropy	GLRLM also assesses the distribution of discretised grey levels in an image, however it assess run lengths, which is the length of a consecutive sequence of voxels with the same grey level along the same direction.

	GLRLM_Run length non-uniformity GLRLM_Normalised run length non-uniformity GLRLM_Run length variance GLRLM_Run percentage GLRLM_Short runs emphasis GLRLM_Short run high grey level emphasis GLRLM_Short run low grey level emphasis	
GREY LEVEL SIZE ZONE MATRIX **GLSZM (N= 16)**	GLSZM_High grey level zone emphasis GLSZM_Low intensity small area emphasis, GLSZM_ Low intensity large area emphasis GLSZM_Intensity non-uniformity GLSZM_Normalised intensity non-uniformity GLSZM_Intensity variance GLSZM_Large area emphasis GLSZM_Low grey level zone emphasis GLSZM_High intensity small area emphasis	GLSZM counts the number of groups (zones) of linked voxels. Voxels are considered linked if the neighbouring voxel has the same discretised grey level.

	GLSZM_High intensity large area emphasis GLSZM_Small area emphasis GLSZM_ Size zone non-uniformity GLSZM_Normalised size zone non-uniformity GLSZM_Size zone variance GLSZM_Size zone entropy GLSZM_Zone percentage	
GREY LEVEL DISTANCE ZONE MATRIX **GLDM (N= 16)**	GLDZM_Zone distance entropy GLDZM_Zone distance non-uniformity GLDZM_Zone distance non-uniformity normalised GLDZM_Zone distance GLDZM_High grey level zone emphasis GLDZM_Low intensity large distance emphasis GLDZM_High intensity small distance emphasis GLDZM_Intensity non-uniformity GLDZM_Intensity non-uniformity normalised GLDZM_Intensity variance	The grey level distance zone matrix counts the number of groups (zones) of linked voxels which share a specific discretised grey level value and have the same distance to the edge of the defined region of interest. The GLDZM describes the relationship between the grey level and location and uses the same grey level zone map as the GLSZM as well as a distance map.

	GLDZM_Large distance emphasis GLDZM_Low grey level zone emphasis GLDZM_Low intensity small distance emphasis GLDZM High intensity large distance emphasis GLDZM_Small distance emphasis GLDZM_Zone percentage	
NEIGHBOURHOOD GREY TONE DIFFERENCE MATRIX **NGTDM (N= 5)**	NGTDM_Busyness NGTDM_Coarseness NGTDM_Complexity NGTDM_Contrast NGTDM_Strength	NGTDSM contains the sum of grey level differences of voxels with a discretised grey level and the average discretised grey level of neighbouring voxels within a defined distance. A valid neighbourhood is considered to exist if there is at least one neighbouring voxel included in the region of interest.
NEIGHBOURING GREY LEVEL DEPENDENCE MATRIX **NGLDM (N= 17)**	NGLDM_Small dependence emphasis NGLDM_Large dependence emphasis NGLDM_Grey-level non-uniformity NGLDM_Normalised grey-level non-uniformity NGLDM_Dependence non-uniformity	NGLDM is an alternate to the grey level co-occurrence matrix. It aims to capture the coarseness of the overall texture. It involves the concept of a neighbourhood around a central voxel.

	NGLDM_Normalised dependence non-uniformity NGLDM_Low grey-level emphasis NGLDM_High grey-level emphasis NGLDM_low grey-level small dependence emphasis NGLDM_High grey level small dependence emphasis NGLDM_Low grey level large dependence emphasis NGLDM_High grey level large dependence emphasis NGLDM_Grey-level variance NGLDM_Dependence variance NGLDM_Dependence Entropy NGLDM_Second moment NGLDM_Second moment 2	
FRACTAL FEATURES **(N= 3)**	Fractal Average Fractal Lacunarity Fractal Standard deviation	Fractal features quantify the geometric complexity by displaying repeating texture patterns at difference size scales [125]

Table 2.2. Radiomics Feature Families (Adapted from [86].

2.3 3D PRINTED PHANTOM CONSTRUCTION

A body structure was generated on an anonymised training CT dataset at St. Luke's Radiation Oncology Network at St. James's Hospital Dublin to represent an anatomical external contour of a pelvis. The CT scan was acquired at 120kVp with slice thickness of 2.5 mm. Four inserts were generated using the contouring module in Varian Eclipse (V.15.5) and post-processing tools in Meshmixer 3.5 (Figure 2.2) and 3DBolus (Adaptiiv Medical Technologies). The inserts were two cylindrical structures, which became the femoral heads in the phantom, a central cylindrical structure to house a simulated whole prostate volume (Figure 2.3 A) and an alternative central cylindrical insert, which was modified to include five cavities 6 mm in diameter, designed to house syringes which were then utilised to mimic individual prostate lesions, created in Meshmaker 3.5 (Figure 2.3 B). The final structures were exported to 3DBolus (Adaptiiv Medical Technologies) for post-processing and conversion to an STL file suitable for 3D printing. STL generation algorithms work at the voxel level [126]. Post-processing included the cleaving of the whole prostate cylindrical insert in 3DBolus (Adaptiiv Medical Technologies) and the ability to create a solid structure in Meshmixer 3.5, which could subsequently be hollowed to hold the inserts. The phantom was printed on an Airwolf Axiom 20 (Figure 2.3 C) using Acrylonitrile Butadiene Styrene (ABS). This 3D printer is capable of producing large models as its printer build envelope is 12.5″ X 12″ X 20″. Other material, such as Thermoplastic Polyurethane (TPU), was trialled for this phantom creation, but was not sufficiently rigid to prevent leakage. TPU is more suitable for the manufacture of customised skin bolus than phantom creation, due to its conformity and elasticity [127]. Supports for the inserts were also 3D printed, sanded and smoothed with a heatgun. In order to mimic soft tissue, the body of the phantom was completely filled with silicone (Z-Dupe- Schein), a vinyl-polysiloxane, which has minimal shrinking properties, excellent consistency of shape and a long shelf life (Figure 2.4). To prevent air bubbles and ensure a homogenous distribution, the phantom was placed on an agitation plate during silicone pouring. In order to mimic bone density, the inserts for the femoral heads were filled with silicone to which crushed chalk (gypsum) had been added. The final phantom is illustrated in Figure 2.5. The print settings utilised are given in Table 2.3.

2.3.1 PHANTOM SCANNING

The phantom was scanned on a 1.5T General Electric (GE) Signa H-DXT MR unit.

The scanning procedure and image analysis are described in Chapter 3.

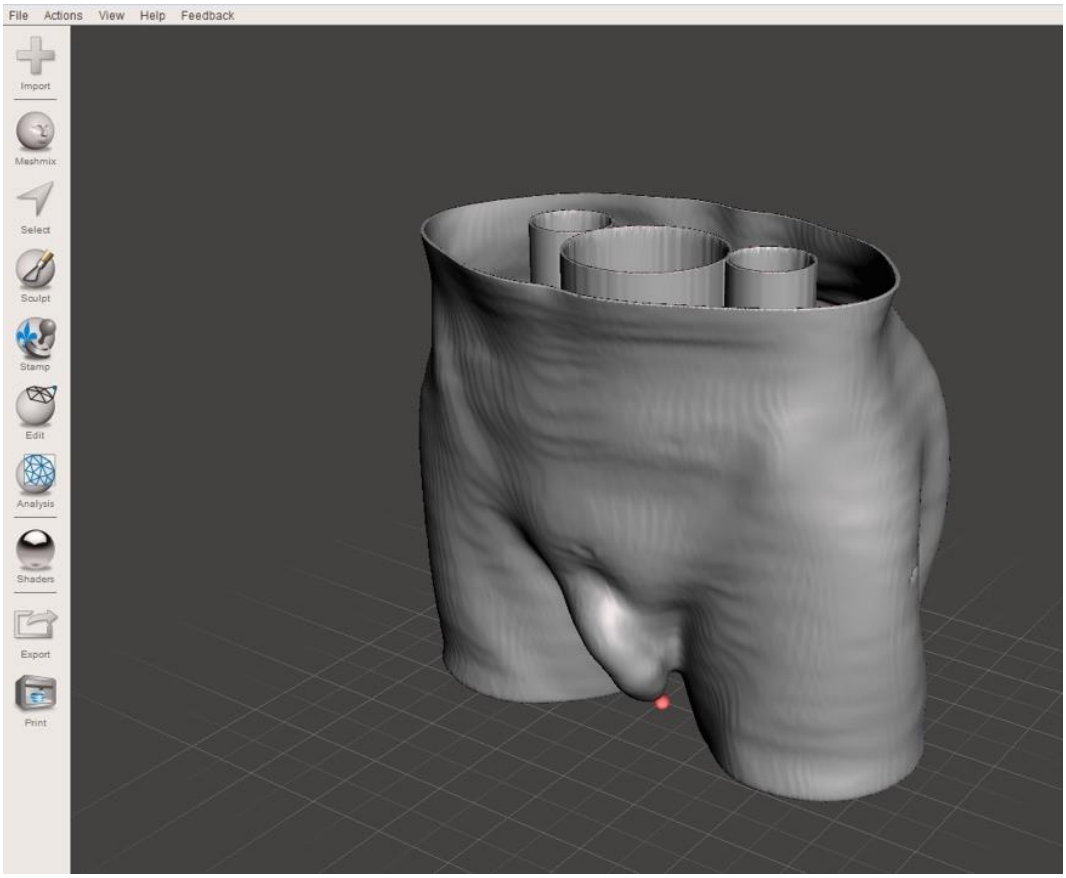
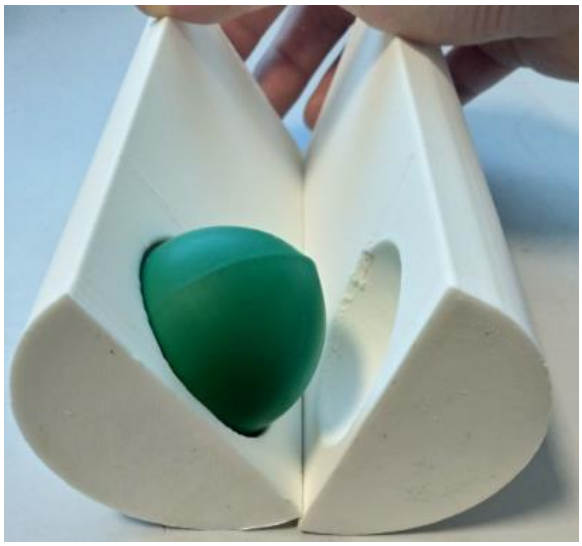


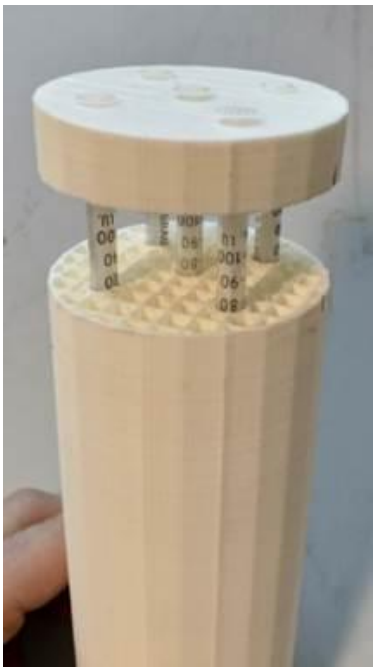
Figure 2.2. Post processing of inserts in Meshmaker 3.5.

Setting	Value
Temperature	250° Celsius
Print Speed	50 mm/second
Bed Temperature	140° Celsius
Nozzle Size	0.5 mm
Fill	0%
Wall Thickness	3 mm
Bottom/Top Thickness	3 mm

Table 2.3. Print settings utilised on Airwolf Axiom 20 for phantom printing.



A



B C

Figure 2.3. A: The insert simulating the whole prostate. This insert was cleaved and a spherical shape hollowed in which a balloon was inserted to hold de-ionised H_2O to simulate a measured prostate volume.

B: The insert which housed the 5 syringes which held de-ionised water to mimic individual prostate lesions of 1 cc to examine geometric distortion of the lesions at different regions of the field.

C: Airwolf Axiom 20 3D printer.

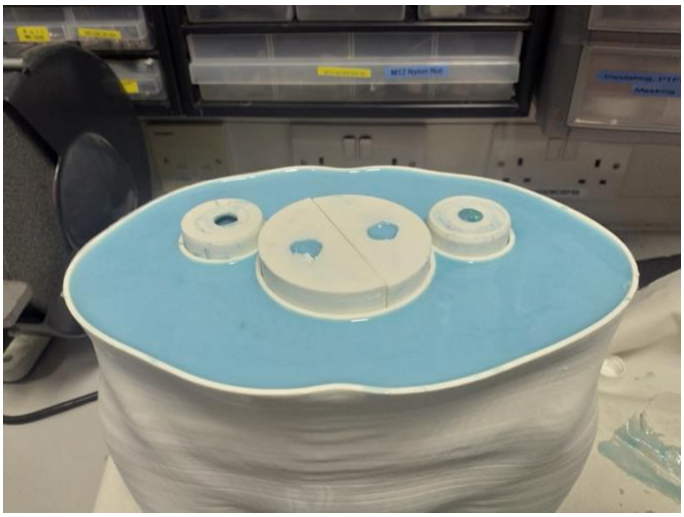


Figure 2.4. Phantom illustrating prostate and femoral heads inserts.

Silicone was utilised to act as a tissue surrogate and a mixture of silicone and gypsum to simulate bone density for the femoral head inserts.

Figure 2.5. Completed 3D printed phantom

2.4 STATISTICAL ANALYSIS

Statistical analysis was performed using SPSS Version 26 (IBM Corp. Released 2019. IBM SPSS Statistics for Macintosh, Version 26.0. Armonk, NY: IBM Corp).

2.4.1 PRINCIPAL COMPONENT ANALYSIS (PCA)

There are four underlying assumptions that must be met prior to conducting PCA [128-130].

1. Multiple variables must be for analysis, which are ordinal or continuous. PCA does not distinguish between independent or dependent variables. For analysis in this thesis, this was not an issue as the PCA was conducted on extracted radiomics features alone, without inclusion of clinical features.
2. A linear relationship must exist between all variables as PCA is based on Pearson correlation coefficients. This was tested by randomly selecting some possible relationships between variables and creating a simple scatterplot.
3. No outliers should be present because of their disproportionate influence on results.
4. Large sample sizes are required for PCA to produce a reliable result. In this thesis, data on 114 patients were included, which is at the lower end of acceptability for PCA but similar or higher in sample size to other published radiomics studies.

Three methods were utilised to determine which components to retain. They were a) the eigenvalue-one criterion, b) the scree plot analysis and c) the analysis of percentage variance explained.

a) An eigenvalue is a measure of the variance that is accounted for by a component. An eigenvalue less than one indicates that the component explains less variance than a variable would and therefore should not be retained [129].
b) The scree plot [130] illustrates the total variance explained by each eigenvalue against its respective component. There are as many components as variables in the scree plot. On analysis of the scree plot, the variables before the inflection point were those that were retained

in the analysis. The inflection point is meant to represent the point where the graph begins to level out and additional components contribute little to the total variance. All data up to this point are included in the analysis (Figure 2.6).

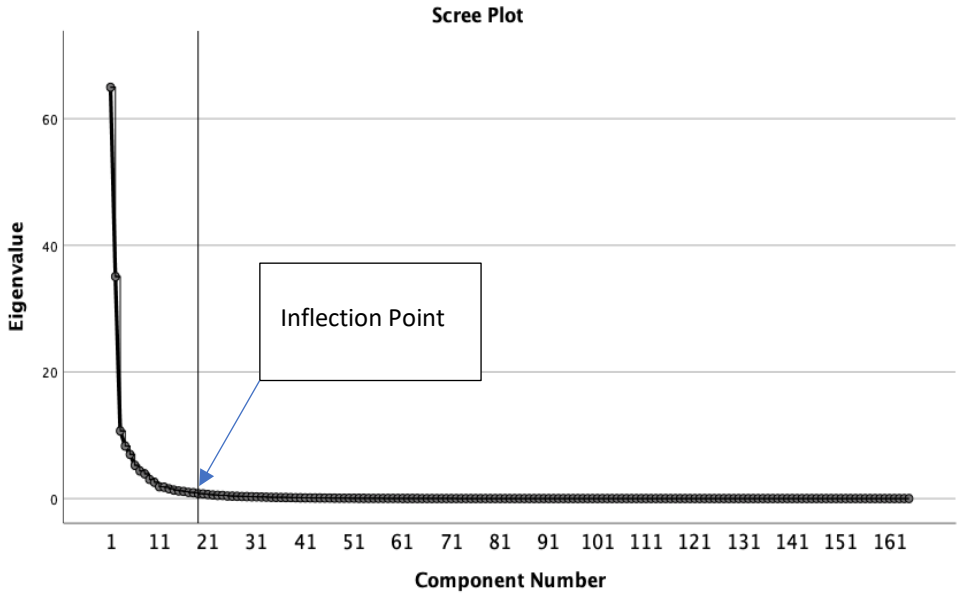


Figure 2.6. Scree Plot Analysis

c) As the component number increases, each subsequent component explains less of the total variance. It has been suggested that a component should only be retained if it explains at least 5% to 10% of the total variance. In this thesis, components that totalled at least 90% of the variance were included; this selection method will be expanded upon per MRI sequence in subsequent chapters.

Once the number of included PCs was determined, a forced factor number PCA was conducted with the number of PCs retained. This was performed in the same manner as the original PCA. This yielded the final components.

2.4.2 SCORING OF COMPONENTS

Component scores are the linear composite of the optimally weighted original variables. The statistical software computed the regression weights and multiplied each variable by the respective weights and summed the products. This resulted in the component score, which is the score that the radiomics

extraction achieves on that particular retained component. This was computed for all retained components. In order to determine which radiomics feature best represented the retained components and would be considered in further regression analysis, Pearson correlations were conducted on the retained PCs and the original 165 extracted features. The original features which had the highest Pearson correlation coefficient with the PCs were included in regression analysis as the PC representatives.

2.4.3 MULTIPLE LINEAR REGRESSION ANALYSIS

Multiple linear regression analysis (hitherto referred to as linear regression analysis) was performed on the retained representative components from the PCA and the clinical features of the patient cohort. In this thesis, linear regression was used to determine how much of the variation in any of the dependent variables analysed could be explained by the independent variables, which were largely the radiomics features determined by PCA. Dependent variables analysed were pimonidazole hypoxic score (PIMO score), Gleason score at biopsy, D'amico risk classification and clinical tumour (T) stage.

There are six underlying assumptions that must be considered before conducting a linear regression.

1. The dependent and independent variables must be measured at the continuous level.
2. A linear relationship must exist between the dependent and independent variables.
3. There must be independence of observations.
4. There should be no significant outliers.
5. Homoscedasticity exists, which means that the variances along the line of best fit remain similar as you move along the line.
6. There is approximate normal distribution of the residuals of the regression line.

Considering the independent variable to be "X" and the dependent variable to be "Y", a simple linear regression model would be as follows:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

where β_0 is the intercept, β_1 is the slope parameter and ε represents the residuals or errors.

To ensure that there is no autocorrelation of residuals, the Durbin-Watson test result was analysed in the linear regressions performed and a value close to 2 illustrated such independence. The assumption of homoscedasticity implies that the variance of the residuals is constant across all the values of the independent variable. This was checked by inspection of a plot of the unstandardised or standardised residuals against the fitted values or standardised fitted values, where a constant spread was expected. Residuals were also checked for normality using normal probability plots.

In order to assess how well the radiomics features extracted at PCA explained the variance in the dependent clinical features, the R^2 and adjusted R^2 statistics were analysed. R^2 is the proportion of the variance explained by the sample and adjusted R^2, which corrects for positive bias and therefore gives the proportion of the variance that would be explained in the population.

The statistical significance of the model was determined by analysing the analysis of variance (ANOVA) table. For this analysis, this indicated whether a radiomics feature can significantly predict a dependent clinical variable. The slope coefficient, β_1, represents the change in the dependent clinical variable for a one unit change in the radiomics feature. Where β_1 was statistically significant, it was interpreted that there was a linear relationship in the population.

2.4.4. AREA UNDER THE RECEIVER-OPERATING CHARACTERISTICS CURVE

The receiver-operating characteristics (ROC) curve is a graph of the true positive fraction (sensitivity) versus the false positive fraction (1-specificity) for continuous threshold values. A value of the area under the curve (AUC) of 1 is considered ideal, while a value of 0.5 is equivalent to random chance [131]. ROC analysis was conducted to determine if radiomics features could predict for clinical endpoints.

These statistical methods will be utilised in the forthcoming studies, commencing with a phantom analysis of the reproducibility of radiomics features under different image acquisition parameters in Study 1.

CHAPTER 3. STUDY 1: REPRODUCIBILITY OF RADIOMICS FEATURES - A PHANTOM ANALYSIS.

3.1 INTRODUCTION

Feature values extracted in radiomics analysis are dependent on scanner factors, including scanner type, voxel and matrix sizes and resolution [132]. In any biomarker research, a pre-requisite is that its value should remain stable and repeatable under standard conditions, as well as being reproducible under varying conditions. In radiomics research, this means under typical scanner noise [122]. Repeatability means comparing feature values under constant conditions whereas reproducibility means comparing values under varying conditions [116]. Features extracted should accurately represent tumour heterogeneity, not changes in scanning parameters [133]. Much research has focused on the impact of inter-observer and intra-observer variability in ROI delineation in the field of radiation oncology, as summarised by Vinod et al [90]. In radiomics, ROI delineation impacts on all subsequent steps downstream and in MRI radiomics an additional issue is the impact of geometric distortion on delineation. Geometric distortion in MR imaging arises mainly from gradient field non-linearity and magnet field ('B_0' field) inhomogeneity [134]. The former is not corrected for in the slice direction by MR manufacturers' software in most 2D or slice-selective pulse sequences and the latter can be further differentiated into those due to residual inhomogeneity in the magnetic field or 'hardware' distortions and those induced by the patient [135]. Modern scanners further compromise distortion through use of a shorter bore magnet and faster and stronger gradient systems to improve patient comfort [136].

Use of phantoms in radiation therapy research provides the basis for 'ground truth' measurements [137]. By restricting the number of experimental variables, which is not possible in a patient model, phantoms can simplify the research question [138]. Phantoms are routinely utilised in radiation therapy research to answer questions pertaining to developments in radiation therapy, such as the spatial accuracy of MR images and accumulation of dose in adaptive radiation therapy [139].

Phantoms can be made from many materials, including polymethylmethacrylate (PMMA) [140], polyvinyl alcohol cryogel (PVA-C) [138, 141, 142] and gadolinium-

infused ballistic gel [143, 144]. Phantoms for use in MRI systems have previously been constructed from water-based agarose gels, together with copper salts, varying their concentrations to modify T2 and T1 relaxation times [145]. Phantoms created specifically for DW-MRI research have utilised solutions of polyvinylpyrrolidone (PVP), an organic polymer, whose concentrations can be varied to produce specified ADC values [146].

Additive manufacturing, or 3D printing, has become increasingly affordable in recent years, offering new opportunities for the creation of phantoms for analysis of many aspects of radiation therapy quality assurance, as well as, more recently, radiomics. The accuracy of 3D printed models is affected by the sum of errors introduced in each step of the process from their creation from medical images. These steps include imaging, ROI definition and any subsequent post-processing of the ROI in the stereolithography (STL) file used for printing [147]. With all 3D printers, layers of material are successively added on top of the previously printed layers. The resolution of most 3D printers typically refers to the print layer thickness only (z-axis) and range from 0.05 mm to 0.3 mm. The most commonly used 3D printing technologies in medicine are binder jetting, where a liquid adhesive is jetted onto a bed of ceramic or gypsum power, powder bed fusion with stereolithography, direct metal laser sintering and electron-beam melting, material extrusion with fused deposition modelling, vat photopolymerisation with selective laser sintering and material jetting. The latter is the technology used in this study as it is known to be most versatile for anatomic models, although vat photopolymerisation and material jetting use the same photoreactive chemistry [147]. In material jetting, droplets of epoxy or acrylic-based liquid photopolymers are jetted onto a tray and polymerised and solidified by ultraviolet light, also known as Fused Deposition Modeling (FDM) [148]. A common material used in material jetting is the high density ($1040 kg/m^3$) acrylonitrile butadiene styrene (ABS), which is also MRI compatible as it contains non-magnetic material [149] and prints a so-called 'negative' mould that can subsequently be filled with a substance with a high concentration of water molecules, such as agarose [147] or silicone-based polymers [150]. One drawback of ABS technology is the tendency for the material to become brittle over time. The accuracy of 3D printing, that is when the STL file is compared to the final printed model, is typically better than 0.5 mm, which is comparable to the spatial resolution of most imaging modalities.

Creation of a 3D printed phantom for MR scanning is challenging as it requires both an anatomically correct phantom from clinical images but also the incorporation of materials that will possess a signal on MRI [151]. Materials such as nylon and gypsum are not visible on MRI, but there is the option to introduce gadolinium during the printing process to overcome this issue [150]. In this study the phantom was 3D printed using ABS.

3.1.1 AIMS AND OBJECTIVES

The aims of this chapter are:

1. To create an MRI-compatible 3D printed pelvic phantom.
2. To determine the impact of geometric distortion on the size of prostate regions of interest.
3. To ascertain which radiomics features are repeatable when matrix size, field of view and slice thickness on T2-weighted MRI are changed.

The objectives of the chapter are:

1. To measure simulated individual lesions in a prostate phantom and compare their size relative to their distance from the isocentre and field position.
2. To calculate the coefficient of variation for radiomics features extracted when scan parameters are altered.

3.2 MATERIALS AND METHODS

The construction of the phantom has been detailed in Chapter 2. In order to create a clinically realistic prostate volume within the phantom, analysis of prostate volumes in previous patient cohorts was performed.

3.2.1 PROSTATE VOLUME SIMULATION

Patient cohort

Eleven intermediate or high-risk patients were selected from an irrevocably anonymised database of 38 patients in the Discipline of Radiation Therapy, who had been treated with androgen deprivation therapy (ADT) and radiation therapy for prostate cancer. As part of their workup, patients had undergone MRI staging and TRUS-based biopsy. 1.5 Tesla T2-weighted MRIs were available for analysis for prostate gland measurements to be used in the phantom creation. To validate the gland size calculated, measurements were also taken on T2-weighted images of 10 additional prostate patient irrevocably anonymised datasets held in the Discipline of Radiation Therapy, who were scanned at 3.0 Tesla. These patients had not received ADT at the time of scanning.

Prostate gland volume measurements

Measurements for prostate gland volume determination were conducted using T2-weighted images in axial and coronal planes and using the method of the two-ellipsoid model [152]. This model, utilising diameters in both axial and coronal planes, yielded correlation coefficients of 0.95-0.96 with automated segmentation methods.

In the current study, measurements were made on the slice with the largest extension of the prostate at mid-axial and mid-coronal sections (Figures 3.1 and 3.2) and calculated using the formula:

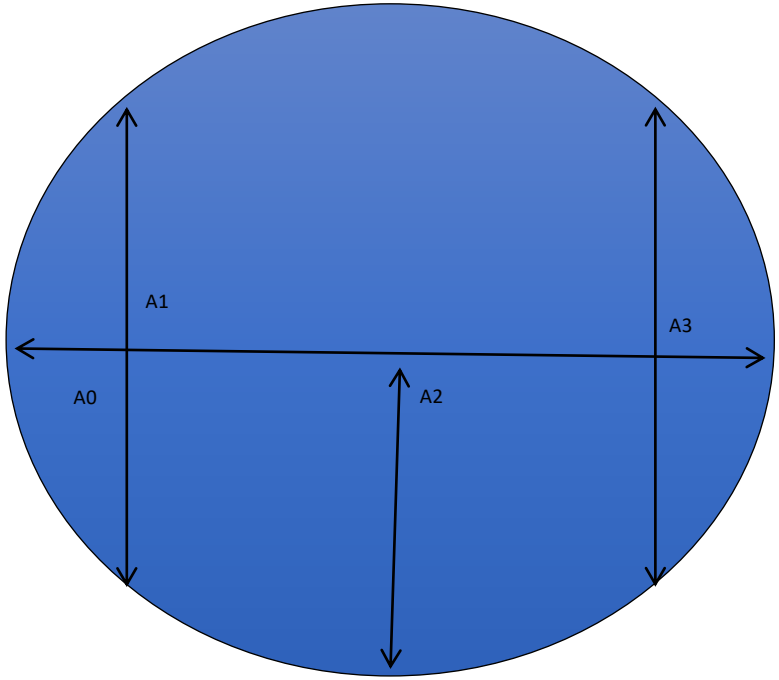
$$V=\alpha*\pi/6*A0+C0/4(A1*C1+A3*C3)+\beta$$

where α= slope of regression model and is 1.1 and β=intercept (cm^3) and is 3.4, as given by Jia et al [152].

The limitations of this method are recognised. Diameter-based models assume that the prostate has a regular ellipsoid shape, which is not always the case clinically and hypertrophied prostates have irregular contours, which are not elliptical. Another drawback of this prostate volume calculation method is the inherent mathematical problem of calculating volume on three linear dimensions whereby a small error in the measurement of one dimension may lead to an amplification of error in the estimate of volume. Despite these accepted limitations, the prostate volume in the study was based not on a single patient, but on a population of patients.

Individual Prostate Lesions

To determine if the size of an individual prostate lesion appreciable on T2-weighted MRI would be affected by geometric distortion because of its relative position both in the phantom and in the field of view, five 1 cc lesions were created in a second phantom insert. 1 cc of de-ionised water was drawn up in each individual syringe and the insert placed in the pelvic phantom, prior to scanning.



A0=longest transverse dimension in the axial images

A1=anterior-posterior diameter in right lobe

A2=anterior-posterior diameter in midline

A3=anterior-posterior diameter in left lobe

Figure 3.1. Prostate gland measurements in axial plane

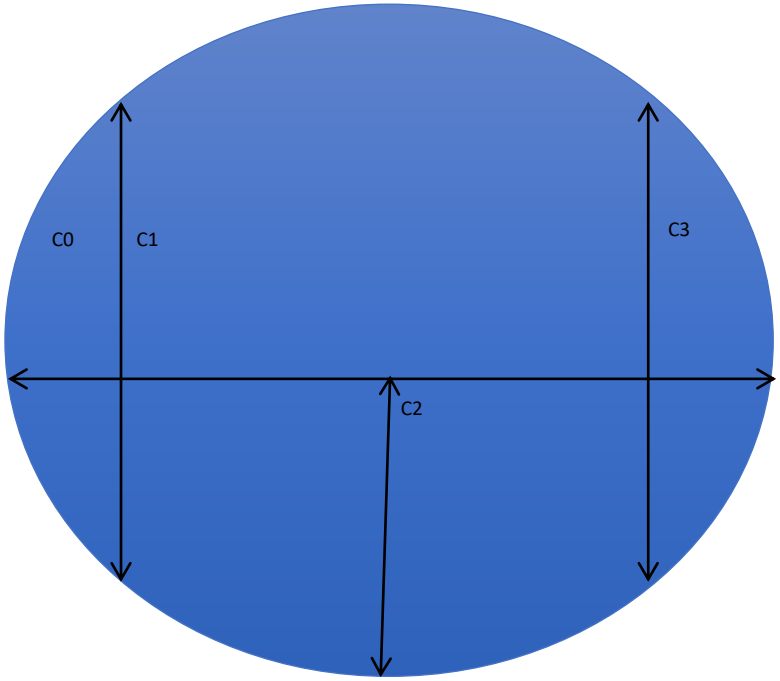


Figure 3.2. Prostate gland measurements in coronal plane

C0=longest transverse dimension in coronal plane

C1=superior-inferior diameter in right lobe

C2=superior-inferior diameter in midline

C3=superior-inferior diameter in left lobe

3.2.2 PHANTOM SCANNING

The phantom was scanned on a 1.5T General Electric (GE) Signa H-DXT MR unit. First, the phantom with the individual lesion inserts was scanned at T2-weighting in the axial plane with the following parameters: TR:5980 ms, TE:120 ms , Matrix:256 X 192, FOV:17 cm^2, Slice thickness: 4 mm and Slice gap: 0.5 mm (Figure 3.3). An eight-channel phased array pelvic coil was used for radiofrequency (RF) signal reception. The same parameters were used as baseline for the first axial scan of the simulated whole prostate gland in the axial plane (Figure 3.4). The frontal view of the phantom, illustrating the inserts and internal supporting architecture, are illustrated in Figure 3.5. T2-weighted scans were also acquired in the axial plane with the matrix changed from 256 X 192 to 320 X 256, then with the FOV changed from 17 cm^2 to 24 cm^2 and finally changed slice thickness from 4 mm to 7mm. Changing both field of view and matrix size impacts on spatial resolution, and these, along with slice thickness impacts on voxel size.

3.2.3 LESION SIZE DETERMINATION

The T2-weighted MRI with individual prostate lesions was imported into OSIRIX LITE™ Digital Imaging and Communications in Medicine (DICOM) viewer. On each slice of the scan, the enhancing lesions X_1 to X_5 were measured in their greatest transverse dimension (Figure 3.6).

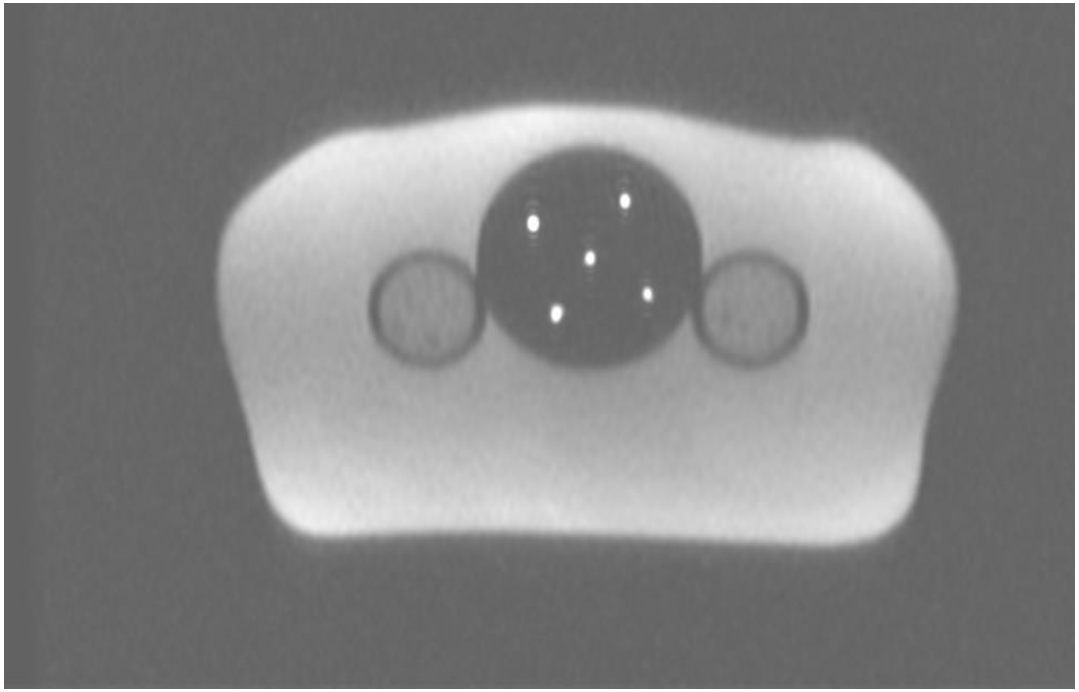


Figure 3.3. Individual lesions scanned at T2 baseline parameters in axial plane

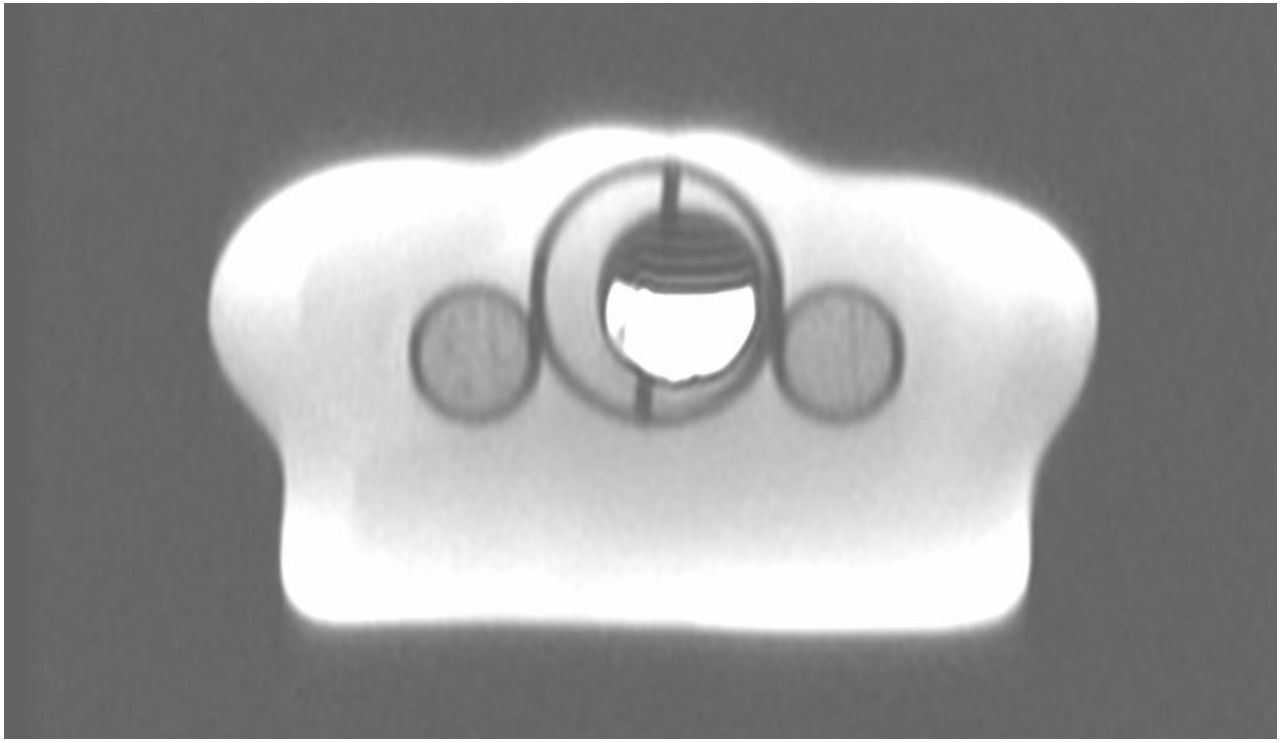
. 

Figure 3.4. Prostate ROI highly enhancing on T2-weighted MRI in axial plane.

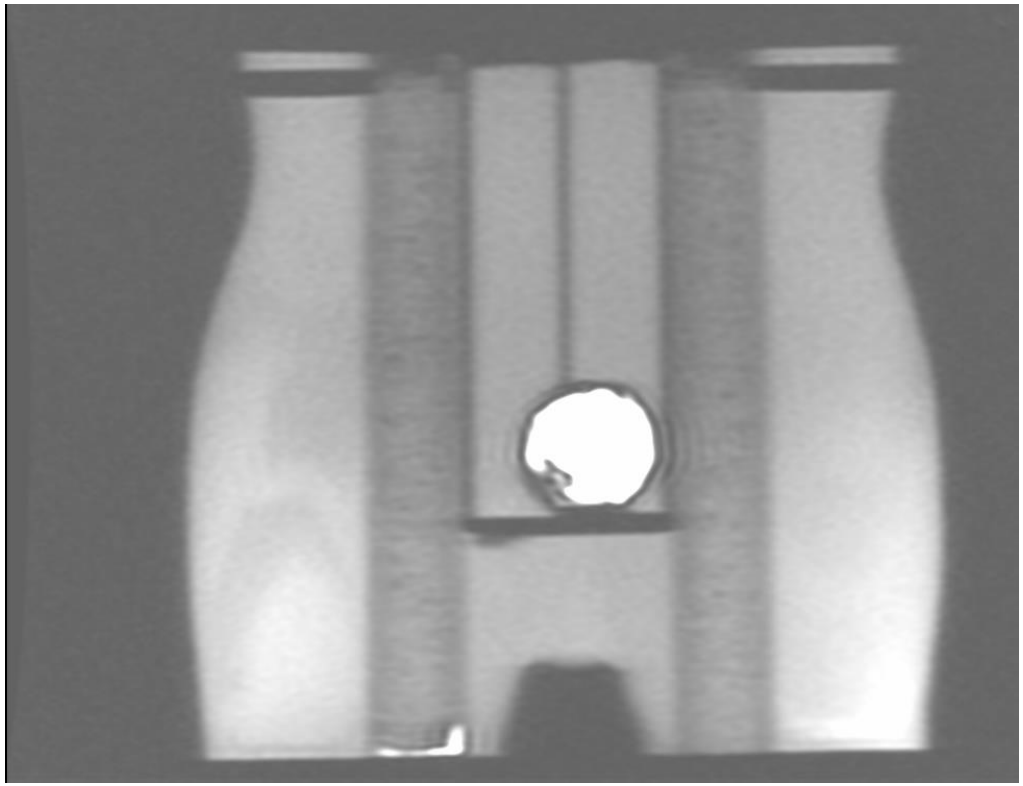


Figure 3.5. Frontal view of prostate phantom. *Note the difference in density between the femoral head inserts (silicone and gypsum) and the surrounding simulated normal tissue (silicone only).*

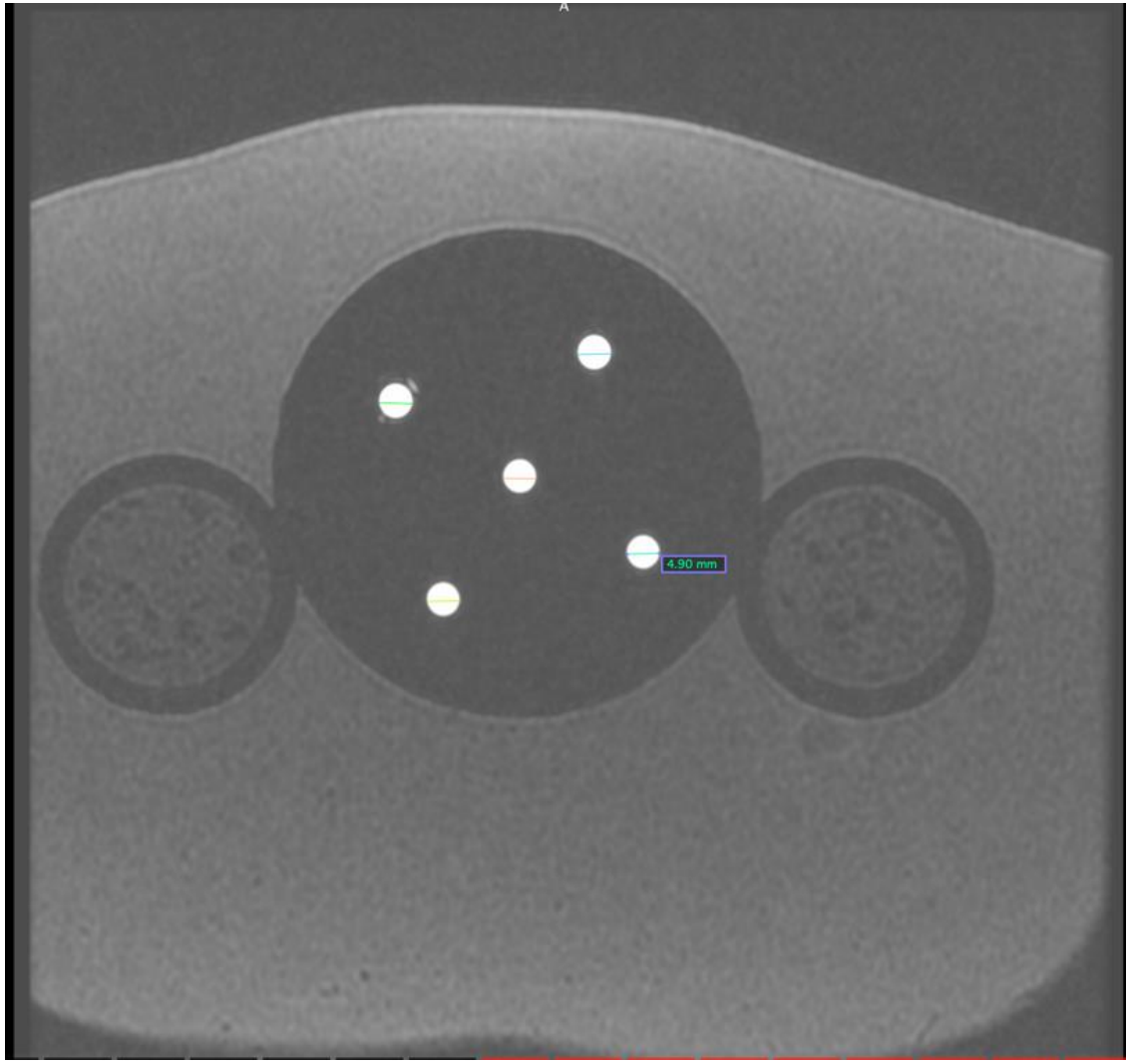


Figure 3.6. Individual lesion measurement on T2-weighted MRI

3.2.4 REGION OF INTEREST DELINEATION

The prostate gland axial images were imported into Varian Eclipse (V 15.5) and the volumes that simulated the prostate delineated as ROIs on the baseline T2-weighted axial scan ($T2_{base}$), the changed matrix scan ($T2_{matrix}$), the changed field of view scan ($T2_{FOV}$) and the increased slice thickness scan ($T2_{slice}$). These ROIs were used for subsequent radiomics analysis.

3.2.5 RADIOMICS FEATURE EXTRACTION

Radiomics feature extraction was conducted on the prostate ROI defined on the 4 T2-weighted MRI scans ($T2_{base}$, $T2_{matrix}$, $T2_{FOV}$ and $T2_{slice}$). The scans were normalised prior to radiomics feature extraction to keep their intensity ranges between -3 and +3 . It has been demonstrated that normalised features are much more stable than unnormalized features [153] and therefore this was kept constant for each changed scanning parameter. This is further explored in Study 2, Chapter 4. The bin width was set to 10 to ensure textural analysis was computed and the linear interpolation algorithm was set for all extractions. No resampling of voxel size was conducted as the parameters under investigation were slice thickness, matrix size and field of view, all of which impact on voxel size. The features extracted were from the radiomics feature families outlined previously, namely, morphological features, local intensity features, first order statistics, intensity histogram features, fractal features and the textural feature families of GLCM, GLRLM, GLSZM, GLZDM, NGTDM and NGLDM.

3.2.6 STATISTICAL ANALYSIS

Thirty-two features, representing each feature family, were randomly selected for reproducibility analysis. The coefficient of variation (COV) was calculated for each of the 32 features using the equation COV= (SD/mean)*100% where the SD is the standard deviation of the feature values measured under each image acquisition setting ($T2_{base}$ and $T2_{matrix}$, $T2_{FOV}$ and $T2_{slice}$) and mean is their arithmetic average.

3.3 RESULTS

3.3.1 ESTIMATION OF PATIENT PROSTATE VOLUMES

The mean prostate gland volume in the 1.5T cohort who had received ADT was 12.44 ccs (SD 3.25). For the comparison 3.0T cohort who had not received ADT, the mean volume was 19.14 ccs (SD 5.68). Results for both cohorts are given in Tables 3.1 and 3.2. As both volumes were small, it was decided to utilise the largest prostate volume calculated in the phantom. Therefore, 19 ml of de-ionised water was pipetted into a balloon and tied off. The balloon was then placed in the cavity for the prostate in the 3D printed insert and secured in the pelvic phantom.

Comparing volumes from cohorts measured using two different field strength MRI units introduces a potential discrepancy in the delineation of the prostate gland between the two cohorts. The increased signal to noise ratio of the 3T images relative to the 1.5T images leads to enhanced spatial resolution ,which may allow for potentially improved accuracy in the delineation of the prostate gland [154]. However, given that the objective was to estimate prostate volume for the purposes of constructing a phantom, it was considered reasonable to use both datasets in this instance.

Pt. ID	A0	A1	A2	A3	C0	C1	C2	C3	Volume (cc)
60060	4.44	1.49	1.37	1.93	4.52	3.01	3.01	3.03	11.306
60061	3.14	1.96	1.88	1.99	2.55	2.37	2.52	2.71	7.496
60062	4.55	3	2.61	3.02	4.88	2.82	2.87	3.07	17.138
60063	4.31	2.05	2.04	2.07	4.71	3.99	3.44	3.93	15.596
60064	4.11	1.56	1.53	1.78	4.08	3.26	3.35	3.55	11.090
60065	4.58	2.51	2.22	2.05	3.25	2.62	2.62	2.91	10.527
60066	3.81	1.84	1.69	1.7	3.86	3.2	2.99	2.7	10.035
60067	4.81	2.3	1.87	2.09	4.07	2.69	3.23	3.07	12.186
60069	5.63	3.02	2.67	2.87	4.17	3.68	3.48	3.18	17.456
60071	4.16	1.82	1.64	1.83	4.18	3.38	3.24	3.13	11.571
Mean Volume									**12.44**
SD									**3.25**

Table 3.1. Measurements from 3.0T T2-weighted MRI for prostate gland size calculations

Pt. ID	A0	A1	A2	A3	C0	C1	C2	C3	Volume (cc)
1020252	4.47	2.25	2.45	2.61	5.65	4.82	4.05	4.34	22.703
1020251	3.91	1.58	1.32	1.36	4.23	2.56	2.85	2.74	8.996
1019912	4.36	2.87	3.18	2.99	4.52	4.39	3.66	4.18	20.926
1019850	5.23	3.03	3.11	3.04	5.32	3.3	3.09	3.63	21.207
1019004	4.01	1.7	1.84	1.69	4.48	3.32	2.95	3	11.245
1019011	4.94	2.28	2.18	2.61	4.79	3.44	3.24	3.07	15.832
1019005	5.42	3.28	3.25	3.2	5.63	3.47	3.27	3.07	22.401
1019018	5.07	3.85	3.82	3.56	5	4.023	3.85	3.62	25.456
1017201	5.68	2.52	2.19	2.72	5.87	4.83	5.13	4.96	27.078
1019851	4.84	2.57	2.33	2.5	4.69	3.19	2.89	3.09	15.590
Mean Volume									**19.14**
SD									**5.68**

Table 3.2. Measurements from 1.5T T2 weighted MRI for prostate gland size calculations.

3.3.2 PROSTATE LESION SIZE ON T2-WEIGHTED MRI

The measurements of individual lesions on each slice of the T2-weighted MRI are given in Table 3.3. Each lesion was labelled as given in Figure 3.7 to describe its location in the phantom, with X_1 simulating isocentre position. Slices 1-2 and 12-13 indicate the edges of the field with slices 6 and 7 indicating the centre of the field. For X_1, the range of measurements was 4.67-4.96, (SD:0.075), for X_2 the range was 4.1-5.15, (SD: 0.26), for X_3 the range 4.44-6.36, (SD: 0.49), for X_4 the range was 4.43-4.95, (SD: 0.13) and for X_5 the range was 4.5-4.9 (SD: 0.1).

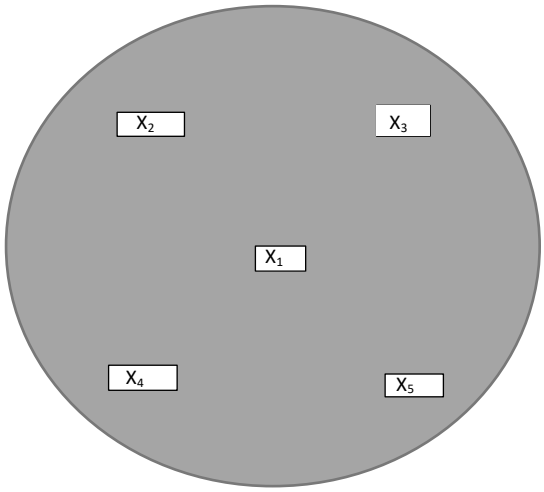


Figure 3.7. Schema indicating individual lesion position within the phantom

Slice	X_1	X_2	X_3	X_4	X_5
1	4.83	4.1	4.44	4.43	4.5
2	4.89	4.8	5.06	4.83	4.8
3	4.75	5	4.86	4.67	4.82
4	4.67	5.11	4.82	4.95	4.86
5	4.8	4.97	4.92	4.79	4.79
6	4.8	5.02	4.94	4.81	4.83
7	4.77	4.93	4.94	4.82	4.9
8	4.96	4.89	4.95	4.75	4.84
9	4.82	5	5.03	4.66	4.86
10	4.83	5.15	4.92	4.81	4.69
11	4.92	4.96	4.83	4.69	4.86
12	4.87	5.03	5.82	4.93	4.81
13	4.82	4.89	6.36	4.8	4.82
SD	0.075	0.26	0.49	0.13	0.10
Range	4.67-4.96	4.1-5.15	4.44-6.36	4.43-4.95	4.5-4.9

Table 3.3. Individual lesion measurements in greatest transverse dimension. *All measurements in mm. SD= Standard deviation*

3.3.3 RADIOMICS REPRODUCIBILITY

The COVs calculated for the 32 radiomics features across $T2_{matix}$, $T2_{FOV}$ and $T2_{slice,}$ relative to T_{base} are given in Table 3.4. The features that were most reproducible (had COV<10%) across all feature families were fractal average, GLCM_average, GLDZM_LISDE, GLRLM_LGRE, GLRLM_RE, NGLDM_SDE, all features calculated from the Shape family, Stats_mean and Stats_rms. Fractal average had the lowest COV across all three scans (0.005 for T_{matrix}, 0.007 for T_{FOV} and 0.010 for T_{slice}. and was therefore the most repeatable.

27/31 (87%) of features had COV<10% following a change in field of view. 29/31 (93.5%) of features had high reproducibility even when the matrix size was increased. Increased slice thickness had a large, negative effect on feature reproducibility with only 12/31 (38.7%) of features with a COV<10%. The features most highly impacted upon were the textural features GLCM_ClusShade, GLSZM_SZN, NGTDM_busyness as well as the feature IH_skewness. The Shape feature family was robust to changes in any of the three parameters, with the COV remaining <10%, despite the change in any of matrix size, FOV or slice thickness.

Feature	T_{FOV}	T_{slice}	T_{matrix}
Fractal Average	0.007	0.010	0.005
Fractal Lacunarity	0.586	***40.961	2.558
GLCM_autocorr	0.677	**15.870	0.043
GLCM_average	0.343	*7.821	0.047
GLCM_clusProm	2.410	***29.833	4.917
GLCM_clusShade	2.744	***103.635	2.370
GLDZM_HIE	*6.298	**19.766	4.126
GLDZM_HISDE	1.036	**15.305	2.878
GLDZM_LISDE	1.918	4.488	1.042
GLRLM_LGRE	*5.414	1.826	2.671
GLRLM_RE	0.553	*6.873	0.030
GLRLM_SRHGE	3.951	**14.683	0.116
GLRLM_SRLGE	*9.673	**14.230	4.176
GLSZM_HIE	*6.298	**19.766	4.126
GLSZM_HISAE	***31.586	**10.122	2.676
GLSZM_IV	**13.537	***39.037	*9.638
GLSZM_SZN	***22.158	***98.564	**10.202
IH_kurtosis	1.495	***60.829	0.966
IH_skewness	**11.311	***110.713	**10.064
NGLDM_GLNN	0.378	***25.971	0.900
NGLDM_HGLDE	0.336	**14.653	0.863
NGLDM_SDE	*2.079	*8.563	0.972
NGTDM_busyness	3.513	***87.189	0.611
Shape_areaDensityBB	0.549	2.421	0.425
Shape_elongation	0.137	1.424	0.214
Shape_flatness	1.764	1.540	0.458
Shape_volume_density_BE	0.773	2.508	0.081
Stats_mean	1.869	0.780	0.177
Stats_rms	1.087	2.228	0.203
Stats_skewness	*5.902	***30.331	3.910
Stats_Var	*9.260	***39.062	0.822

Table 3.4. Coefficient of Variation (COV) (%) of radiomics features when imaging parameters altered

*FOV= Field of View. Results in plain text are ≤5% COV, results denoted by * are >5%, ≤10% COV, results denoted by ** are >10%, ≤20% COV and results denoted by *** are >20% COV.*

3.4 DISCUSSION

3.4.1 GEOMETRIC DISTORTION IN LESION DETECTION

Most patient MR images are expected to be distortion corrected using the vendor's correction algorithms prior to their use in radiation therapy. However, distortion caused by B_0 inhomogeneities and gradient coil nonlinearities is not corrected for [155]. Gradient non-linearities of the x, y and z gradient coils can be separated from other types of spatial distortion using the reverse gradient technique. Here, two images are acquired using the same parameters except for a reversal in the read gradient polarity [156]. As well as the issues pertaining to ROI definition due to geometric distortion in MRI, it has been illustrated that a gradient non-linearity distortion correction can improve voxel-based image intensity of data on brain MRI [157]. However, in clinical practice it is more common to check geometric distortions at a specific plane using a phantom and having a set passing criterion [158].

The practical implications of geometric distortion on ROI delineation, which has a downstream effect on radiomics analysis, were analysed in this study. Prostate lesions were created in a 3D phantom with a fixed volume and their appreciable size on T2-weighted MRI measured and compared, depending on their location both within the phantom and within the field of view. All lesions in the 3D phantom were created using 1 cc of de-ionised water using 1 cc plastic syringes. The lesions were positioned at equidistant points from the central lesion in the phantom. X_1 was at the isocentre and X_2-X_5 were equidistant from X_1 (2.25 cm). However, some of the lesions were closer to the edge of the field of view than others due to the scanning plane (Figure I). The lesion volumes were too small to appreciate on the Eclipse (V. 15.5) treatment planning system, in that it was not possible to compute a volume on a slice by slice basis. They were therefore measured in the maximum transverse diameter on each individual axial slice using OSIRIX LITE™. In Table 3.3, slices 6 and 7 are representative of the centre of the imaging field and the slices 12 and 13 the edge of the field. Overall, the lesions measured across all 13 slices varied in maximum transverse diameter between 4.1-6.36 mm, depending on the location in the phantom and how close to the field edge the lesion being measured was located. The least variation was seen for X_1 , the lesion at the isocentre (4.67-4.96 mm). Although each lesion was

equidistant from the isocentric lesion, some appeared greater enhancing than others, at the edge of the field of view. The most extreme case was X_3, which measured 5.82 mm on slice 12 and 6.36 mm on slice 13, even though it was very stable at 4.94 mm on both slices 6 and 7. Although its measurement is beyond the scope of this project, the likely explanation for these results is the impact of geometric distortion on the appreciable lesion size. In simulated MR geometric distortion in SBRT for liver cancer, it was found that the closer the tumour to the surface of the patient, the larger the margin may need to be considered during target volume delineation, but if the tumour is closer to the centre of the image, compensating for distortion is not necessary [155], which supports the results of the current analysis. Distortions are considered to be minimal at the centre of a closed bore magnet and increase towards the radial edges of the volume being scanned [136]. This is true regardless of scanner manufacturer, field strength or protocol parameter [159]. Distortions of up to 25 mm over a 24 cm field of view for a 1.5T field strength has been reported [156]. A 24 cm FOV would not be unreasonable for a prostate cancer patient undergoing radiation therapy and was utilised in this study as the changed FOV for radiomics analysis of a prostate phantom. It has been clearly stated that MRI users should be cognisant of large geometric distortions when imaging peripheral areas in patients [134]. These results are important not only for radiomics analysis in localised prostate cancer, but in cases where pelvic lymph nodes may also be treated.

Patient-induced distortion, which includes geometric distortions from chemical shift and susceptibility artefacts due to the variations in the magnetic properties of the tissues in the body, could not be considered in this phantom study. However, it has been shown in breast radiation therapy that maximum distortion values were on average 3.8 times greater than those measured in a phantom, indicating that patient-induced distortions are larger than MRI scanner residual distortions [160]. Therefore, the range of measurements observed in this phantom study (4.1-6.36 mm) is likely to be wider in a patient-based analysis. However, patient-based analyses do not offer the opportunity for 'ground truth' measurements, as was possible in this phantom study.

3.4.2 REPRODUCIBILITY OF RADIOMICS FEATURES IN A PHANTOM

Reproducibility of radiomics features under varying scanner parameters, as well as reproducibility under constant conditions, are important aspects of radiomics research, in order to make models generalisable to radiation therapy practice [116]. Phantom or 'ground truth' studies can determine whether individual radiomic features are robust to variations in MR image acquisition and reconstruction parameters [91]. 3D printing of phantoms is quite cost-effective relative to the purchase of commercial phantoms and this technology also offers the possibility to print various geometries [150], as seen in this study and also reported in cardiac medicine research [161]. However, there are some limitations of 3D printed phantoms which must be noted. First, the anatomical model created was limited in terms of accuracy because of the CT slice thickness of 2.5 mm, on which the phantom was modelled. While this is standard for pelvic CT for radiation therapy treatment planning, a slice thickness of 0.625 mm has been reported as preferable for 3D printing [150]. Defining the ROI boundary during definition was at times challenging owing to the motion of de-ionised water volume in the balloon in the prostate insert of the phantom. The partial volume effect also influences ROI definition [150]. Assembling the phantom including support structures for the inserts was challenging and pouring of the silicone into the mould may have resulted in changes of position of some of the inserts. Others have reported similar issues [150]. Very high temperatures are required to extrude the ABS material, meaning that printing a 100% solid object is not possible due to deformation of the plastic upon cooling [161].

The aim of this phantom study was to ascertain which radiomics features were reproducible when field of view, matrix size and slice thickness on T2-weighted MRI were changed. In radiomics analysis, features with low coefficients of variation (COV) are considered potential candidates for reproducible tumour quantification. COVs can be quantified as very small ($\leq$ 5%), small (> 5%, $\leq$ 10%), intermediate (> 10%, $\leq$ 20%) or large (> 20%)[133, 162], or as large (COV > 50%), diminished variation (< 30%) and moderate variation (< 50%) [163].

The baseline T2-weighted parameters used for scanning the prostate phantom in the axial plane were TR:5980ms, TE:120ms, Matrix:256 X 192, FOV:17 cm^2, Slice thickness: 4 mm and Slice gap: 0.5 mm. The scans were repeated 3 times,

changing one parameter each time. First the matrix was changed from 256 X 192 to 320 X 256. The FOV was then changed from 17 cm^2 to 24 cm^2 and finally slice thickness was changed from 4 mm to 7mm. Changes in FOV and matrix size impact on spatial resolution and these, together with slice thickness, impact on voxel size.

The features that were most reproducible (had COV<10%) across all feature families were fractal average, GLCM_average, GLDZM_LISDE, GLRLM_LGRE, GLRLM_RE, NGLDM_SDE, all features calculated from the Shape family, Stats_mean and Stats_rms. Fractal average had the lowest COV across all three scans and was therefore the most reproducible. Fractal features examine tumour structural geometry [8] and fractal analysis has been cited as a predictor of pathological complete response in patients with rectal cancer [125], making it an important parameter in radiomics analysis. Fractal analyses are based on higher order statistical methods, where filter grids are imposed on the image and the number of grid elements containing voxels of a specific value are computed [83]. In a previous MRI phantom study with T2-weighted images at 3.0T, the co-occurrence matrix was found to be the most robust of all texture features analysed when the imaging parameters TR, TE, number of acquisitions (NA) and sampling bandwidth (SBW) were modified (TR 900-4500ms, TE 25-125 ms, NA 1-5 and SBW 26-101kHz) [164].

27/31 (87%) of features had COV<10% following a change in field of view. Spatial resolution has been found to be an important factor to consider for textural feature analysis in a study where repetition time (TR) was altered to represent a change in spatial resolution, and this was independent of field strength [124]. In a CT phantom study, a FOV change from 25 to 55 cm resulted in 71% of radiomics features yielding a COV of <10% and 77.9% a COV <15% [165], which is slightly more conservative than found in the current analysis.

29/31 (93.5%) of features had high reproducibility when the matrix size was increased, which directly affects voxel size. Interestingly, it has been reported that voxel size had a negative effect on feature stability in a virtual brain phantom [166]. In the current analysis, GLSZM_SZN and IH_skewness had lesser reproducibility with a change in matrix size but with COV still <15%. In a positron emission tomography (PET) radiomics study on reproducibility, Leijenaar et al reported similar results in a patient study of non-small cell lung cancer with

features from GLCM and GLRLM families being more reproducible than GLSZM [167]. In a PET phantom study, 56% of all features had COV $\geq$ 20%, with the exceptions being NGLCM Entropy, Intensity, SUVmax, SUV Entropy and SULpeak and GLCM Entropy [133].

Increased slice thickness had a large, negative effect on feature reproducibility with only 12/31 (38.7%) of features with a COV<10%. The features most highly impacted upon were the textural features GLCM_ClusShade, GLSZM_SZN, NGTDM_busyness, as well as the feature IH_skewness. It has been found previously that another member of the GLSZM family, GLSZM_SAE was highly influenced by slice thickness and was unstable in a test-re-test analysis on CT in lung cancer [168]. Lu et al [169] examined the effect of three different slice thicknesses (1.25 mm, 2.5 mm and 5 mm) and two reconstruction algorithms (lung and standard) in CT radiomics of a 32 patient lung cancer dataset. Inter-setting agreement was measured by the concordance correlation coefficient (CCC). The best result was found comparing 1.25 mm and 2.5 mm slice thickness using the same reconstruction algorithm and the worst was with a comparison of the 1.25 mm slice thickness and the lung algorithm compared to the 5 mm slice thickness and the standard algorithm, having an average CCC< 0.5. Textural feature families such as GLCM showed the lowest agreement levels. Although it can be argued that increased slice thickness decreases image noise, larger partial volume effects and poor spatial resolution in the axial direction are also caused by larger slice thickness and can blur images, diminishing textural detail [115], which was observed in the results of the current analysis. In cases where images have been obtained using large slice thickness, resampling to obtain isotropic voxels is cautioned against in radiomics analysis and a slice by slice analysis is recommended instead to prevent inaccuracies [8].

The Shape feature family was robust to changes in any of the three parameters, with the COV remaining <10%, despite the change in any of matrix size, FOV or slice thickness. This was also reported by Fiset et al in their study on repeatability and reproducibility of MRI-based radiomic features in patients with cervical cancer, where test-re-test simulation MRIs, repeat MRIs within 8 days and interobserver variation in contouring were analysed. Shape features were the most stable across the three cohorts [116]. Shape features from CT were also found to be the most reproducible together with the GLCM feature family in the

Reference Image Database to Evaluate Therapy Response (RIDER) lung cancer dataset and a clinical dataset of 40 rectal cancer patients under test-retest conditions at varying timepoints [170]. A similar result on the reproducibility of Shape features was obtained by Zhao et al, also in lung cancer on CT images that later became part of the RIDER dataset [115].

3.4.3 CONCLUSION

In this chapter, a 3D printed pelvic phantom with two inserts, one for simulation of the entire prostate and one for individual prostate lesions, was created. Using the prostate lesion insert, the impact of geometric distortion on the appreciable lesion size was demonstrated, with lesions close to the centre unaffected by distortion and those on the periphery appearing larger at the edge of the field of view compared with those at the centre of the field. As radiomics feature extraction is dependent on ROI definition, this should be considered in future MRI radiomics analyses.

Using the whole prostate gland insert, the scanning parameters matrix, field of view and slice thickness were independently changed during T2-weighted scanning. Overall, Fractal average, all Shape features and GLCM_average, GLDZM_LISDE, GLRLM_LGRE, GLRLM_RE, NGLDM_SDE, Stats_mean and Stats_rm were the most reproducible features, with reproducibility significantly decreased with slice thickness, relative to increases in matrix or field of view sizes.

CHAPTER 4. STUDY 2: IMPACT OF IMAGE PRE-PROCESSING ON REPRODUCIBILITY OF RADIOMICS FEATURES - A PATIENT ANALYSIS

4.1 INTRODUCTION

Radiomics features have potential as image biomarkers both in the diagnosis and treatment of prostate cancer [171]. Radiomics features that illustrate high prediction value should also be highly repeatable and reproducible in order to build a clinically usable radiomics signature [89]. Set standards for radiomics feature extraction in CT and PET have been proposed by the Image Biomarker Standardisation Initiative [86]; however, to date there is none set for feature extraction in MRI. There are a vast number of parameters that can vary in MR radiomics analysis. MR image intensities vary depending on voxel size, pulse sequence, vendor, reconstruction algorithm and voxel size [88]. The values of textural features in particular have been shown to change their values when computed using various quantisation methods [172]. Textural features describe the pattern distribution of voxels and thereby measure intra-tumour heterogeneity and include second order and higher features [173]. Second order features (eg. GLCM or 'Haralick's feature family) measure relationships between direct neighbouring voxel intensity values, while higher order features (e.g. GLDZM, GLSZM and NGTDM feature families) quantify regional relationships between voxels. First order features (e.g. Stats feature family) measure the distribution of intensity values [174].

Feature extraction is not only influenced by the selection of imaging parameters and pre-processing but also by random noise, which can alter the image metrics and change the relationship between voxels [175]. Noise varies between manufacturers, scanners and institutions and as a result, resampling of voxels, may be necessary to reduce image noise [162]. Normalisation in MR images corrects the non-parametric intensity nonuniformity and limits grey level analysis to the range of mean $\pm$ 3 (standard deviation) [165]. Radiomics feature extraction without normalisation in MR will therefore hypothetically consider a much larger range. Interpolation algorithms translate image intensities from the original image grid to an interpolation grid, where voxels are spatially represented by their centres. The nearest neighbour algorithm assigns the intensity of the most nearby voxel in the original grid to each voxel in the

interpolation grid and has been cited as leading to 'blockiness', and bias in textural features [86].

4.1.1 AIMS AND OBJECTIVES

The aims of this chapter are

1. To calculate the reproducibility of radiomics features in patient cohorts across MR image sequences when image pre-processing parameters are varied.
2. To make recommendations about which features are most robust to changes in image pre-processing.

The objectives are:

1. To change the image pre-processing parameters of normalisation, resampling and interpolation algorithm independently prior to extraction of 32 radiomics features on T2-weighted MRI, DW-MRI and ADC maps.
2. To calculate intraclass correlation coefficients (ICC) for each changed feature relative to its baseline.
3. To determine which features are most repeatable across the 3 MR sequences.

4.2 METHODS

Three MRI patient datasets (T2-weighted, DW-MRI and ADC maps) were utilised in this study to determine the reproducibility of radiomics features across MRI sequences when pre-processing parameters were changed.

4.2.1 T2-WEIGHTED DATASETS

The T2-weighted datasets analysed were acquired using the T2 **P**eriodically **R**otated **O**verlapping **P**aral**lEL L**ines with **E**nhanced **R**econstruction (PROPELLER) technique. The PROPELLER or BLADE technique uses radial k-space sampling with parallel data lines rotating around the centre of k-space at each time of repetition, leading to oversampling, which can correct for phase, rotation, translation and weighting to reduce spatial inconsistencies. Signal-to-noise ratio

is increased and there is a significant reduction of artefacts relative to turbo spin echo (TSE) techniques due to the correction of phase-encoded aberrations [176]. PROPELLER imaging parameters of the scan protocol are given in Table 4.1. T2-weighted images for 114 patients were available. There were errors in image transfer in 5 patients, an error in reading the DICOM series in 3 patients and the ROI was undefined in 2 patients. Radiomics-specific issues included an error in constructing the binary structure array in 4 patients, in calculation of shape features for 7 patients and in calculation of intensity features for 5 patients. Therefore the final T2-weighted radiomics analysis was conducted on 88 patient datasets.

	T2 PROPELLER
TR(ms)/TE (ms)	4144.87/141.76
Flip Angle °	0/90
Acquired matrix (mm)	512 X 512
Slice thickness (mm)	3
Voxel size (mm^3)	1.5

Table 4.1. T2 image scanning parameters. *TR= Time of repetition, TE= Echo time.*

4.2.2 DW-MRI AND ADC DATASETS

DW-MRI was performed with *b* values of 0, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 and 1500 s/mm^2. The DW images utilised in this analysis were acquired with a *b* value of 1500 s/mm^2. The patients were placed as close to the magnet isocentre as possible to minimise bias in the DW-MRI parameters caused by potential gradient nonlinearity [177]. As described by Hompland et al [120], the ADC maps were developed using in-house software developed with Matlab (MathWorks). ADC was calculated for each pixel by fitting the equation $SI=SI_0$ X exp (-*b* X *D*) to plots of ln $(S(b)/S_0)$ versus *b*, using a linear least squares algorithm. The degree of signal attenuation *SI*, depends on the signal intensity of the T2-weighted image with no diffusion gradient applied SI_0, the degree of diffusion weighting known by the *b value* and the diffusion coefficient *D*. The b values used were from 200 to 800 s/mm^2, based on the linearity analysis of ln $(S(b)/S_0)$. Goodness of fit was determined by Pearson Correlation coefficients with pixel with $R^2 < 0.9$ being excluded. 3 patients were excluded as poor fitting occurred for more than 20% of the tumour pixels and 5 patients were excluded because of motion during the DW-MRI acquisition. There were therefore 106 patient datasets available from the FUNCTPROST study available for radiomics analysis.

For the DW-MRI analysis, there were errors in image transfer in 1 patient and an error in reading the DICOM series in 1 patient. Therefore the final DW-MRI radiomics analysis was conducted on 104 patient datasets. For the ADC analysis, there were errors in image transfer in 3 patients. Radiomics-specific issues included an error in constructing the binary structure array in 3 patients and in calculation of shape features for 6 patients. Therefore the final ADC radiomics analysis was conducted on 94 patient datasets.

4.2.3 RADIOMICS FEATURE EXTRACTION

As discussed in Chapter 2, the index tumour was the ROI on which feature extraction was conducted and this was delineated on a single axial slice by an experienced radiologist using the histological specimen as a guide on T2-weighted images, ADC maps and DW-MR images. Examples of index lesions delineated in various anatomical locations within the prostate gland for T2 weighted images are seen in Figure 4.1 and for ADC maps and DW-MRI in Figure 4.2.

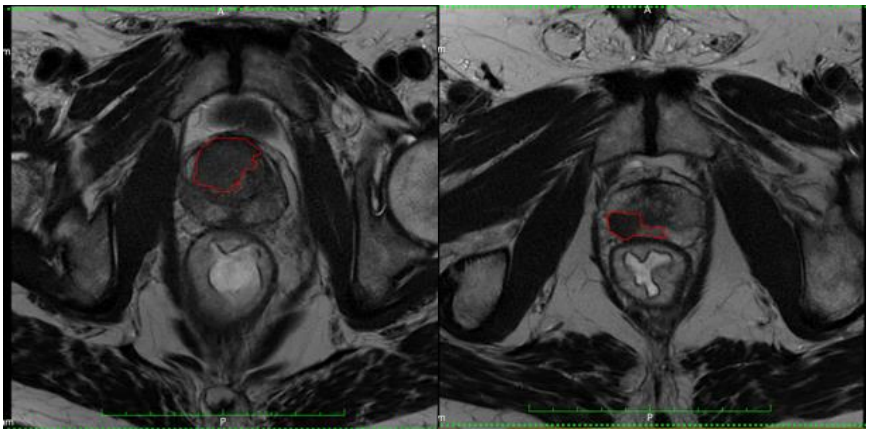


Figure 4.1. Index lesions delineated on T2-weighted sequences in the transition and peripheral zones.

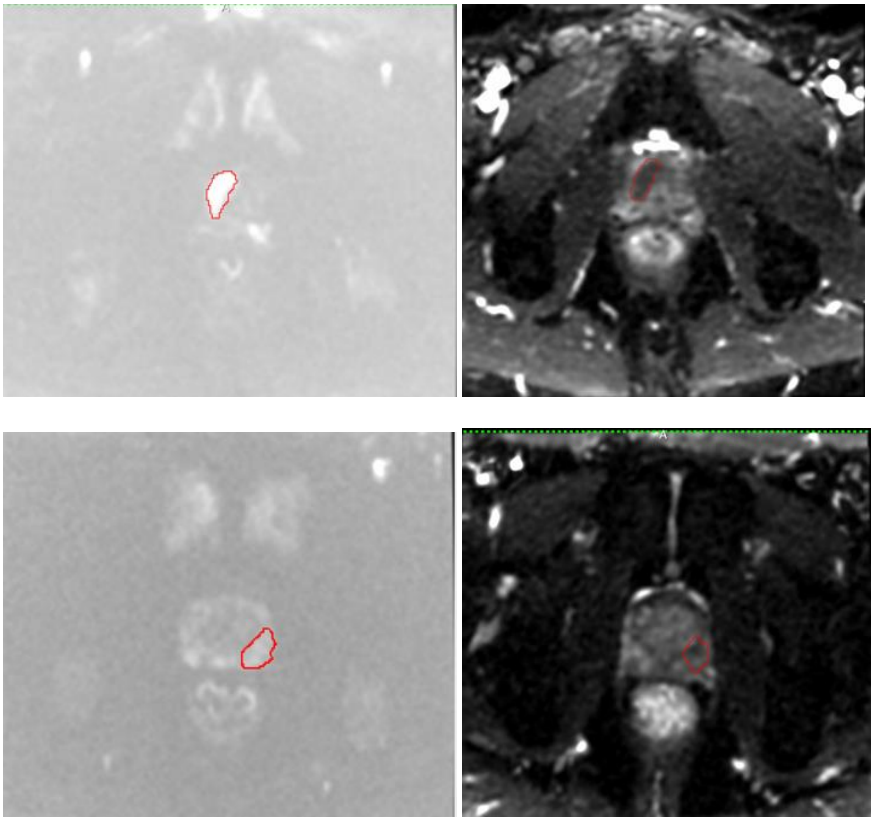


Figure 4.2. Index lesions delineated on DW-MRI (left panels) and ADC (right panels).

33 radiomics features were extracted in this study from the following feature families: Fractal (n=2), GLCM (n=4), GLDZM (n=3), GLRLM (n=4), GLSZM (n=4), IH (n=4), NGLDM (n=3), NGTDM (n=1), Shape (n=4), Stats (n=4). These features were the same as those of the phantom study in Study 2.

First, all MRI sequence datasets were pre-processed as follows:

1. Normalised to zero mean unit variance
2. Resampled to 3 mm isotropic voxels with slice spacing 3 mm
3. Bin width was selected as 0.1
4. Linear interpolation algorithm was utilised

For each dataset, features extracted using these parameters were denoted as "baseline" for that sequence against which the features with changed pre-processing parameters were compared.

In the first change, all pre-processing parameters were kept the same, except normalisation of images was not performed. In the second change, all parameters were kept constant but resampling of voxels was not performed. In the final change, all parameters were kept as at baseline, except nearest neighbour interpolation algorithm was utilised instead of linear. See Figure 4.3 for a summary of methods. Features were extracted on an evaluation version of RadiomiX (RadiomiX Research Toolbox version 20180831(OncoRadiomics SA, Liège, Belgium)) for non-clinical use.

Feature results were exported to SPSS Version 26 (IBM Corp. Released 2019. IBM SPSS Statistics for Macintosh, Version 26.0. Armonk, NY: IBM Corp) and intraclass correlation coefficients (ICC) were calculated to determine the reproducibility of each feature following each pre-processing change, relative to baseline for each MR sequence. In essence, the ICC calculates the correlation between the feature values at baseline and the feature values for the changed parameter and its calculated value represents the strength of the agreement. While there are no standard values for acceptable reliability using ICC, values < 0.5 are considered poor agreement, between 0.5 and 0.75 moderate agreement, between 0.75 and 0.9 good agreement and >0 .9 excellent agreement [178]. There are 10 different classes of ICC [179]. The class used in this analysis was the two-way random effects model with absolute agreement, as utilised in other Radiomics analysis [119] and given by the formula:

$$\mathbf{MS_R\text{-}MS_E/MS_R + (k\text{-}1)MS_E + \mathit{k}/\mathit{n}(MS_C\text{-}MS_E)}$$

Where MS_R=mean square for rows, MS_E=mean square for error, MS_C=mean square for columns, *k*= number of measurements, *n*=number of subjects.

Absolute agreement was selected instead of linear agreement as each feature value extracted with and without pre-processing was directly compared with one another.

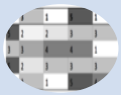
(a) Feature extraction with baseline pre-processing parameters	(b) Change in pre-processing parameters	c) Feature extraction with changed pre-processing parameters	(d) ICC calculation for repeatability
 First Order features  Texture features	1. Normalisation not performed 2. Resampling of voxels not performed 3. Nearest neighbour interpolation algorithm replaces trilinear	 First Order features  Texture features	Two-way random effects, absolute agreement <0.5: Poor 0.5-0.75: Moderate 0.75-0.9: Good >0.9: excellent

Figure 4.3. Summary of methods for ICC reproducibility determination

4.3 RESULTS

4.3.1 PATIENT COHORTS

The clinical features of analysed T2-weighted, DW-MRI and ADC patient datasets are given in Appendices 1-3 respectively. The image datasets came from an overall patient cohort of 114 prostate cases. Therefore, there is overlap of patients included in the three image cohorts. In some instances, a particular sequence for a given patient may not have been available for analysis, hence the patient cohorts are described per MR sequence.

One patient had an initial prostate specific antigen level of 145 ng/mL, but was concluded not to have metastatic disease on bone scan. Therefore, this patient was included in the analysis.

To determine if there was a relationship between patient characteristics, the T2-weighted cohort was also stratified according to presence or absence of hypoxia (PIMO score ≥ 3 and PIMO score <3, respectively) and this was cross-tabulated with clinical T stage, (Table 4.2), Gleason score at biopsy (Table 4.3) and PSA threshold of <10 and $\geq$10 ng/mL (Table 4.4). 29 patients with PIMO score < 3 had a stage of T1c but no other clear patterns were observed. For example, 15

Gleason score 8 patients had PIMO score <3 while only 9 had PIMO score of 3 or above.

T2-weighted Patient Characteristics

The mean age was 64 years (range 47-75 years). The mean pre-op PSA was 12.4 (range 2.2-145). The mean Gleason score at biopsy was 7 (range 6-9). The Clinical (c) T stages were as follows: cT1c (n=37), cT2a (n=8), cT2b (n=8), cT2c (n=4), cT3a (n=7), cT2x (n=14) and no cT stage was recorded for 5 patients. 2 patients were classified as low risk, 47 as intermediate and 39 as high risk as per the D'amico risk classification. The mean PIMO score was 1.9 (range 0-5) and 10 patients in this cohort had not received PIMO, as discussed in Chapter 2.

DW-MRI Patient Characteristics

The mean age was 64 years (range 47-76 years). The mean pre-op PSA was 11.9 (range 2-145). The mean Gleason score at biopsy was 7 (range 6-9). The Clinical (c) T stages were as follows: cT1c (n=44), cT2a (n=8), cT2b (n=8), cT2c (n=7), cT3a (n=9), cT2x (n=15), cT3x (n=7) and no cT stage was recorded for 6 patients. 3 patients were classified as low risk, 52 as intermediate and 49 as high risk as per the D'amico risk classification. The mean PIMO score was 2.7 (range 0-5) and 12 patients in this cohort had not received PIMO.

ADC Patient Characteristics

The mean age was 64 years (range 45-75 years). The mean pre-op PSA was 13.3 (range 2-145). The mean Gleason score at biopsy was 7 (range 6-9). The Clinical (c) T stages were as follows: cT1c (n=37), cT2a (n=9), cT2b (n=9), cT2c (n=6), cT3a (n=9), cT2x (n=13), cT3x (n=4) and no cT stage was recorded for 6 patients. 2 patients were classified as low risk, 46 as intermediate and 45 as high risk as per the D'amico risk classification. The mean PIMO score was 2.8 (range 0-5) and 11 patients in this cohort had not received PIMO.

4.3.2 INTRACLASS CORRELATION COEFFICIENTS

ICCs and their 95% confidence intervals for each feature across all MR sequences are given in Tables 4.5-4.7. The purpose of reporting 95% confidence intervals for ICCs is to accurately interpret the size of the heterogeneity [180].

Figure 4.4 illustrates the ICC for all features when normalisation was not performed. Figure 4.5 illustrates the ICC for all features when resampling was not performed and Figure 4.6 depicts the ICC when the interpolation algorithm was changed from linear to nearest neighbour.

4.3.2.1 RESAMPLING OF VOXELS

Where resampling was not performed prior to feature extraction, reproducibility across all features for ADC maps was poor with 1/33 (3.03%) features having good reproducibility (Stats_var) and 3/33 (9.09%) having moderate reproducibility (Shape_elongation, Stats_mean and Stats_rms). For DW-MRI, the reproducibility was more promising. 2 GLCM features, 1 GLRLM feature, 1 GLSZM, 1 IH and 1 Stats feature showed moderate reproducibility (18.18%), while 1 GLSZM, 1 Shape and 1 Stats feature showed moderate reproducibility (9.09%). 2 Stats features (6.06%) showed excellent reproducibility with ICC > 0.9 (Stats_mean and Stats_rms). Similar to ADC analysis, the reproducibility of T2-weighted features when resampling was not performed was also limited. GLCM_ClusShade and Stats_var were the only features with moderate reproducibility (6.06%), while Stats_mean and Stats_rms (6.06%) showed excellent reproducibility, each with an ICC of 0.98.

4.3.2.2 IMAGE NORMALISATION

No feature was repeatable on ADC maps when normalisation of ADC maps was not performed prior to radiomics feature extraction. In fact, without normalisation, limited discrimination between features was observed. For DW-MRI, only Fractal Average illustrated a moderate agreement without prior normalisation (3.03%). For T2-weighted MRI, 2 IH features had good agreement (IH_kurtosis and IH_skewness) (6.06%), 3 Shape features had perfect agreement and 1 Stats feature also had perfect agreement (ICC=1) (12.12%).

4.3.2.3 INTERPOLATION ALGORITHM

When the nearest neighbour interpolation algorithm was utilised in feature extraction, reproducibility between features was generally high across all sequences and many feature families. For the ADC dataset, 8/33 (24.24%) of features demonstrated moderate reproducibility (Fractal lacunarity, GLSZM_IV,

IH_qcod, IH-skewness, NGLDM_SDE, Shape_Area_Density_BB, Stats_skewness and Stats_var). 13/33 (39.39%) features had good reproducibility (Fractal average, GLDZM_HIE, GLDZM_HISDE, GLDZM_LISDE, GLRLM_LGRE, GLRLM_SRGHE, GLRLM_SRLGE, GLSZM_HIE, GLSZM_HISAE, GLSZM_SZN, NGLDM_HGLDE, NGLDM_GLNN, Shape_elongation). 7/33 features (21.21%) showed excellent agreement between interpolation algorithms (GLCM_average, GLRLM_RE, NGTDM_busyness, Shape_flatness, Shape_volume_Density_BE, Stats_mean, Stats_rms).

For the DW-MRI analysis, 9/33 (27.27%) showed moderate reproducibility (Fractal average, GLCM_ClusShade, GLDZM_LISDE, GLRLM_RE, GLRLM_SRLGE, GLSZM_HISAE, IH_kurtosis, Shape_Area_Density_BB, Shape_elongation) and 10/33 (30.30) showed good agreement (Fractal lacunarity, GLCM_auto, GLCM_average, GLCM_ClusProm, GLDZM_HIE, GLDZM_HISDE, GLRLM_SRGHE, GLSZM_HIE, GLSZM_IV and GLSZM_SZN).

T2-weighted MRI analysis had moderate reproducibility in 4/33 features (12.12%) (GLRLM_RE, GLSZM_HISAE, Stats_skewness, Stats_Var and good reproducibility in 8/33 (24.24%) (Fractal average, Fractal lacunarity, GLCM_ClusShade, GLCM_ClusProm, GLDZM_LISDE, GLSZM_SZN, Shape_volume_Density_BE and Stats_rms). Excellent agreement between features extracted with the nearest neighbour algorithm compared with the linear algorithm was observed in 8/33 features on T2-weighted MRI (24.24%) (GlCM-autocorr, GLCM_average, GLDZM_HIE, GLDZM_HISDE, GLRLM_SRGHE, GLSZM_HIE, GLSZM_IV and Stats_mean).

PIMO * T stage Crosstabulation									
		Tstage							Total
		T1c	T2a	T2b	T2c	T3a	T3x	T2x	
PIMO	<3	29	2	3	4	3	2	12	55
	3 and above	7	5	3	3	4	1	3	26
Total		36	7	6	7	7	3	15	81

Table 4.2. PIMO stratification and clinical T stage

PIMO * Gleasons core Crosstabulation							
		Gleasonscore					Total
		Gleason 6	Gleason 3+4	Gleason 4+3	Gleason 8	Gleason 9	
PIMO	<3	4	25	14	15	4	62
	3 and above	3	10	3	9	1	26
Total		7	35	17	24	5	88

Table 4.3. PIMO stratification and Gleason score at biopsy.

PIMO * PSATHRESH Crosstabulation				
		PSATHRESH		Total
		<10	10 and above	
PIMO	<3	37	25	62
	3 and above	16	10	26
Total		53	35	88

Table 4.4. PIMO stratification and PSA thresholding <10 ng/mL

Radiomics Feature	ADC ICC	DW-MRI ICC	T2-weighted ICC
Fractal Average	-0.37 (-0.134, 0.082)	0.032 (-0.035, 0.125)	0.111 (-0.099, 0.311)
Fractal Lacunarity	0.082 (-0.125, 0.282)	0.214 (0.032, 0.385)	0.242 (-0.063, 0.495)
GLCM_auto	0.034 (-0.135, 0.211)	*0.542 (-0.28, 0.789)	0.118 (-0.074, 0.338)
GLCM_average	0.079 (-0.83, 0.249)	*0.602 (-0.086, 0.857)	0.124 (-0.058, 0.385)
GLCM_ClusShade	0.066 (-0.133, 0.262)	0.408 (0.234, 0.556)	*0.506 (0.271, 0.67)
GLCM_ClusProm	0.054 (-0.145, 0.250)	0.291 (0.090, 0.466)	0.371 (0.152, 0.55)
GLDZM_HIE	0.358 (0.077, 0.567)	-0.32 (-0.176, 0.127)	0.195 (-0.086, 0.493)
GLDZM_HISDE	0.037 (-0.121, 0.206)	-0.32 (-0.176, 0.127)	0.195 (-0.086, 0.493)
GLDZM_LISDE	0.085 (-0.061, 0.244)	-0.54 (-0.173, 0.085)	0.062 (-0.053, 0.214)
GLRLM_LGRE	0.093 (-0.60, 0.256)	0.146 (-0.-48, 0.335)	0.031 (-0.044, 0.129)
GLRLM_RE	0.299 (-0.91, 0.594)	0.238 (-0.065, 0.58)	0.17 (-0.055, 0.481)
GLRLM_SRGHE	-0.36 (-0.212, 0.150)	*0.576 (-0.006, 0.809)	0.145 (-0.081, 0.39)
GLRLM_SRLGE	0.010 (-0.93, 0.132)	0.145 (-0.067, 0.354)	0.002 (-0.049, 0.07)
GLSZM_HIE	-0.27 (-0.182, 0.142)	-0.11 (-0.299, 0.086)	0.195 (-0.086, 0.493)
GLSZM_HISAE	0.008 (_0.180, 0.2)	*0.697, (0.219, 0.860)	0.314 (-0.086, 0.606)
GLSZM_IV	0.321 (0.123, 0.494)	**0.809 (0.369, 0.919)	0.496 (-0.081, 0.809)
GLSZM_SZN	0.231 (0.033, 0.412)	0.086 (-0.059, 0.244)	0.019 (-0.079, 0.138)
IH_kurtosis	0.236, (0.044, 0.413)	0.373 (0.177, 0.536)	0.393 (0.039, 0.627)
IH_maxGradl	0.420 (0.24, 0.573)	-0.117 (-0.278, 0.068)	0.046 (-0.053, 0.168)
IH_qcod	0.151 (-0.48, 0.341)	0.218 (-0.042, 0.442)	0.189 (-0.013, 0.379)
IH_skewness	0.412 ().223, 0.571)	*0.598 (0.459, 0.709)	0.419 (0.228, 0.579)
NGLDM_GLNN	0.471 (0.073, 0.698)	0.303 (-0.071, 0.578)	0.178 (-0.087, 0.446)
NGLDM_HGLDE	0.161 (-0.06, 0.371)	0.187 (-0.087, 0.472)	0.08 (-0.057, 0.268)
NGLDM_SDE	0.156 (-0.062, 0.367)	0.267 (-0.095, 0.565)	0.102 (-0.059, 0.272)

NGTDM_busyness	0.000 (-0.203, 203)	0.057 (-0.134, 0.246)	-0.022 (-0.15, 0.126)
Shape_Area_density_BB	0.192 (-0.089, 0.475)	0.179 (-0.072, 0.481)	-0.041 (-0.139, 0.081)
Shape_elongation	**0.805 (0.720, 0.866)	**0.788 (0.702, 0.852)	-0.1 (-0.303, 0.112)
Shape_flatness	0.000 (-0.89, 0.110)	0.000 (-0.111, 0.128)	0.000 (-0.154, 0.169)
Shape_volume_density_BE	-0.064 (-0.234, 0.12)	0.000 (-0.183, 0.186)	-0.016 (-0.145, 0.132)
Stats_mean	**0.719 (0.604, 0.804)	***0.951 (0.923, 0.969)	***0.98 (0.949, 0.99)
Stats_rms	**0.771 (0.672, 0.843)	***0.947 (0.896, 0.97)	***0.98 (0.949, 0.99)
Stats_skewness	0.442 (0.264, 0.592)	*0.615 (0.48, 0.722)	0.491 (0.318, 0.633)
Stats_var	*0.546 (0.374, 0.679)	**0.832 (0.548, 0.92)	*0.608 (-0.084, 0.857)

Table 4.5.Radiomics feature analysis when resampling of voxels was not performed prior to extraction.

*95% confidence intervals are given in parentheses. *=moderate agreement between feature at baseline and without resampling, **= good agreement, ***=excellent agreement. Note that the features highlighted have large heterogeneity as given by their 95% confidence intervals*

Radiomics Feature	ADC ICC	DW-MRI ICC	T2-weighted ICC
Fractal Average	0.101 (-0.075, 0.28)	*0.538 (-0.071, 0.801)	0.226 (0.036, 0.403)
Fractal Lacunarity	-0.004 (-0.057, 0.066)	0.061 (-0.055, 0.199)	-0.05 (-0.075, 0.085)
GLCM_auto	0.000 (-0.105, 0.126)	0.000 (-0.102, 0.121)	0.000 (-0.101, 0.121)
GLCM_average	0.000 (-0.035, 0.049)	0.000 (-0.046, 0.061)	0.000 (-0.034, 0.048)
GLCM_ClusShade	0.000 (-0.192, 0.195)	0.000 (-0.184, 0.187)	0.000 (-0.194, 0.196)
GLCM_ClusProm	0.000 (-0.192, 0.195)	0.000 (-0.158, 0.169)	0.000 (-0.192, 0.195)
GLDZM_HIE	0.000 (-0.095, 0.115)	0.000 (-0.098, 0.116)	0.000 (-0.093, 0.113)
GLDZM_HISDE	0.000 (-0.095, 0.115)	0.000 (-0.098, 0.116)	0.000 (-0.093, 0.113)
GLDZM_LISDE	-0.23 (-0.086, 0.061)	0.069 (-0.092, 0.236)	0.149 (-0.071, 0.424)
GLRLM_LGRE	0.018 (-0.05, 0.105)	0.039 (-0.131, 0.215)	0.151 (-0.081, 0.407)
GLRLM_RE	-0.028 (-0.097, 0.063)	0.067 (-0.081, 0.225)	0.159 (-0.066, 0.451)
GLRLM_SRGHE	0.000 (-0.095, 0.115)	0.000 (-0.098, 0.117)	0.000 (-0.093, 0.113)
GLRLM_SRLGE	-0.024 (-0.097, 0.07)	0.047 (-0.124, 0.222)	0.185 (-0.085, 0.473)
GLSZM_HIE	0.000 (-0.095, 0.115)	0.000 (-0.098, 0.116)	0.000 (-0.093, 0.113)
GLSZM_HISAE	0.000 (-0.095, 0.115)	0.000 (-0.097, 0.116)	0.000 (-0.093, 0.113)
GLSZM_IV	0.000 (-0.092, 0.113)	0.000 (-0.097, 0.116)	0.000 (-0.108, 0.128)
GLSZM_SZN	-0.09 (-0.104, 0.106)	-0.029 (-0.197, 0.149)	0.026 (-0.086, 0.157)
IH_kurtosis	-0.059 (-0.261, 0.148)	-0.009 (-0.206, 0.187)	**0.882 (0.827, 0.92)
IH_maxGradl	0.000 (-0.046, 0.063)	0.000 (-0.053, 0.07)	0.000 (-0.054, 0.072)
IH_qcod	0.069 (-0.068, 0.221)	-0.41 (-0.225, 0.149)	0.282 (-0.079, 0.56)
IH_skewness	-0.041 (-0.242, 0.164)	0.155 (-0.043, 0.34)	**0.86 (0.79, 0.907)
NGLDM_GLNN	-0.029 (-0.092, 0.061)	0.088 (-0.069, 0.252)	0.073 (-0.054, 0.247)
NGLDM_HGLDE	0.000 (-0.094, 0.115)	0.000 (-0.1, 0.119)	0.000 (-0.092, 0.113)
NGLDM_SDE	-0.001 (-0.011, 0.015)	0.01 (-0.069, 0.09)	0.001 (-0.01, 0.016)
NGTDM_busyness	0.000 (-0.203, 0.203)	0.014 (-0.177)	0.006 (-0.096, 0.129)

Shape_Area_density_BB	0.006 (-0.2, 0.21)	0.01 (-0.196, 0.197)	0.041 (-0.165, 0.242)
Shape_elongation	-0.1 (-0.215, 0.195)	-0.124 (-0.314, 0.074)	***1
Shape_flatness	-0.026 (-0.23, 0.179)	0.036 (-0.162, 0.23)	***1
Shape_volume_density_BE	0.029 (-0.177, 0.232)	0.183 (-0.014, 0.365)	***1
Stats_mean	0.000 (-0.027, 0.037)	0.000 (-0.003, 0.004)	0.000 (-0.01, 0.015)
Stats_rms	0.000 (-0.011, 0.017)	0.000 (-0.003, 0.004)	0.000 (-0.01, 0.015)
Stats_skewness	-0.018 (-0.223, 0.187)	0.152 (-0.046, 0.338)	***1
Stats_var	0.000 (-0.-89, 0.110)	0.000 (-0.096, 0.115)	0.00 (-0.106, 0.126)

Table 4.6. Results from radiomics feature analysis when normalisation of images was not performed prior to extraction.

*95% confidence intervals are given in parentheses. *=moderate agreement between feature at baseline and without resampling, **= good agreement, ***=excellent agreement. Note that the features highlighted have large heterogeneity as given by their 95% confidence intervals.*

Radiomics Feature	ADC ICC	DW-MRI ICC	T2-weighted ICC
Fractal Average	**0.718 (0.602, 0.804)	*0.536 (0.382, 0.66)	**0.759 (0.664, 0.830)
Fractal Lacunarity	*0.514 (0.349, 0.649)	**0.775 (0.671, 0.847)	**0.81 (0.713, 0.873)
GLCM_autocorr	0.000 (-0.105, 0.126)	**0.777 (0.687, 0.844)	***0.969 (0.953, 0.979)
GLCM_average	***0.926 (0.871, 0.955)	**0.752 (0.654, 0.825)	***0.984 (0.975, 0.99)
GLCM_ClusShade	0.159 (-0.044, 0.350)	*0.703 (0.59, 0.789)	**0.883 (0.832, 0.919)
GLCM_ClusProm	0.235 (0.037, 0.417)	**0.787 (0.677, 0.859)	**0.846 (0.78, 0.894)
GLDZM_HIE	**0.892 (0.83, 0.93)	**0.793 (0.706, 0.856)	***0.971 (0.956, 0.98)
GLDZM_HISDE	**0.892 (0.83, 0.93)	**0.793 (0.706, 0.856)	***0.971 (0.956, 0.98)
GLDZM_LISDE	**0.801 (0.713, 0.864)	*0.627 (0.494, 0.732)	**0.815 (0.738, 0.871)
GLRLM_LGRE	**0.731 (0.621, 0.813)	0.000 (-0.088, 0.106)	-0.01 (-0.172, 0.177)
GLRLM_RE	***0.922 (0.885, 0.948)	*0.703 (0.59, 0.789)	*0.651 (0.524, 0.749)
GLRLM_SRGHE	**0.896 (0.836, 0.933)	**0.796 (0.712, 0.857)	***0.978 (0.967, 0.985)
GLRLM_SRLGE	**0.763 (0.66, 0.837)	*0.665 (0.541, 0.76)	0.333 (0.15, 0.494)
GLSZM_HIE	**0.892 (0.83, 0.93)	**0.793 (0.706, 0.856)	***0.964 (0.947, 0.976)
GLSZM_HISAE	**0.846 (0.759, 0.901)	*0.606 (0.428, 0.73)	*0.524 (0.36, 0.655)
GLSZM_IV	*0.553 (0.396, 0.679)	**0.838 (0.753, 0.893)	***0.967 (0.95, 0.978)
GLSZM_SZN	**0.784 (0.581, 0.879)	**0.82 (0.745, 0.875)	**0.898 (0.853, 0.93)
IH_kurtosis	-0.06 (-0.211, 0.198)	*0.624 (0.49, 0.729)	0.056 (-0.138, 0.246)
IH_maxGradl	-0.134 (-0.329, 0.071)	0.143 (-0.044, 0.323)	**0.755 (0.657, 0.827)
IH_qcod	*0.547 (0.387, 0.675)	0.134 (-0.061, 0.319)	-0.051 (-0.243, 0.145)
IH_skewness	*0.642 (0.505, 0.747)	-0.117 (-0.279, 0.064)	-0.028 (-0.229, 0.175)
NGLDM_GLNN	**0.805 (0.72, 0.867)	0.257 (0.067, 0.429)	0.345 (0.163, 0.504)
NGLDM_HGLDE	**0.763 (0.574, 0.861)	0.012 (-0.142, 0.176)	0.207 (0.023, 0.379)
NGLDM_SDE	*0.706 (0.512, 0.819)	0.089 (-0.058, 0.247)	-0.056 (-0.227, 0.123)
NGTDM_busyness	***1	0.006 (-0.189, 0.2)	-0.049 (-0.241, 0.146)

Shape_Area_density_BB	*0.701 (0.509, 0.813)	*0.546 (0.213, 0.731)	0.129 (-0.064, 0.314)
Shape_elongation	**0.857 (0.791, 0.903)	*0.739 (0.637, 0.815)	0.02 (-0.174, 0.213)
Shape_flatness	***0.92 (0.882, 0.946)	0.441 (0.263, 0.588)	0.122 (-0.073, 0.308)
Shape_volume_density_BE	***0.944 (0.917, 0.963)	0.223 (0.034, 0.397)	**0.788 (0.702, 0.851)
Stats_mean	***0.992 (0.988, 0.995)	0.193 (-0.001, 0.372)	***0.942 (0.916, 0.96)
Stats_rms	***0.962 (0.932, 0.977)	0.107 (-0.088, 0.295)	**0.801 (0.719, 0.861)
Stats_skewness	*0.674 (0.547, 0.771)	0.031 (-0.164, 0.224)	*0.647 (0.519, 0.746)
Stats_var	*0.619 (0.476, 0.731)	0.061 (-0.132, 0.249)	*0.54 (0.387, 0.664)

Table 4.7. Results from radiomics feature analysis across all MR sequences when nearest neighbour interpolation algorithm used instead of linear algorithm.

*95% confidence intervals are given in parentheses. *=moderate agreement between feature at baseline and without resampling, **= good agreement, ***=excellent agreement*

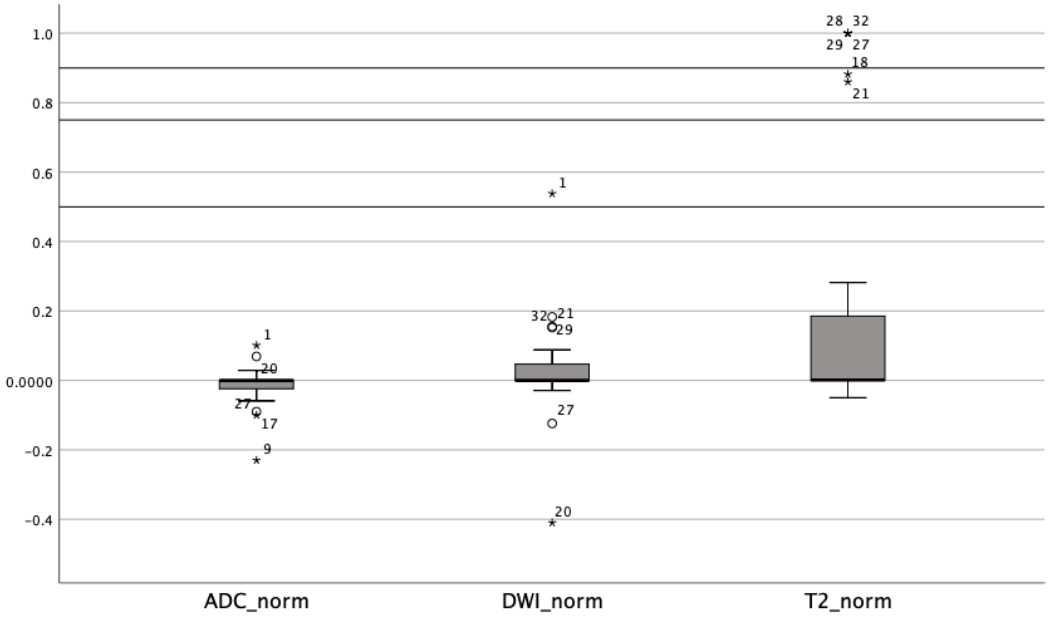


Figure 4.4. ICC when normalisation not performed. *ICC thresholds of 0.5, 0.75 and 0.9 are highlighted. The outliers highlighted are mainly Shape features for T2-weighted imaging. Overall, reproducibility of radiomics features on MRI is not possible without image normalisation.*

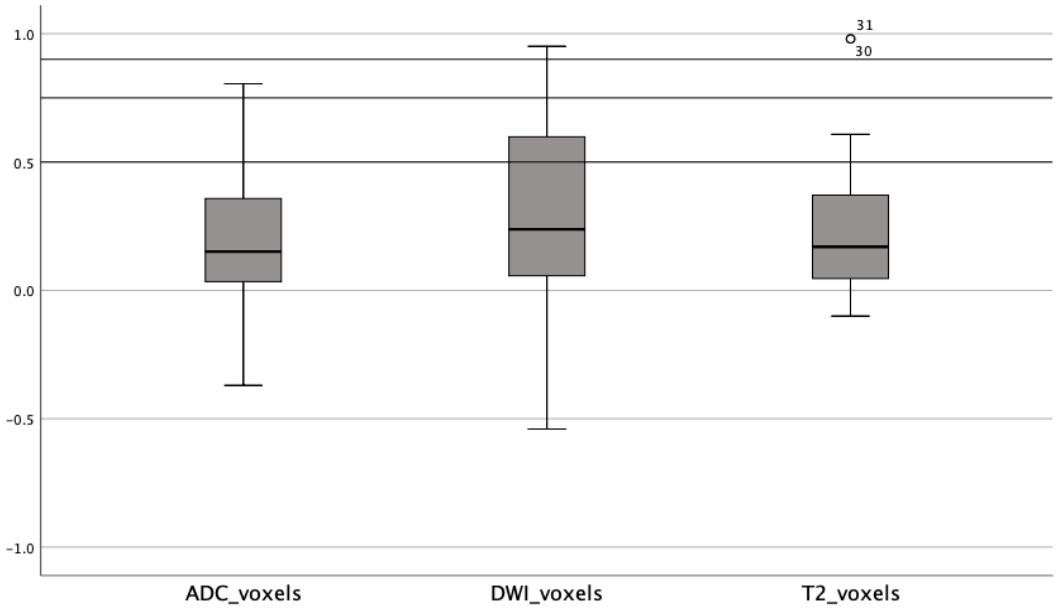


Figure 4.5. ICC when resampling of voxels was not performed. *ICC thresholds of 0.5, 0.75 and 0.9 are highlighted. We can interpret that reproducibility of features overall when resampling is not performed is very poor, especially for ADC maps and T2-weighted imaging.*

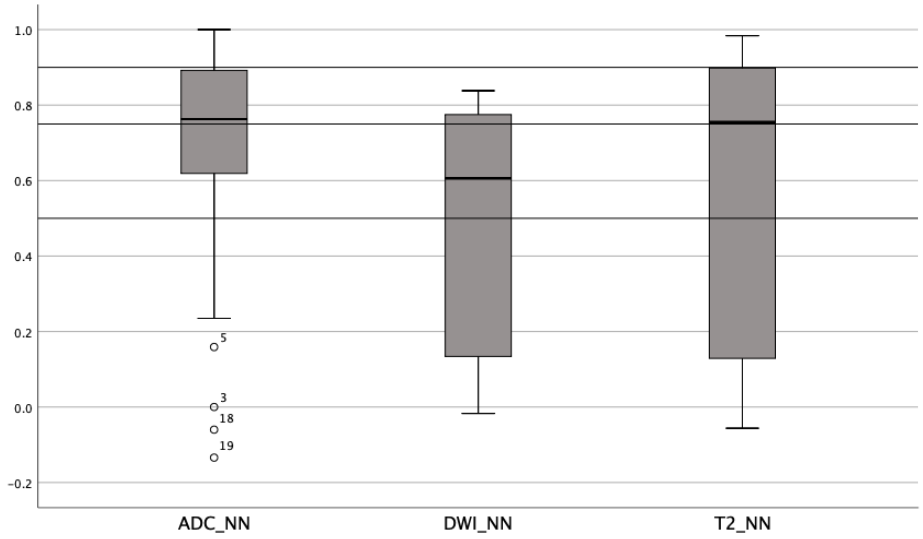


Figure 4.6. ICC when interpolation algorithm was changed from linear to nearest neighbour. *ICC thresholds of 0.5, 0.75 and 0.9 are highlighted. Outliers on ADC are from the IH and GLCM feature families. Note that the number of features having a value of >0.5 is higher than in Figures 4.4 and 4.5.*

4.4 DISCUSSION

The image intensity of MR images is usually relative and not considered comparable between images for different patients or for longitudinal studies, meaning that intensity normalisation is required [122]. This is supported by the results of this analysis when normalisation was not performed on ADC maps, DW-MRI or T2-weighted images prior to feature extraction. As can be seen in Figure 4.4, only some features from the Shape family were reproducible when normalisation was not performed, with the majority unable to discriminate across remaining feature families (ICCs of 0). Using Pyradiomics software, it has been noted in analysis of ADC maps in cervical cancer that Shape family features were left unchanged by normalisation, filtering and quantisation unless they were volume independent [119]. However, others have found that voxel variation can even impact first order features [163]. In this analysis, Shape features were seen to be very reproducible for ADC when nearest neighbour algorithm was utilised instead of linear interpolation and on T2-weighted images when normalisation of the image was not performed prior to feature extraction, the latter most likely because only one observer delineated all ROIs in this analysis and Shape features describe the volume that has been delineated in the ROI [173]. High repeatability and reproducibility of Shape features have been previously reported in cervical cancer radiomics analysis [116] and are also more robust to image noise than GLCM and GLSZM textural features in lung cancer [175].

In this study, images were normalised by centring the image at its mean value and comparing this with no normalisation. There is no consensus on the use of normalisation for ADC maps at present [122]; however the current study has clearly illustrated that reproducibility of features on ADC maps before and after normalisation was poor, with no feature being considered reproducible on ICC analysis.

Others have compared normalisation methods in prostate cancer radiomics on T2-weighted MRI. These include normalisation of the image at its mean value, normalisation of grey level values to a defined ROI that has a fixed value or adjusting to a reference histogram. These have found that most parameters are very sensitive to normalisation, with the exceptions being first order features and a single GLCM feature . Normalising to a defined ROI with a fixed value most

affected the reproducibility of radiomics features in this study by Scalco et al [89], indicating that either the histogram method or normalising to the mean of the image, as was done in the current analysis, is a better option. Reproducibility on T2-weighted MRI without normalisation in this study was extremely poor, except for the first order Stats family, illustrating the importance of this image pre-processing step in radiomics research.

First order features are considered global statistical descriptors and textural features local descriptors as they analyse patterns inside grey values that can have considerably altered values between the original image and the pre-processed image [172]. The spatial variation of textural features is quantified in voxel intensities. Resampling of voxels either increases (down-sampling) or decreases (up-sampling) the distance between voxels. Larger voxel sizes (down-sampling) means that information is lost and smaller voxel sizes (up-sampling) can lead to the creation of false or artificial information [174]. In a study of ADC radiomics in rectal cancer, it was found that when pixel size changed, GLCM features were moderately repeatable and the majority of GLSZM features were poorly repeatable. In the current analysis, neither GLCM nor GLDZM features were considered reproducible on ADC without voxel resampling with very low ICCs (0.034-0.079 for GLCM and 0.037-0.085 for GLDZM). 2 GLCM features were moderately reproducible on DW-MR (GLCM_autocorr and GLCM_average) and 1 on T2-weighted imaging (GLCMClusShade) for the same changed pre-processing parameter. No GLDZM feature was reproducible across all 3 sequences when resampling of voxels was not performed. Similar to Traverso et al [172], it was also established in the current study that on ADC analysis, first order features, such as the Stats feature family, were robust to changes in voxel size, with 3 out of 4 features showing either moderate (Stats_var 0.546) or good (Stats_mean 0,719, Stats_rms 0.771) reproducibility.

The ADC maps in this analysis were created using *b* values from 200 to 800 s/mm^2 and the ROI delineated on the maps based on the pathological specimen. In radiation therapy practice, where no pathological specimen would be available for such delineation purposes, it would be expected that radiomics features extracted from ADC maps would be dependent on the magnitude of the gradients used when generating the ADC map. However, excellent reproducibility of ROIs delineated by two observers on ADC maps across 31 *b* value combinations (ICC

0.92-0.998) has previously been reported in rectal cancer, indicating that *b* value consistency across institutions may not be necessary, although a radiomics analysis was not conducted on the resulting ROIs in this analysis [181].

Some authors advocate for resampling of voxels before any radiomic analysis as the variability in feature values is as low as possible when the voxel size is equal for all included images [168] and noise is reduced [182]. This hypothesis has been proven when comparing the results of radiomics feature reproducibility in rectal cancer, where CT slice thickness was 5 mm, and in lung cancer, where CT slice thickness was 1.25 mm, but when resampling of data into isotropic voxels of 3 mm was performed, the results did not significantly vary [170]. It was also observed in the phantom study of this thesis where increased slice thickness, which directly impacts on voxel size, had a very negative impact on radiomics feature reproducibility, with 61.3% of the 32 features analysed not being repeatable. Of note is that when slice thickness was changed in the phantom study and voxel resampling was not performed in this patient analysis (both T2-weighted MRI), only some features from the first order Stats family were reproducible, most likely due to noise.

When either voxel size resampling or image normalisation was not performed in this analysis, reproducibility across all MR sequences was extremely poor, with values of 0 reported across many textural feature families. This can be attributed to the fact that the information conveyed by the feature values without such image pre-processing is mainly related to the volume of the ROI and not to texture or other intervoxel relationships [163].

Currently, no definitive recommendation for a specific interpolation algorithm is given for radiomics analysis. Therefore the impact of algorithm selection was explored in this analysis. Changing from linear interpolation to nearest neighbour interpolation algorithm for resampling method yielded the highest number of reproducible results. For the ADC dataset, 28/33 features were reproducible, 19/33 features on DW-MRI were reproducible, as were 20/33 for T2-weighted analysis. The results of this analysis showed good to excellent reproducibility for GLDZM features, poor reproducibility for NGLDM and NGTDM features, some positive results for GLRLM and poor reproducibility for the IH feature family. However, the Image Biomarker Standard Initiative [86] warns of the potential of

bias in textural features using this algorithm; hence these results are difficult to fully interpret.

Similar results on the impact of image acquisition parameters on radiomics feature extraction are seen in other imaging modalities. Image noise in CT, (specifically tube current, noise index and reconstruction algorithm) has been shown to impact on the reproducibility of radiomics features. An increase in tube current improved reproducibility while increasing reconstruction caused a blurring in the image, which reduced reproducibility [183]

. An intra-CT phantom analysis also illustrated the negative effect of modification of the reconstruction kernel, yielding a reproducibility of 43.1% (76 out of 177 features), while varying the pitch had a lesser effect on reproducibility of 89.3% (158 out of 177 features).

4.4.1 CONCLUSION

Based on this analysis, it can be concluded that normalisation of MR images prior to radiomics feature extraction is a pre-requisite in order to keep grey levels within a specified range. Image resampling is particularly important when data are across different scanners, manufacturers and institutions, but it also helps in noise reduction even within the same institute. However, the impact of up-sampling and down-sampling must be understood. First order features such as the Stats feature family appear to be robust to changes in voxel size. Finally, although the nearest neighbour algorithm did yield the overall highest number of reproducible results in this analysis, some non-reproducible textural features were also observed, in line with previous reports that this algorithm has the potential to introduce bias in textural feature analysis [183].

CHAPTER 5. STUDY 3: RADIOMICS FOR PREDICTION OF CLINICAL FEATURES IN PROSTATE CANCER: A T2-WEIGHTED MRI ANALYSIS

5.1 INTRODUCTION

T2-weighted MRI sequences provide anatomical information on the prostate, with excellent spatial resolution [184]. The differences in hydrogen transverse (T2-weighted) MRI relaxation times between normal and pathological tissues are key to image contrast and discrimination of cancer [185]. T2-weighted MRI alone can localise prostate cancer with sensitivity ranging from 54-82% and specificity from 46-91% [186].

Typically, these images are acquired in the axial and coronal planes, at a minimum. Axial slices are acquired perpendicular to the posterior surface of the prostate and coronal slices perpendicular to this plane and in the axis of the seminal vesicles. Adding a sagittal plane gives an optimal view of the cranio-caudal extent of prostate cancer lesions, as well as the prostatic apex [187]. As the imaging in each of these planes can take up to 4 minutes on a 1.5 Tesla MRI, rectal motion should be minimised using an enema as well as an anti-peristaltic, such as glucagon. The field of view should be 12-20 cm to encompass the entire prostate gland and seminal vesicles. A pelvic phased array coil with a minimum of 16 channels is a requirement, but an endorectal coil is not a prerequisite, according to the European Society of Urogenital Radiology (ESUR) guidelines. These guidelines also indicate that T2-weighted images utilised for detection in the axial and sagittal planes should be acquired in 4 mm slices, with in-plane resolution between 0.5 X 0.5 mm to 0.7 X 0.7 mm at 1.5 Tesla. For staging, slice thickness should be 3 mm with 0 mm gap at 1.5 Tesla and the in-plane resolution from 0.3 X 0.3 mm to 0.7 X 0.7 mm [63].

Prostate cancers have specific features on T2-weighted sequences, depending on the anatomical location of the lesion within the prostate. In the peripheral zone, a hyposignal, a homogenous signal area, a nodular shape or an area with convex edges, arouses suspicion. T2 values in the peripheral zone have been shown to overlap with those of benign prostatic hyperplasia [188], thereby making diffusion-weighted MRI the dominant sequence in this region, as it is very unusual for diffusion to be restricted in the peripheral zone for reasons other

than cancer [189]. For example, Langer et al [190] reported that T2 values in prostate cancer range from 68.9-102.7 ms in the cancerous peripheral zone and from 82-162 ms in the normal peripheral zone. Low signal can also be attributable to acute or chronic prostatitis, haemorrhage, previous treatment and atrophy post-inflammation [189].

In the transition zone, the T2-weighted sequence is difficult to analyse due to the overlap of signal intensity characteristics of the normal and cancerous zone but is considered the dominant MRI sequence for detection in this region. Poor circumscription, known as the 'erased charcoal sign' or nodules with a lenticular shape, are considered suspicious and if present are usually located in the anterior half [5]. When T2-weighted imaging is equivocal in the transition zone, positive diffusion weighted imaging increases the likelihood of clinically significant cancer [191]. For the anterior fibromuscular stroma, the absence of a very marked T2-weighted hyposignal warrants guided biopsy in this region [187]. It has been reported that 17% of tumours in the apex are missed on standard TRUS biopsy [184].

T2-weighted MRI is also used to evaluate the prostatic capsule, seminal vesicles and posterior bladder wall for tumour invasion. Seminal vesicle invasion yields a low T2 signal intensity and filling-in of the prostate-seminal vesicle angle. Extra-capsular extension (ECE) is identified through bulge, loss of capsule and capsular enhancement, obliteration of the recto-prostatic angle, abutment and neurovascular bundle thickening. All of these are also scored individually from 1-5 in the PI-RADS scoring system [192].

5.1.1 AIMS AND OBJECTIVES

The aim of this chapter is to establish that radiomics features from T2-weighted images features can predict specific clinical endpoints.

The objectives are:

- To identify the principal components of extracted radiomics features from the prostate index lesion following feature reduction.
- To relate radiomics features from T2-weighted MRI to PIMO score, Gleason score at biopsy, D'amico risk score and clinical tumour stage.

5.2. MATERIALS AND METHODS

5.2.1 T2-WEIGHTED DATASETS

The acquisition of the T2-weighted MRI datasets and subsequent ROI delineation have been discussed in Chapter 2. In line with the reproducibility results of the phantom chapter, the images were acquired maximising signal-to-noise ratio, using a matrix of 512 X 512, slice thickness of 3 mm and voxel size of $1.5mm^3$. T2-weighted images and associated regions of interest (ROI) were analysed using RadiomiX (RadiomiX Research Toolbox version 20180831 (OncoRadiomics SA, Liège, Belgium)) software.

5.2.2 RADIOMIC FEATURE EXTRACTION PARAMETERS

165 features were extracted after data had been normalised to zero mean, unit variance. T2-weighted MRI scans are known to illustrate intensity-related drift artefacts, even within the same patient and the same scanner, resulting in variances in signal intensities [31], hence the need for normalisation of images prior to feature extraction, as illustrated in the analysis in Study 3. Resampling was performed using 3 mm isotropic voxels, bin width was set to 0.1 to ensure textural analysis was computed and the linear interpolation algorithm was set. The analysis in Study 3 illustrated that linear interpolation is favourable to nearest neighbour interpolation, as does the literature [86] The features extracted were from the radiomics feature families or matrices outlined in Chapter 2, namely, morphological features, local intensity features, first-order statistics, intensity histogram features, fractal features and the textural feature families of GLCM, GLRLM, GLSZM, GLZDM, NGTDM and NGLDM.

5.2.3 STATISTICAL ANALYSIS

Feature reduction was performed by Principal Components Analysis (PCA) in SPSS Version 26 (IBM Corp. Released 2019. IBM SPSS Statistics for Macintosh, Version 26.0. Armonk, NY: IBM Corp). Pearson correlations were then run on the PCA component scores and the original 165 features extracted to determine the features which best represented the Principal Components (PCs). Features which were the highest correlated with the PCs were then utilised in linear regression analysis.

Multiple linear regression analysis was conducted to ascertain if radiomics features could predict PIMO score, Gleason score at biopsy, D'amico risk score and clinical tumour (T) stage. For PIMO score, a β_1 co-efficient of > 0.5 determined which independent variables were retained in subsequent linear regression analyses. PIMO score ranged from 0-5 and therefore utilising a β_1 co-efficient of > 0.5 was deemed conservative. For Gleason score at biopsy, D'amico risk stratification, pre-operative PSA level and clinical T stage, the independent variables with the 3 highest β_1 co-efficients were utilised in subsequent linear regression analyses. In order to determine if radiomics features could predict for these clinical endpoints, a ROC curve for each endpoint was calculated and the AUC reported.

5.3 RESULTS

5.3.1 PATIENT CHARACTERISTICS

The clinical features of all 88 analysed patients were given in Chapter 4. To summarise, the mean age was 64 years (range 47-75 years). The mean pre-op PSA was 12.4 (range 2.2-145). The mean Gleason score at biopsy was 7 (range 6-9). The Clinical (c) T stages were as follows: cT1c (n=37), cT2a (n=8), cT2b (n=8), cT2c (n=4), cT3a (n=7), cT2x (n=14) and no cT stage was recorded for 5 patients. 2 patients were classified as low risk, 47 as intermediate and 39 as high risk as per the D'amico risk classification. The mean PIMO score was 1.9 (range 0-5) and 10 patients in this cohort had not received PIMO, as discussed in Chapter 2.

5.3.2 PRINCIPAL RADIOMICS FEATURES IDENTIFIED ON T2-WEIGHTED MRI

A principal components analysis (PCA) was run on the 165 extracted components. The suitability of PCA was assessed prior to analysis. Inspection of the correlation matrix showed that all variables had at least one correlation coefficient greater than 0.3, which illustrated linearity between all variables.

PCA revealed 16 components that had eigenvalues greater than one and which explained 39.369%, 21.248%, 6.477%, 5.032%, 4.229%, 3.187%, 2.655%, 2.374%, 1.833%, 1.563%, 1.130%, 1.091%, 0.936%, 0.811%, 0.730% and 0.682% of the

total variance, respectively. Visual inspection of the scree plot (Figure 5.1) indicated that 16 components should be retained.

A forced factor number PCA on these 16 components was conducted. The 16-component solution explained 93.345% of the total variance (Table 5.1 and full Table Appendix 1). Pearson correlations were conducted on the sixteen retained PCs and the original 165 extracted features. The radiomics features which had the highest Pearson correlation coefficients with the PCs were as follows: 3 from first order statistics (Stats_var, Stats_p10, Stats_qcod), 2 from the grey level size zone feature family (GLSZM_LAE, GLSZM_IN), 1 from the grey level distance zone feature family (GLDZM_IN), 2 from the grey level co-occurrence feature family (GLCM_clusShade, GLCM_inverse var), 2 from the neighbouring grey level dependence matrix (NGLDM_LGE, NGLDM_SDE), 3 from the intensity histogram family (IH_uniformity, IH_kurtosis, IH_maxGradl,) and 3 based on shape features (Shape_flatness, Shape_areaDensityBB, Shape_elongation). These were independent variables in subsequent regression analysis.

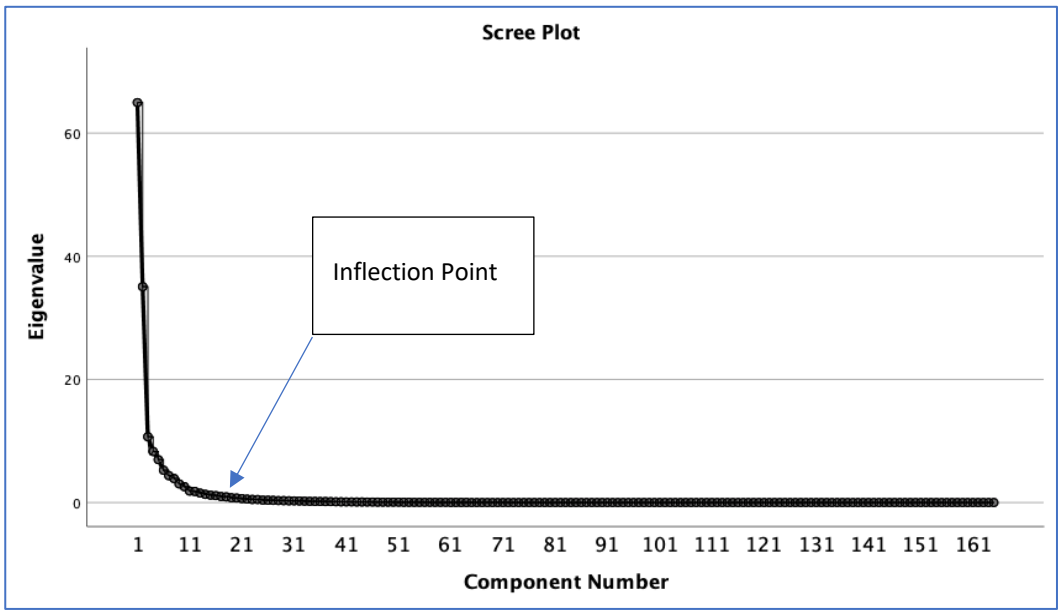


Figure 5.1. Scree Plot of PCA Analysis showing 16 components prior to the inflection point.

Feature Number	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	*% of Variance*	*Cumulative %*	*Total*	*% of Variance*	Cumulative %	*Total*	% of Variance	*Cumulative %*
1	64.959	39.369	39.369	64.959	39.369	39.369	50.670	30.709	30.709
2	35.060	21.248	60.618	35.060	21.248	60.618	20.937	12.689	43.398
3	10.686	6.477	67.094	10.686	6.477	67.094	18.518	11.223	54.621
4	8.303	5.032	72.127	8.303	5.032	72.127	16.386	9.931	64.552
5	6.979	4.229	76.356	6.979	4.229	76.356	8.133	4.929	69.481
6	5.258	3.187	79.543	5.258	3.187	79.543	7.816	4.737	74.218
7	4.380	2.655	82.197	4.380	2.655	82.197	7.722	4.680	78.897
8	3.917	2.374	84.571	3.917	2.374	84.571	4.451	2.697	81.595
9	3.024	1.833	86.404	3.024	1.833	86.404	3.679	2.230	83.825
10	2.578	1.563	87.967	2.578	1.563	87.967	2.779	1.684	85.509
11	1.865	1.130	89.097	1.865	1.130	89.097	2.604	1.578	87.087
12	1.801	1.091	90.188	1.801	1.091	90.188	2.581	1.564	88.651
13	1.544	.936	91.123	1.544	.936	91.123	2.383	1.444	90.095
14	1.338	.811	91.934	1.338	.811	91.934	1.971	1.195	91.290
15	1.204	.730	92.664	1.204	.730	92.664	1.831	1.110	92.400
16	1.125	.682	93.345	1.125	.682	93.345	1.561	.946	93.345

Table 5.1. Total Variance Explained Table from Principal Components Analysis of 16 features. *The full table of all 165 features is available in Appendix 1.*

5.3.3 PREDICTION OF PIMO SCORE

A multiple linear regression was run to assess the predictive potential of radiomics feature(s) for PIMO score. Assumption tests identified the presence of homoscedasticity (Figure 5.2), normality of the residuals (Figure 5.3) and independence of residuals (Durbin-Watson statistic of 1.903). Similar assumption tests were run for all other linear regression analyses and similar results were achieved.

A radiomics model that included the features NGLDM_SDE, Stats_var, Stats_p10, GLCM_inverseVar, IH-uniformity, Shape_flatness and Shape_elongation significantly predicted PIMO score, $F(7, 83) = 2.632$ $p = 0.017$ accounting for 19.5% of the variation in PIMO score with adjusted $R^2 = 12.1$ %, a medium size effect according to Cohen [193]. The impact of each component of the model is given in Table 5.2.

A second linear regression, including clinical parameters, as well as the radiomics model features, was run. The clinical features included were: pre-operative PSA level, clinical T stage, Gleason score at biopsy, D'amico risk score and lymph node status of patients. Homoscedasticity and normality of the residuals were once again confirmed. The model changed, with some clinical parameters superseding radiomics features (Stats_var,Stats_p10, GLCM_InverseVar, NGLDM_SDE, IH_uniformity, Shape_flatness, Shape_elongation, Pre-op PSA, cT, Gleason score at biopsy, D'amico risk classification and N status $F(12, 77) = 1.960$, $p = 0.043$, accounting for 26.6% of the variation in PIMO score with adjusted $R^2 = 13\%$. The component values for each model parameter are given in Table 5.3. On receiver operating characteristic (ROC) analysis, this model yielded an area under the curve (AUC) of 0.591 when PIMO score was stratified into PIMO <3 as not hypoxic and PIMO score ≥ 3 as hypoxic (Figure 5.4).

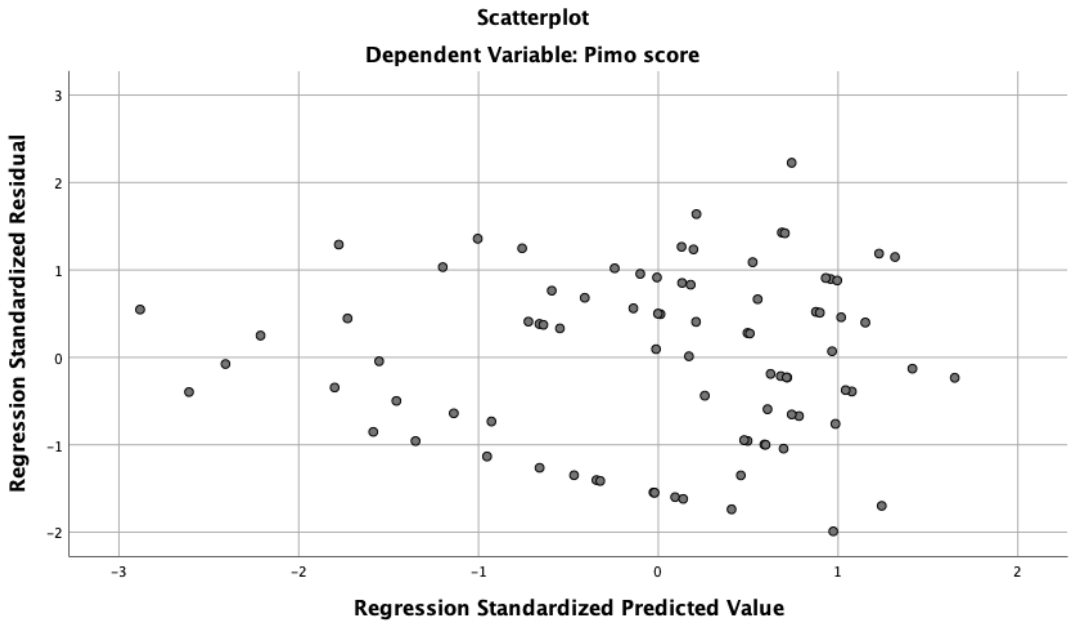


Figure 5.2. Homoscedasticity of residuals where PIMO score is the dependent variable- note: points of the plot above exhibit no pattern and are approximately constantly spread.

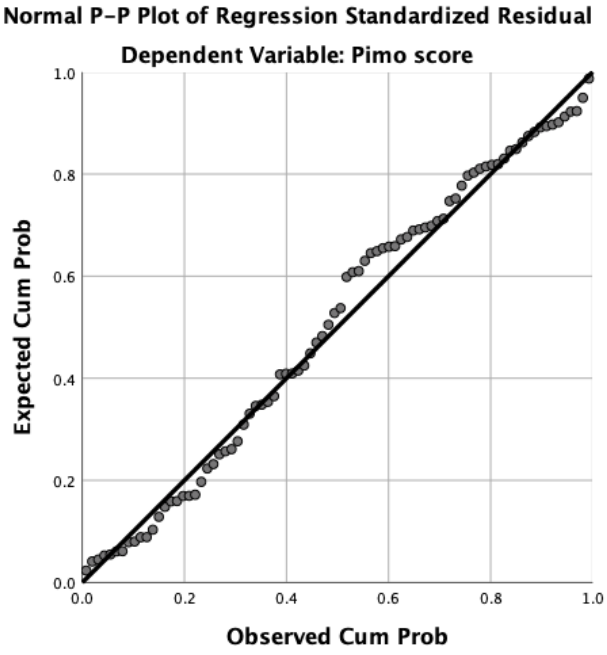


Figure 5.3. Normal Probability Plot illustrating independence of residuals.

PIMO score radiomics parameters					
Model		Unstandardised Coefficients		Standardised Coefficients	t
		B	Std. Error	Beta	
	(Constant)	4.901	1.447		3.387
	NGLDM_SDE	-3.027	.990	-.407	-3.057
	Stats_var	-2.123	13.319	-.025	-.159
	Stats_p10	.656	.910	.084	.721
	GLCM_inverseVar	-1.910	1.408	-.172	-1.357
	IH_uniformity	-.533	1.509	-.055	-.353
	Shape_flatness	-1.336	.842	-.183	-1.586

	Shape_elongation	-.690	.765	-.099	-.902

Table 5.2 Radiomics model parameters for prediction of PIMO score.

Model		Unstandardised Coefficients		Standardised Coefficients	t
		B	Std. Error	Beta	
	(Constant)	2.899	2.233		1.298
	NGLDM_SDE	-2.740	1.080	-.376	-2.536
	Stats_var	.270	13.584	.003	.020
	Stats_p10	.298	.958	.039	.311
	GLCM_inverseVar	-.946	1.571	-.083	-.602
	IH_uniformity	-.289	1.593	-.030	-.182
	Shape_elongation	-.502	.884	-.071	-.568
	Shape_flatness	-1.306	1.003	-.168	-1.302
	PreOp PSA	-.011	.009	-.151	-1.308

	cT	-.005	.051	-.011	-.089
	Risk_classification	.254	.383	.108	.663
	Glscore at biopsy	.070	.269	.042	.259
	N_status_	.767	.519	.183	1.476

Table 5.3 Radiomics and clinical feature model parameters for prediction of PIMO score.

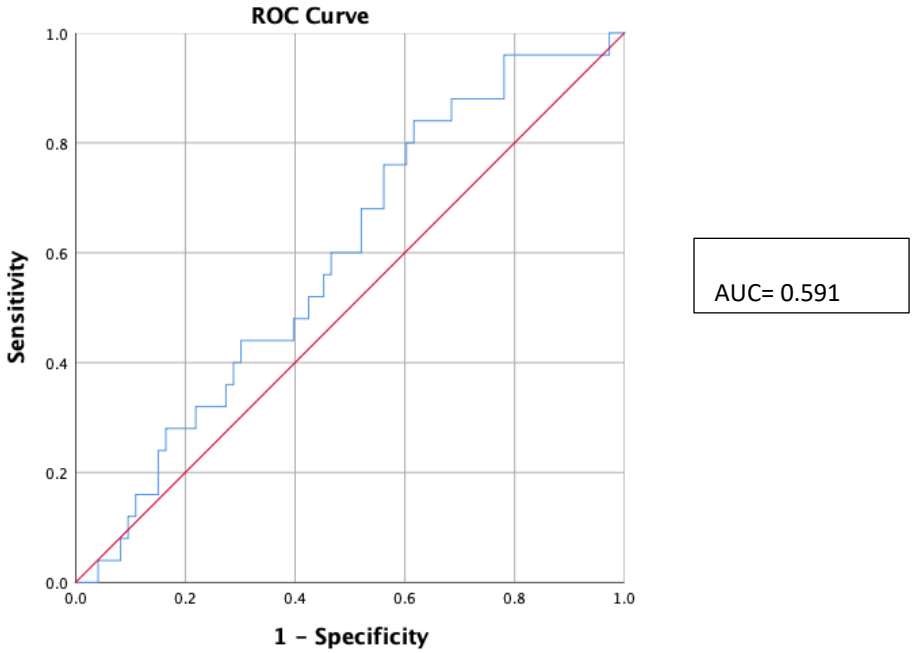


Figure 5.4. ROC curve for prediction of PIMO score by combined radiomics-clinical parameter model.

5.3.4 PREDICTION OF GLEASON SCORE AT BIOPSY

A radiomics model with the features Stats_Var and GLCM_inverseVar was found to statistically significantly predict for Gleason score at biopsy on linear regression, $F(2,94) = 3.386$, $p = 0.038$, accounting for 6.9% variation in Gleason score at biopsy and an adjusted $R^2 = 4.8\%$. The component values are given in Table 5.4. On ROC analysis, when Gleason score was stratified according to Gleason 7 (4+3) and above, compared with Gleason 7 (3+4) and below, the AUC was 0.580 (Figure 5.5).

Model		Unstandardised Coefficients		Standardised Coefficients	t
		B	Std. Error	Beta	
	(Constant)	7.936	.343		23.169
	Stats_var	-7.049	3.713	-.196	-1.898
	GLCM_inverseVar	-1.438	.667	-.222	-2.155

Table 5.4. Radiomics model parameters for prediction of Gleason score.

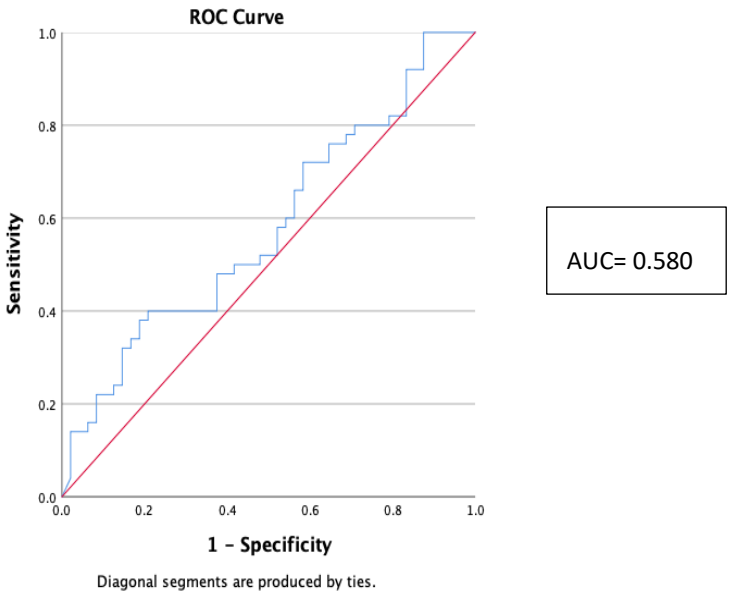


Figure 5.5 ROC curve for prediction of Gleason score 7 (4+3) and above by radiomics model.

5.3.5 PREDICTION OF CLINICAL TUMOUR STAGE

A model consisting of Stats_Var,GLCM_inverseVar and Shape_areaDensityBB significantly predicted Clinical Tumour (T) stage $F(3, 89) = 4.116$, $p = 0.009$ accounting for 12.6% of the variation in clinical tumour stage, with adjusted R^2 = 9.5 % (Table 5.5). ROC analysis was equivocal when clinical T stage was stratified according to a threshold of T3a or above (T3a indicating unilateral or bilateral extracapsular extension and impacting on treatment decision), with an AUC of 0.501 (Figure 5.6).

Model		Unstandardised Coefficients		Standardised Coefficients	t
		B	Std. Error	Beta	
	(Constant)	17.946	10.004		1.794
	Stats_var	-57.670	55.721	-.107	-1.035
	GLCM_inverseVar	-35.832	10.608	-.351	-3.378
	Shape_areaDensityBB	7.814	14.006	.057	.558

Table 5.5. Radiomics model parameters for prediction of clinical Tumour stage.

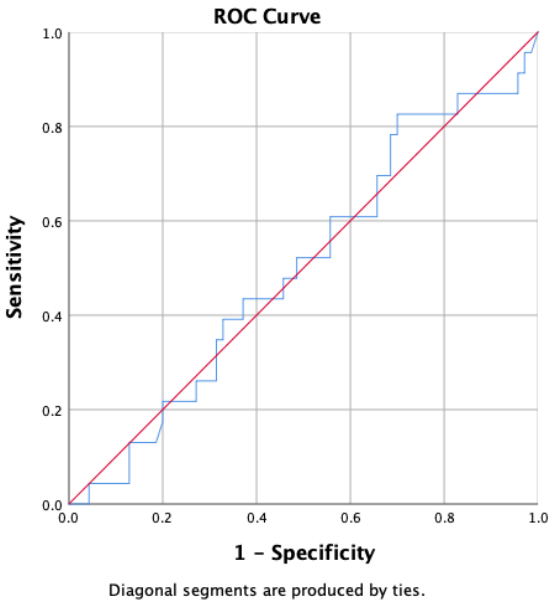


Figure 5.6. ROC curve for prediction of clinical T3a and above by radiomics model.

5.3.6 PREDICTION OF D'AMICO RISK STRATIFICATION SCORE

A radiomics model including all 16 radiomics variables significantly predicted D'amico risk stratification $F(15,94)=1.887$, $p = 0.037$ accounting for 26.4% of the variation in risk stratification, with adjusted R^2 = 12.4% (Table 5.6). On ROC analysis, the model had moderate prediction, with an AUC of 0.753 when dichotomised into low or intermediate risk versus high risk (Figure 5.7).

D'amico Risk Stratification Score					
Model		Unstandardised Coefficients		Standardised Coefficients	t
		B	Std. Error	Beta	
	(Constant)	1.828	.833		2.195
	Stats_var	-3.959	6.446	-.155	-.614
	Shape_areaDensityBB	1.411	.813	.222	1.736
	GLCM_inverseVar	-.867	.672	-.189	-1.290
	GLSZM_LAE	-1.503E-5	.000	-.003	-.031
	GLDZM_IN	.044	.080	.065	.548
	IH_uniformity	.524	.812	.128	.645

	NGLDM_LGE	-.982	.471	-.305	-2.083
	NGLDM_SDE	-.171	.456	-.056	-.376
	Stats_p10	-.315	.365	-.101	-.863
	IH_kurtosis	-.008	.043	-.022	-.190
	Shape_flatness	-.251	.383	-.079	-.655
	GLCM_clusShade	5.252E-5	.003	.004	.019
	Stats_qcod	.013	.080	.018	.166
	Shape_elongation	.345	.316	.118	1.090
	IH_maxGradI	.004	.058	.008	.071

Table 5.6. Radiomics model parameters for prediction of D'amico risk stratification score.

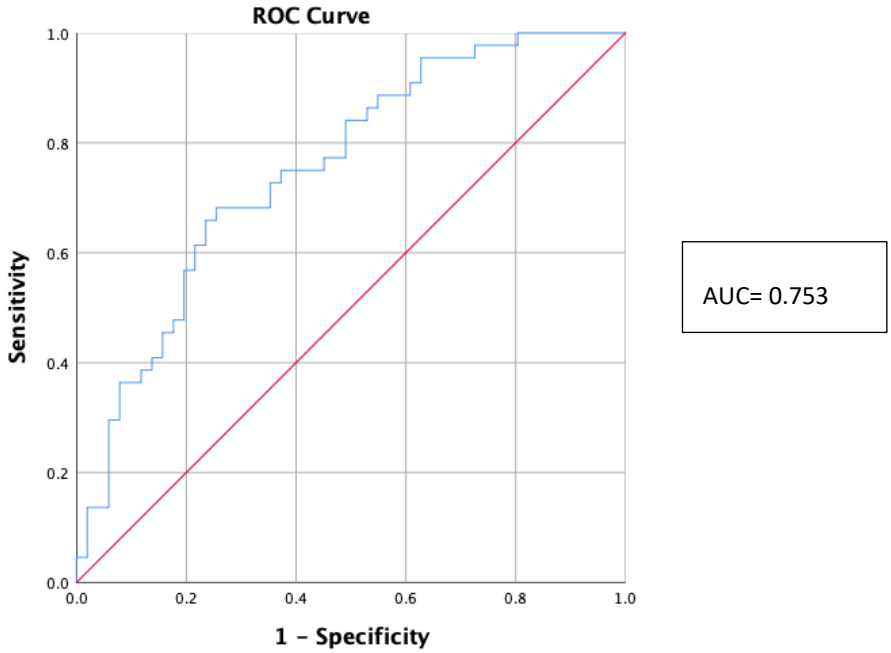


Figure 5.7. ROC curve for prediction of D'amico risk stratification score by radiomics model.

5.4 DISCUSSION

5.4.1 RADIOMICS PREDICTION OF HYPOXIA ON T2-WEIGHTED MRI

Radiomics-based analysis of hypoxia is in its infancy but has the potential to be of high clinical relevance as conventional measurements of tumour hypoxia using functional imaging, molecular biomarkers or direct probes are expensive, time-consuming and, in some instances, invasive [194]. As discussed in Chapter 1, hypoxia is one of the leading causes of failure of radiation therapy and progression of disease [195]. PIMO staining is an established method for the surrogate assessment of the presence and extent of tumour hypoxia [71].

In this analysis, a 7-feature radiomics-only model significantly predicted for PIMO score (p=0.017). The textural feature neighbouring grey level dependence matrix_small dependence emphasis (NGLDM_SDE) was most prominent in the model with a β coefficient of -3.027 (p=0.003). In the combined radiomics-clinical model, which also significantly predicted for PIMO score (p= 0.043), NGLDM_SDE also had the highest β coefficient of -2.740 (p=0.014). NGLDM_SDE is a feature of the neighbouring grey level dependence feature family, which is an alternate to the grey level co-occurrence matrix. This feature family captures the coarseness of the overall texture of the ROI with the small dependence emphasis specifically focused on the low neighbouring grey level dependence counts [86]. In their analysis of the same patient cohort, Hompland et al [120] found that PIMO staining correlated with pathologic tumour grade, pathologic Gleason score, lymph node status and largest tumour extent, leading to their conclusion that a high PIMO score indicated an aggressive tumour phenotype in this cohort. From the current analysis, it can be postulated that the presence of the radiomics feature NGLDM_SDE could be indicative of an aggressive phenotype; however with an AUC of 0.591 in the prediction of PIMO score on ROC analysis, this remains a hypothesis.

Beig et al [196] found that radiomics textural features were predictive of a hypoxia enrichment score, based on the 21 genes implicated in the hypoxia pathway of glioblastoma multiforme (GBM). Using 180 multiparametric images from GBM patients, three ROIs were defined: oedema, including the non-enhancing tumour; tumour necrosis; and enhancing tumour. On T2-weighted

images of the tumour necrosis ROI, the textural feature *information measure of correlation 1,* which captures co-occurrences and quantifies structural heterogeneity, was identified to be one of the top 8 features most associated with hypoxia enrichment score.

In methodology similar to that utilised in the current analysis, 14 non-small cell lung cancer (NSCLC) patients who were administered intravenous PIMO within 24 hours of pneumonectomy had their computed tomography (CT) images analysed for textural radiomics features [197]. Unlike this study, the ROIs were delineated on the CT images using semi-automation, without reference to the surgical specimens. The textural features that most strongly correlated with PIMO staining were the standard deviation of all pixel values and the mean value of positive pixels.

In the current analysis, GLCM_inversevar was the third most prominent feature in both the radiomics-only model (β=-1.910, p=0.179) and combined radiomics-clinical parameters model (β=-0.946, p=0.549).

A recently published pilot study on the association between hypoxia gene profiles and multiparametric MRI in prostate cancer identified 16 T2-weighted textural features which had significant correlations with 3 hypoxia gene profiles using both sequencing and immunohistochemical techniques [198]. GLCM, GLRLM, GLSZM and local binary patterns were extracted from multiparametric MRI and 3 hypoxia-related gene sets were selected for investigation, which were a 15-gene universal hypoxia gene set (GENE15), a 32-gene prostate specific hypoxia gene set (GENE32) and a 44-gene high frequent hypoxia gene set (GENE 44), the latter including HIF1A and VEGFA. Although limited to 6 patients, the results hold promise for future larger scale analysis.

It is apparent that there is structural heterogeneity in tumours that exhibit hypoxia which cannot be determined by traditional analysis of conventional imaging techniques. All studies cited above have illustrated that radiomics features extracted from T2-weighted MR images have potential in predicting for PIMO score or hypoxia gene profiles. As textural features quantify the occurrence of various patterns in an image and indicate how the signal response at any given voxel relates to those around it [46], these could be considered to

have potential as a surrogate for the structural heterogeneity displayed by the hypoxic tumour regions.

In the current radiation therapy workflow, even where ART is routine, a methodology to determine regions of hypoxia and thereby the likely sites of treatment failure is largely absent. Knowledge of the specific radiomics features that can predict for surrogates of hypoxia, which, from the current analysis and the literature, are likely to be texture-based features, gives the potential for radiomics-based adaption at the treatment unit, specifically the MRI-linear accelerator. A radiomics based targeted radiotherapy (Rad-TRaP) framework for prostate cancer brachytherapy and external beam treatment planning has been proposed previously [31]. However, the radiomic features in this Rad-TRaP study predicted only the location of the prostate cancer lesions, not the regions of hypoxia within the lesions. Such a framework however could be extended in future for the purposes of analysing and adapting therapy based on radiomics features predictive of hypoxia. In nasopharyngeal carcinoma, it has been postulated that utilising pre-treatment MRI images can predict patients whose treatment will require adaption during therapy. A T2-weighted model had an AUC of 0.750 when predicting for eligibility for ART in nasophayngeal carcinoma, with 6 selected radiomics features - sphericity, elongation, kurtosis and three textural features from the GLDM and NGTDM families [199]. However, this focused on adaption based on geometric or dosimetric features only and not on the presence of hypoxia.

The current analysis was performed on T2-weighted MRIs that had been acquired for diagnostic purposes, so before such a proposal could be deemed feasible, a similar analysis would have to be conducted on the available sequences on the MR-linear accelerator and confirmed in a validation cohort. Adoption of radiomics analysis into cone beam computed tomography (CBCT) workflow is dependent upon the ability to define the ROI on images known to be subject to limited quality and noise, making the MR-linear accelerator a more favourable option. Although it has been determined that some radiomics features are robust to CBCT [200], CBCT radiomics analysis to date has been to ascertain prognosis rather than to adapt or intensify treatment, particularly in lung cancer [201], which is similar to radiomics analysis in positron emission tomography (PET) [202].

5.4.2 PREDICTION OF GLEASON SCORE AT BIOPSY

In this T2-weighted analysis, a model with Stats_Var and GLCM_inverseVar was identified to predict Gleason score, with Stats_Var having the largest β coefficient of -7.049 (p=0.061). It is interesting to note that the second order textural feature GLCM_inverseVar, which was also in the PIMO prediction models, was a component of the model predicting for Gleason score with a β value of -1.438 (p=0.034). When Gleason score was dichotomised into Gleason 7 (4+3) and above versus Gleason (3+4) and below, the model yielded an AUC of 0.580 on the ROC curve.

There are conflicting data in the literature on this topic. Wibmer et al [102], also using linear regression analysis in their 3-T Haralick texture analysis (Energy, Entropy, Correlation, Difference Homogeneity, Contrast) of 147 patients who had biopsy-proven prostate cancer and were treated with radical prostatectomy reported that T2-weighted textural features, could differentiate benign from cancerous tissues but could not discriminate between Gleason scores. Of note is that the delineation of the tumour region of interest in the study of Wibmer et al also utilised the whole prostate specimen as guidance, as in the current analysis. However, in a follow up study, using the same cohort but a machine learning automatic classification, Fehr et al [203] were able to discriminate between Gleason 3+4 and Gleason 4+3. Also utilising the same method of ROI delineation, an Italian study of 49 evaluable prostate cancer cases reported significantly different Spearman correlation coefficients between GLCM features of Gleason score 6 lesions and Gleason score $\geq$ 7 lesions [204].

Gnep et al [96] reported significant Spearman's correlation coefficients in 74 patients between Gleason score and the textural features Difference Entropy 90°, Difference Variance 45° and Difference Variance 90° on T2-weighted MRI analysis. The potential for multicollinearity between the latter two features identified should be noted. The tumour ROI in this study was delineated using multiparametric MRI and histological reports only and, unlike the current analysis, all tumours arose solely in the peripheral zone.

Ginsburg et al [41], in a study of 80 patients from 3 institutions, also did not find any correlation between Gleason score and radiomics features extracted from T2-weighted MR images. However, the classifier under investigation in this study

was specifically designed to distinguish benign from cancerous prostate tissue only. The same group has postulated that radiomic features for detecting prostate cancer differ between transition and peripheral zones [41].

Analysis of a 99-patient dataset from The Cancer Imaging Archive (TCIA) found that GLSZM features of large size zone emphasis, zone size non-uniformity and [205]zone size emphasis can discriminate between three groupings of Gleason score (Group 1: Gleason $\leq$ 6, Group 2: Gleason 3+4, Group 3: Gleason $\geq$ 4+3) [206]. In this study, the ROI defined was automatically based on the centroid coordinates of all lesions in the T2-weighted images that defined this dataset. Others have reported a mean AUC for T2-weighted prediction of Gleason score of 0.739 in a cohort of 33 patients whose ROIs were delineated using 3D Slicer and who went on to receive IMRT for their prostate cancer . Chen et al [207] found that a T2-weighted radiomics classifier could outperform the PI-RADS v2 in distinguishing high-grade and low-grade prostate cancer with an AUC of 0.808. Of note is that the AUC greatly improved when the synthetic minority oversampling technique (SMOTE) to generate a sample from the joint weighting of multiparametric features was applied relative to the initial logistic regression model analysis, which yielded an AUC of 0.618, which is closer to the result of the current analysis with an AUC of 0.580.

Overall, there is no clear consensus on whether radiomics features derived from T2-weighted MRI can predict Gleason score, with some authors supporting the results of the current analysis and others not. The AUC for the current analysis was low at 0.580. The methods used to define the ROI for radiomics feature extraction (based on prostatectomy specimens or on multiparametric MRI with histology reports only, or on pre-defined centroid coordinates of lesions) as well as the Gleason score defined by samples gained at biopsy, which are heterogeneous [208], are likely reasons for the cause of these differing results as is considerable overlap between Gleason scores [102]. Interestingly, the ability of advanced statistical and machine learning methods to interrogate the data in a more comprehensive manner than traditional statistical methods should also be considered, as seen in the analysis of the same dataset by Wibmer [102] and Fehr [203] and the conflicting results achieved as well as the improvement in AUC achieved by Chen et al when moving from simple logistic regression to SMOTE

[207]. This is a consideration for future research on the dataset used in the current analysis.

5.4.3 PREDICTION OF CLINICAL TUMOUR (T) STAGE

A model with the radiomics features Stats_Var, GLCM_inverseVar and Shape_areaDensityBB predicted for clinical T stage. Similar to the models for PIMO score and Gleason score, GLCM_inverseVar is once more present in this model with a β value of -35.832 (p=0.001), second only to Stats_Var (β=-57.670, p=0.304). The GLCM feature family expresses how combinations of discretised grey levels of neighbouring pixels (2D) or voxels (3D) are distributed along one of the image directions [86]. In an analysis of T2-weighted MRI radiomics classifiers that included 13 GLCM features, including GLCM_inverseVar, a moderate AUC of 0.625 for the prediction of prostate stage was reported [205] . In the current analysis when clinical stage was stratified into T3a and above versus T2c and below, the AUC on ROC analysis was equivocal (0.501). The presence or absence of extracapsular extension (ECE) is an important determinant in the staging of prostate cancer and subsequent treatment decision making. Those found to have unilateral or bilateral ECE are staged as T3a; therefore this was used for stratification purposes on ROC analysis but the model could not discriminate between these patients and those with less advanced disease. A recently developed least absolute shrinkage and selection operator (LASSO) model on T2-weighted images found that ECE could be predicted with an AUC of 0.906 in a training dataset of 74 patients and with an AUC of 0.821 in the corresponding validation dataset of 45 patients. The radiomics features in the model included 3 GLCM features, one of which was GLCM_variance [209], again indicating that this feature appears to hold potential in determining aggressive features of prostate cancer.

Similar results have also been reported in other tumour sites such as rectal cancer. Using LASSO techniques to select radiomics features from 1029 extracted on T2-weighted images of 152 patients with rectal cancer, Ma et al [210] trained 6 machine-learning classifiers with the radiomics signature and found that the multilayer perceptron classifier predicted T stage with an AUC of 0.809, sensitivity of 76.2% and specificity of 74.1%. GLCM, GLRLM and GLSZM textural features, together with shape and first order features, formed this radiomics signature for

T stage. However, others have found using a similar supervised machine learning technique that the ability to discriminate between clinical T1-T2 and T3-T4 in rectal cancer had a mean AUC of 0.706 but a 95% confidence interval of 0.487-0.925 [211]. In other clinical sites such as non-small cell lung cancer (NSCLC), authors have reported that the combination of a radiomics signature, extracted from tumoral and peri-tumoral lung parenchyma ROIs on CT can add to the prediction of likely postsurgical recurrence compared with the prediction from TNM staging alone [212]. Morphological, intensity based and textural features (GLCM, GLRLM, GLSZM) formed the radiomics signature in this study.

5.4.4 PREDICTION OF D'AMICO RISK CLASSIFICATION SCORE

Previous work has illustrated the correlation between suspicious lesions on multiparametric MRI and D'amico risk classification in 101 patients[213] as well as the added benefit of including MRI data to D'amico risk groups to predict those who will fail treatment [214]. These provided the rationale for hypothesising that radiomics features would therefore have a relationship with risk classification score, as was found in this analysis. The D'amico risk classification focuses on 3 clinical parameters: initial PSA, clinical Gleason score and clinical Stage. Low risk patients are classified as those with PSA<10 ng/mL, Gleason < 7 and Stage T1-T2a. Intermediate risk are those patients with PSA of between 10-20 ng/mL, Gleason 7 or Stage T2b. High risk patients are those with an initial PSA > 20ng/mL, Gleason 8-10 or Stage >T2c [215]. Gleason score as a determinant of high-risk disease has many limitations as there is considerable variability in prostate tumour cellularity that cannot be considered by Gleason score alone, especially where benign and cancerous tissues overlap [216].

A 15-feature radiomics model significantly predicted risk stratification score in this cohort of 88 patients, accounting for 26.4% of the variation in risk stratification in the cohort and 12.4% in the population. The features with the largest β values in the model were Stats_Var (β=-3.959, p=0.541), Shape_area_Density_BB (β=1.411, p=0.087), HGLDM_LGE (β=-0.982, p=0.04) and GLCM_InverseVar (β=-0.867, p=0.201). On ROC analysis, the model resulted in an AUC of 0.753 for the prediction of high risk disease versus intermediate or low risk. Again, the second order feature GLCM_InverseVar and the first order

feature Stats_Var are components of the model for risk score, as in all other clinical parameters assessed in this analysis.

Only one recent prostate cancer study has investigated the relationship between textural featural analysis on MRI and D'amico risk classification [4]. It specifically found a significant negative correlation between T2-weighted texture features of skewness and kurtosis and D'amico score.

5.4.5 CONCLUSION

The aim of this chapter was to extract radiomics features from T2-weighted images and to examine the predictive potential of these features for specific clinical parameters. Statistically significant models were generated for prediction of PIMO score, Gleason score at biopsy, clinical T stage and D'amico risk classification score. ROC analysis of these models identified their predictive potentials ranging from equivocal for clinical T stage (AUC=0.501) to moderate for D'amico risk classification score (AUC=0.753). The first order feature Stats_Var and the second order feature GLCM_InverseVar featured in all models predicting for PIMO score, Gleason score at biopsy, clinical T stage and D'Amico risk stratification score, indicating that they may have potential in the stratification of aggressive from non-aggressive disease in prostate cancer. The literature largely supports these results, particularly for GLCM features, although the quantity of data is modest and is conflicting for Gleason score. Given the difference in the clinical management of patients who present with Gleason 3+4, rather than Gleason 4+3 disease, further radiomics analysis on Gleason score is warranted in future studies; however the known limitations of Gleason score in the distinction of low from high-risk disease may negate the benefit of such analysis.

The most clinically meaningful results from this chapter are that radiomics features derived from T2-weighted imaging have the potential to predict for PIMO score, Gleason score at biopsy and D'amico risk stratification. However, it must be cautioned that these results have not been validated in a separate cohort, which is a significant limitation of any biomarker research. Nevertheless, the potential to predict for such markers of aggressive disease may impact on future workflow of radiation therapy and in determining the patients who should receive intensified treatment regimens, as well as those who should be considered for less active treatment and warrants further investigation.

CHAPTER 6. STUDY 4: RADIOMICS FOR PREDICTION OF CLINICAL FEATURES IN PROSTATE CANCER: DIFFUSION-WEIGHTED MRI AND APPARENT DIFFUSION COEFFICIENT MAP ANALYSES

6.1 INTRODUCTION

After an initial negative standard Trans Rectal Ultrasound (TRUS) biopsy, many men present with persistently increased PSA levels [217]. This may be due to the low positive predictive value of PSA, which accounts for a high proportion of unnecessary biopsies [218-220]. However, the negative predictive value (NPV) of a TRUS-guided biopsy has been reported as being between 36% and 89% [192], with some patients also being intolerant of this invasive procedure [62]. Where a negative biopsy is associated with continued suspicious clinical symptoms (continued rising PSA > 4.0 ng/mL, suspicious digital rectal examination, abnormal PSA velocity), repeat biopsies suffer an even greater risk of negative rate for cancer detection than the initial one (81-83%) [7], with second, third and fourth re-biopsies detecting cancer in only 25% to 27%, 5% to 24% and 4% to 21% of cases, respectively [221].

Efforts to improve the rates of cancer detection of the traditional sextant biopsy have resulted in the initiation of extended biopsies, where attention is given to the lateral peripheral zone as well as the anterior apex [184]. The considerable limitations of TRUS biopsy have led to investigation of the potential of DW-MRI and ADC maps in assisting in the detection of prostate cancer.

Standard MRI for prostate cancer includes high-resolution T2-weighted images, which allow for assessment of the prostate and the tumour. DW-MRI images should be analysed together with the T2-weighted images and the ADC maps. When ADC maps and DW-MRI are considered together, they can provide functional tissue characterisation, which supplements the anatomical and morphological data given by T2-weighted imaging.

A prostate tumour will typically illustrate high signal intensity because of diffusion limitation on high *b*-value images and yield a low ADC value, meaning that it appears darker on ADC maps. This correlates with higher cellularity of prostate tumours than normal prostate tissue on histology. ADC maps can be useful because T2 ‘shine through’ from the normal high signal peripheral zone on the DW-MRI images can be problematic. This occurs in tissues with long T2

relaxation times and the strong T2 signal may be incorrectly assumed to be due to restricted diffusion [61]. The variances in reported ADC values for normal prostate tissue is likely due to the use of different *b*-values, field strengths, imaging protocols and MRI scanners.

The ability to detect a lesion on DW-MRI depends on the tumour site, its size and composition. Tumours <5 mm are difficult to detect and inflammatory processes in the prostate can produce low ADC values and mimic prostate cancer, leading to false positives [222]. It is particularly difficult to detect tumours in the transition zone and here ADC maps, in conjunction with T2-weighted images, can improve their detection [223].

Franiel et al. [7] examined 178 areas suspicious for prostate cancer, where the prostate was divided into 20 zones to coincide with histological reporting. Some 64% of these were at the peripheral zone, with the remainder (36%) in the central gland. Fifty-three areas were positive for prostate cancer; of these, 53% were in the peripheral zone and 47% were in the transition zone. A median ADC value of 1.03 x 10^{-3} mm^2/sec was measured for prostate cancer in the peripheral zone. In this region normal prostate tissue had a median ADC of 1.42 x 10^{-3} mm^2/sec and prostatitis 1.26 x 10^{-3} mm^2/sec. For the transition zone, median ADC for prostate cancer was recorded as 0.76 x 10^{-3} mm^2/sec, normal prostate tissue was 0.96 x 10^{-3} mm^2/sec and prostatitis 1.06 x 10^{-3} mm^2/sec.

AbdelMaboud et al. [224] reported on 36 patients where the data of T2W and DW-MRI were correlated with histology post-TRUS-guided biopsy and/or prostatectomy. ADC values for the prostate tumours were similarly decreased relative to the healthy prostate tissue (0.737 ± 0.154 X $10^{-3}$$mm^2$/s and 1.484 ± 0.289 X $10^{-3}$$mm^2$/s, respectively). Afaq et al. [59] report highest mean ADC values of 1.54-2.99 x10^{-3} mm^2/s in the peripheral zone, followed by the central gland (0.9 -2.14 x10^{-3} mm^2/s) and then prostate cancer (0.8 -1.66 x10^{-3} mm^2/s). Again, this group noted that there was some overlap between the ADC values for prostate cancer and normal prostate tissue. Chan et al [225] reported benign mean peripheral zone ADC values as 1.6 x10^{-3} mm^2/s and mean ADC values in cancer tissue as 1.43 x10^{-3} mm^2/s.

6.1.1 SENSITIVITY AND SPECIFICITY OF DW-MRI IN THE DETECTION OF PROSTATE CANCER LESIONS.

In an attempt to determine whether sensitivity and specificity were increased when combining DW-MRI with T2-weighted imaging relative to T2-weighted imaging alone, Morgan et al [226] compared diagnoses based on imaging with the results of sextant biopsies. Each sextant's histology was compared with the T2-weighted images alone and then the T2-weighted and ADC maps together. Sensitivity, specificity, positive and negative values were calculated and compared. A kappa value for interobserver agreement was calculated for T2-weighted imaging alone and T2-weighted images and ADC maps together. There was an improvement in overall accuracy, sensitivity and specificity for the most experienced observer with the addition of the ADC maps to the T2-weighted images. The sensitivity for tumour detection by an experienced observer improved from 46.5% to 71% on inclusion of the ADC maps in the evaluation. However, T2-weighted imaging alone showed a moderate overall interobserver agreement (Kappa 0.53), which was surprisingly better than for T2-weighted imaging and ADC maps (Kappa 0.334) ($p<0.001$).

A meta-analysis on the value of diffusion-weighted imaging in the detection of prostate cancer speculated that, given the aggressive nature of peripheral zone (PZ) tumours, a separate protocol specific to PZ tumours might lead to improved accuracy [227, 228]. Such a protocol may have merit as 70-75% of prostate cancers arise in the PZ [227, 228]. The PZ data were analysed separately for 8 studies. Sensitivity was increased relative to the entire gland (79% compared to 62%). However, the overall diagnostic accuracy was not improved, as there was still significant heterogeneity in the 8 trials chosen for this subgroup analysis.

Sato et al [229] studied prostate cancer detection with T2-weighted and DW-MRI in comparison to those cancers detected by T2-weighting alone and found that the combined imaging modality had an area under the receiver operating characteristic curve (AUC) of 0.89 in comparison to 0.81 of T2-weighted alone. However, the addition of DW-MRI to T2 weighted MRI improves the positive predictive value (PPV) and specificity of prostate cancer diagnosis, particularly in transitional zones [230].

Arsov et al [231] performed a study in patients with prior negative transrectal ultrasound-guided biopsy (TRUS-GB) and increased PSA. This illustrated that targeted biopsies based on functional MRI findings were superior to the traditional transrectal saturation biopsies. In 37.9% (22/58), functional MRI showed evidence of prostate cancer. Sixteen of the 22 patients underwent TRUS-GB with additional targeted biopsies. Prostate cancer was diagnosed in 68.8% (11/16). Follow up was 49.1 weeks. There was no significant elevation in PSA or detection of significant PSA increase in patients with normal functional MRI findings. The study showed a reduction in the mean number of biopsy cores required, relative to what would be required for traditional transrectal saturation biopsies.

A meta-analysis [232] showed no difference in overall detection between targeted and random biopsies; however the MRI targeted biopsy (MRI-TBx) had a higher rate of detection of significant prostate cancer (sensitivity 0.91, 95% CI 0.87-0.94), compared to TRUS biopsy (0.76, 95% CI 0.64-0.84) and a lower rate of detection of insignificant prostate cancer (sensitivity 0.44, 95% CI 0.26-0.64), compared to TRUS biopsy (0.83, 95% CI 0.77-0.87). Differentiation between significant and insignificant prostate cancer was ascertained using either Epstein criteria or Gleason score.

Willis et al [233] compared two diagnostic pathways; TRUS guided biopsy for all with multiparametric MRI (mpMRI) for all, followed by MRI-targeted biopsy, if positive. The study suggested that mpMRI and consequent MRI-targeted biopsy might result in biopsies that are more accurate and fewer in number as compared to TRUS-guided biopsies. The former not only accurately detects more clinically significant cancers but also correctly identifies more men without clinically significant disease in probabilistic sensitivity analysis.

6.1.2 RADIOMICS STUDIES ON DW-MRI AND APPARENT DIFFUSION COEFFICIENT MAPS

DW-MRI analysis in prostate cancer is mainly based on visual interpretation and ADC analysis on threshold optimisation. Intratumoural heterogeneity that is imaged by DW-MRI and ADC is currently not assessable using such methods. Radiomics analysis is however, well placed to quantitatively exploit these sequences and use this information on heterogeneity to plan for therapy [8].

The spatial resolution of functional MRI is important in the ROI delineation process [234], which impacts downstream on radiomics feature extraction. DW-MRI images can be of low spatial resolution and are sensitive to motion and magnetic susceptibility. They are also dependent on *b* -values with deviation of signal decay at high values, and dependent on TE and diffusion time [38, 171]. Variability of approximately 5% has been reported in ADC measurements across examinations and platforms under phantom analysis [235]. Despite these limitations, delineation of the dominant lesion in prostate cancer on DW-MRI and ADC maps has previously been validated with histopathology [236].

It has been demonstrated that radiomics analysis of multiparametric MRI, including DW-MRI, improves the performance of the PI-RADS in determining aggressive from indolent disease [237]. Many studies in prostate cancer have shown the ability of radiomics features derived from ADC maps to distinguish between benign and malignant disease [102] as well as correlation of radiomics features from ADC maps with Gleason score [83, 203, 204]. Correlation between ADC values and hypoxic level has been reported in B I 6FI murine melanoma tumours, where hypoxic level was determined using HIF1-α antibody staining [238].

Therefore, DW-MRI and ADC maps provide imaging data about the tumour microenvironment in prostate cancer that may be further exploited with radiomics analysis, potentially assisting with decision making about biopsy location and the intensification or de-intensification of therapeutic approaches [49].

6.1.3 AIMS AND OBJECTIVES

The aim of this chapter is to establish whether radiomics features from DW-MRI and Apparent Diffusion Coefficient (ADC) maps can predict specific clinical features.

The objectives are:

- To determine if ADC values are correlated with clinical parameters, including PIMO score.

- To identify the principal components of extracted radiomics features from the prostate index lesion following feature reduction on DW-MRI and ADC maps.

- To build radiomics models from both image sequences that are predictive of PIMO score, Gleason score at biopsy, D'amico risk stratification score and clinical tumour (T) stage.

6.2. MATERIALS AND METHODS

6.2.1. DATASETS AND RADIOMICS FEATURE EXTRACTION

The acquisition of the DW-MRI and ADC datasets and subsequent ROI delineation have been discussed in Chapter 4. 104 DW-MRI datasets and 94 ADC datasets were available for radiomics analysis using RadiomiX (RadiomiX Research Toolbox version 20180831 (OncoRadiomics SA, Liège, Belgium)) software. As in Chapter 4, the index lesion was the ROI on which feature extraction was conducted and this was delineated on a single axial slice on both DW-MRI and ADC images by an experienced radiologist using the histological specimen as a guide. Feature extraction was performed after the same pre-processing parameters were applied as the analysis on T2-weighted MRI dataset in Study 3.

6.2.2 ADC VALUES

The mean ADC value of the index lesion for each patient dataset was calculated using OSIRIX LITE™ Digital Imaging and Communications in Medicine (DICOM) viewer. The values were subsequently analysed in multiple linear regression analysis, together with radiomics and clinical features, for determining a potential model of hypoxia as well as in linear regression analysis to determine whether it was independently predictive of hypoxia.

6.2.3 STATISTICAL ANALYSIS

Statistical analysis was performed on both datasets as previously described in Study 3. In short, feature reduction was performed followed by multiple linear regression and ROC analyses.

6.3 RESULTS

6.3.1 PATIENT CHARACTERISTICS

The clinical features of all 104 DW-MRI and 94 ADC analysed patient datasets were given in Chapter 4. A summary of both is given below.

DW-MRI Patient Characteristics

The mean age was 64 years (range 47-76 years). The mean pre-op PSA was 11.9 (range 2-145). The mean Gleason score at biopsy was 7 (range 6-9). The Clinical (c) T stages were as follows: cT1c (n=44), cT2a (n=8), cT2b (n=8), cT2c (n=7), cT3a (n=9), cT2x (n=15), cT3x (n=7) and no cT stage was recorded for 6 patients. 3 patients were classified as low risk, 52 as intermediate and 49 as high risk as per the D'amico risk classification. The mean PIMO score was 2.7 (range 0-5) and 12 patients in this cohort had not received PIMO.

ADC Patient Characteristics

The mean age was 64 years (range 45-75 years). The mean pre-op PSA was 13.3 (range 2-145). The mean Gleason score at biopsy was 7 (range 6-9). The Clinical (c) T stages were as follows: cT1c (n=37), cT2a (n=9), cT2b (n=9), cT2c (n=6), cT3a (n=9), cT2x (n=13), cT3x (n=4) and no cT stage was recorded for 6 patients. 2 patients were classified as low risk, 46 as intermediate and 45 as high risk as per the D'amico risk classification. The mean PIMO score was 2.8 (range 0-5) and 11 patients in this cohort had not received PIMO.

6.3.2 ADC VALUES CALCULATED

The mean ADC value for the cohort was $0.829 \pm 0.225 \times 10^{-3}$ mm^2/sec.

6.3.3 PRINCIPAL RADIOMICS FEATURES IDENTIFIED ON DW-MRI AND ADC

A principal components analysis (PCA) was run on the 165 extracted components for both DW-MRI and ADC datasets. The suitability of PCA was assessed prior to analysis. Inspection of the correlation matrices showed that all variables had at least one correlation coefficient greater than 0.3, which illustrated linearity between all variables.

DW-MRI PCA Analysis

PCA revealed 14 components that had eigenvalues greater than one and which explained 71.312%, 31.853%, 13.892%, 7.880%, 6.817%, 5.285%, 3.543%, 2.944%, 2.5%, 1.86%, 1.573%, 1.468%, 1.082% of the total variance, respectively. Visual inspection of the scree plot indicated that 14 components should be retained and a forced factor number PCA on these 14 components was conducted.

The 14-component solution explained 93.536% of the total variance (Table 6.1). Pearson correlations were conducted on the 14 retained PCs and the original 165 extracted features. The radiomics features which had the highest Pearson correlation coefficients with the PCs were as follows: 1 from first order statistics (Stats_ min), 3 from the grey level size zone feature family (GLSZM_IV, GLSZM_INN, GLSZM_LAE), 1 from the grey level run length feature family (GLRLM_RLV), 2 from the neighbouring grey tone difference matrix (NGTDM_busyness, NGTDM_contrast), 3 from the intensity histogram family (IH_skewness, IH_kurtosis, IH_minGradl,) and 4 based on shape features (Shape_leastaxislength, Shape_surface, Shape_areaDensityBB, Shape_elongation). These were inputted as independent variables in subsequent regression analysis.

ADC PCA Analysis

PCA yielded 19 components that had eigenvalues greater than one and which explained 68.732%, 34.153%, 10.739%, 8.056%, 5.771%, 4.981%, 4.026%, 3.369%, 2.957%, 2.489%, 2.260%%, 1.954%, 1.464%, 1.277%, 1.176%, 1.081%, 0.922%, 0.847%, 0.714% of the total variance, respectively. As previously, visual inspection of the scree plot indicated that 19 components should be retained and a forced factor number PCA with 19 components was conducted.

The 19-component solution explained 95.133% of the total variance (Table 6.2). Pearson correlations were conducted on the 19 retained PCs and the original 165 extracted features. The radiomics features which had the highest Pearson correlation coefficients with the PCs were as follows: 1 from first order statistics (Stats_ min), 2 from the grey level size zone feature family (GLSZM_IV, GLSZM_LAE), 1 from the grey level run length feature family (GLRLM_RLV), 4 from the grey level distance zone matrix (GLDZM_INN, GLDZM_DZNN,

GLDZM_LDE, GLDZM_SDE), 2 from the neighbouring grey tone difference matrix (NGTDM_busyness, NGTDM_contrast), 4 from the intensity histogram family (IH_skewness, IH_kurtosis, IH_minGradl, IH_min) and 5 based on shape features (Shape_leastaxislength, Shape_surface, Shape_areaDensityBB, Shape_elongation, Shape_volnumber). These were considered independent variables in subsequent regression analysis.

Total Variance Explained									
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	71.312	43.219	43.219	71.312	43.219	43.219	52.111	31.582	31.582
2	31.853	19.305	62.524	31.853	19.305	62.524	26.930	16.321	47.904
3	13.892	8.419	70.944	13.892	8.419	70.944	20.927	12.683	60.587
4	7.880	4.776	75.720	7.880	4.776	75.720	20.479	12.412	72.998
5	6.817	4.131	79.851	6.817	4.131	79.851	5.480	3.321	76.320
6	5.285	3.203	83.054	5.285	3.203	83.054	5.066	3.070	79.390
7	3.543	2.147	85.202	3.543	2.147	85.202	5.018	3.041	82.431
8	2.944	1.784	86.986	2.944	1.784	86.986	3.755	2.276	84.707
9	2.500	1.515	88.501	2.500	1.515	88.501	3.496	2.119	86.825

10	2.323	1.408	89.909	2.323	1.408	89.909	2.983	1.808	88.634
11	1.860	1.127	91.037	1.860	1.127	91.037	2.545	1.542	90.176
12	1.573	.953	91.990	1.573	.953	91.990	2.205	1.336	91.512
13	1.468	.890	92.880	1.468	.890	92.880	1.847	1.119	92.632
14	1.082	.656	93.536	1.082	.656	93.536	1.492	.904	93.536
15	.965	.585	94.120						
16	.883	.535	94.656						

Table 6.1. Total Variance Explained from Principal Components Analysis of features for DW-MRI. *The full table of all 165 features is available in Appendix 4.*

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	68.732	41.656	41.656	68.732	41.656	41.656	54.754	33.184	33.184
2	34.153	20.699	62.355	34.153	20.699	62.355	25.914	15.706	48.890
3	10.739	6.509	68.864	10.739	6.509	68.864	18.269	11.072	59.962
4	8.056	4.882	73.746	8.056	4.882	73.746	15.097	9.150	69.112
5	5.771	3.497	77.243	5.771	3.497	77.243	5.972	3.619	72.731
6	4.981	3.019	80.262	4.981	3.019	80.262	5.039	3.054	75.785
7	4.026	2.440	82.702	4.026	2.440	82.702	4.632	2.808	78.592
8	3.369	2.042	84.744	3.369	2.042	84.744	4.234	2.566	81.158
9	2.957	1.792	86.536	2.957	1.792	86.536	3.580	2.169	83.328
10	2.489	1.509	88.045	2.489	1.509	88.045	3.303	2.002	85.330
11	2.260	1.370	89.415	2.260	1.370	89.415	2.932	1.777	87.107

12	1.954	1.184	90.599	1.954	1.184	90.599	2.744	1.663	88.770
13	1.464	.887	91.486	1.464	.887	91.486	2.051	1.243	90.013
14	1.277	.774	92.260	1.277	.774	92.260	1.641	.995	91.007
15	1.176	.713	92.973	1.176	.713	92.973	1.453	.881	91.888

Table 6.2. Total Variance Explained Table from Principal Components Analysis of features for ADC maps. *The full table of all 165 features is available in Appendix 5.*

6.3.4 PREDICTIVE RADIOMIC MODELS BASED ON DW-MRI AND ADC MAPS

6.3.4.1 PREDICTION OF PIMO SCORE

On linear regression analysis, ADC value did not independently predict for PIMO score (p=0.579).

A multiple linear regression was run to assess the predictive potential of radiomics features for PIMO score. There was homoscedasticity in both DW-MRI and ADC results, indicating that the variance of the residuals was constant across all the values of the independent variable as seen in Figure 6.1 for ADC and Figure 6.2 for DW-MRI. There was also normality of the residuals as seen in Figure 6.3 for ADC and Figure 6.4 for DW-MRI. There was independence of residuals, as assessed by a Durbin-Watson statistic of 2.052 for DW-MRI and 1.899 for ADC. As for the T2-weighted analysis in Study 3, similar assumption tests were run for all other linear regression analyses and comparable results achieved.

A radiomics model with all 14 radiomics features significantly predicted PIMO score in the DW-MRI dataset, $F(14, 82) = 2.072$ $p = 0.025$ accounting for 29.9% of the variation in PIMO score with adjusted $R^2 = 15.5$ %. The features with the highest β coefficients in the model were GLRLM_RLV (-3.939, p=0.347), GLSZM_INN (β=-3.664, p=0.215) and Stats_min (β=-1.458, p=0.054) (Table 6.3). This model yielded an area under the ROC curve of 0.582 (Figure 6.5). In order to ascertain if the model would change when combined with clinical features, a further regression, including clinical parameters was run. The clinical features included were: pre-operative PSA level, clinical T stage, Gleason score at biopsy and D'amico risk classification and lymph node status of patients. Homoscedasticity and normality of the residuals were once again observed. The resultant model included only the radiomics features Stats_min, GLSZM_INN and GLRLM_RLV $F(3, 82) = 3.191$, $p = 0.028$, accounting for 10.8% of the variation in PIMO score with adjusted $R^2 = 7.4$ %. Table 6.4 illustrates that GLRLM_RLV had the largest β coefficient (3.080, p=0.087).

For ADC analysis, a model with 13 radiomics parameters significantly predicted hypoxia $F(13,82) = 1.905$ $p = 0.044$, accounting for 26.4% of the variation in

hypoxia score with adjusted R^2 = 12.5%. The model included the following radiomics parameters: GLSZM_IV, GLSZM_LAE, GLRLM_RLV, GLDZM_INN, NGTDM_busyness, NDTDM_contrast, IH_skewness, IH_kurtosis, IH_minGradl, Shape_leastaxislength, Shape_surface, Shape_areaDensityBB and Shape_elongation. The most prominent features in the model according to β coefficient were GLDZM_INN (-4.736, p= 0.117), GLRLM_RLV (-2.019, p=0.636) and Shape_elongation (-1.496, p=0.16). When the clinical parameters and ADC values were added to the model, the model was not significant and contained 3 radiomics features only (GLRLM_RLV, GLDZM_INN, Shape_areaDensityBB) $F(3,82) = 2.673$, p= 0.053 and explained 9.2% of the variation in PIMO score with an adjusted R^2 = 5.8%. β coefficient in this model was highest for GLRLM_RLV (4.157, p=0.018).

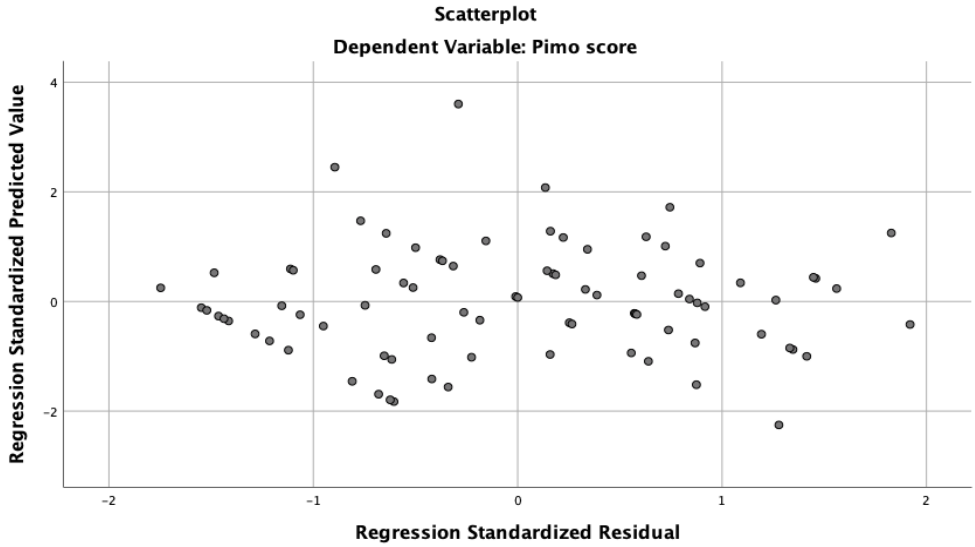


Figure 6.1. Homoscedasticity of residuals for ADC where PIMO score is the dependent variable.

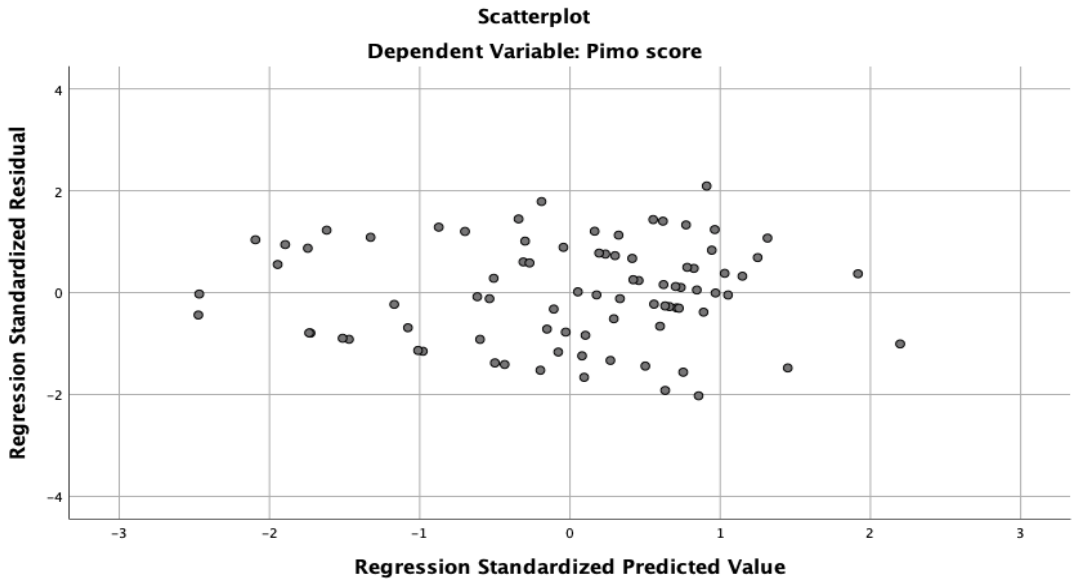


Figure 6.2. Homoscedasticity of residuals for DWI where PIMO score is the dependent variable.

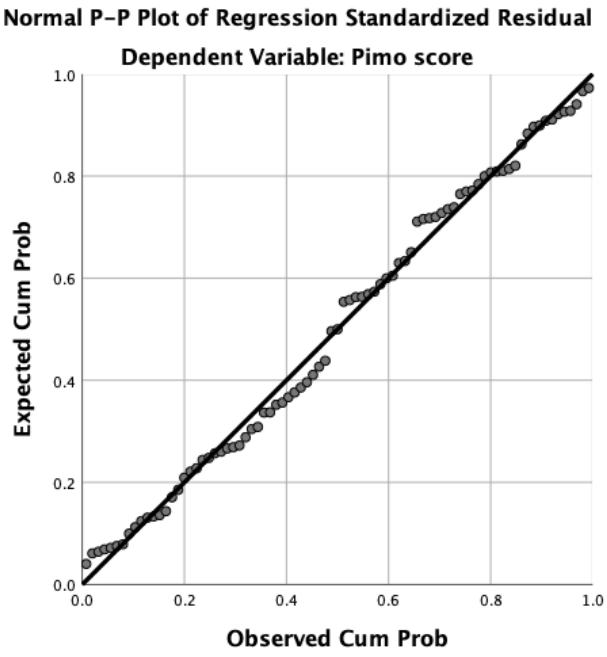


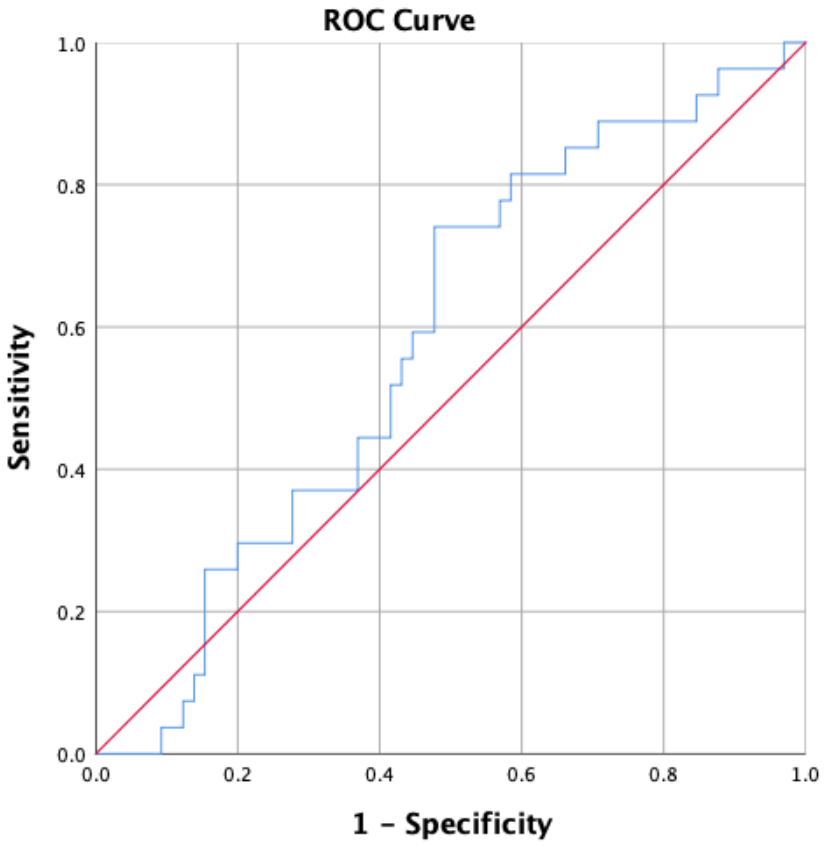
Figure 6.3. Normal Probability Plot for ADC with dependent variable PIMO score illustrating normality of residuals.

Figure 6.4. Normal Probability Plot for DW-MRI with dependent variable PIMO score illustrating normality of residuals.

Model		Unstandardised Coefficients		Standardised Coefficients	t
		B	Std. Error	Beta	
	(Constant)	3.572	1.848		1.933
	Stats_min	-1.458	.742	-.270	-1.964
	GLSZM_IV	.080	.075	.201	1.069
	GLSZM_INN	-3.664	2.930	-.231	-1.251
	GLSZM_LAE	.019	.011	.451	1.720
	GLRLM_RLV	-3.939	4.162	-.253	-.946
	NGTDM_busyness	-2.208E-6	.000	-.012	-.113
	NGTDM_contrast	-.510	1.459	-.062	-.350
	IH_skewness	.355	.418	.133	.849
	IH_kurtosis	-.201	.175	-.161	-1.152

	IH_minGradI	-.260	.128	-.389	-2.030
	Shape_leastaxislength	-.560	.597	-.123	-.938
	Shape_surface	.015	.065	.045	.231
	Shape_areaDensityBB	1.219	2.276	.077	.536
	Shape_elongation	-1.241	1.044	-.166	-1.189

Table 6.3. Radiomics-only model parameters for prediction of PIMO score on DW-MRI



AUC= 0.582

Figure 6.5. ROC curve for prediction of PIMO score by radiomics-only model on DW-MRI.

Model		Unstandardised Coefficients		Standardised Coefficients	t
		B	Std. Error	Beta	
	(Constant)	1.757	.530		3.313
	Stats_min	-.809	.617	-.150	-1.311
	GLSZM_INN	-2.391	1.688	-.151	-1.417
	GLRLM_RLV	3.080	1.777	.197	1.733

Table 6.4. Radiomics and clinical parameter model parameters for prediction of PIMO score on DW-MRI.

Note that the final model includes radiomics features only

PIMO Score Radiomics only					
Model Coefficients		Unstandardised Coefficients		Standardised Coefficients	t
		B	Std. Error	Beta	
	(Constant)	3.914	1.866		2.097
	GLSZM_IN	.405	.418	.258	.969
	GLSZM_LAE	.018	.011	.444	1.652
	GLRLM_RLV	-2.019	4.248	-.130	-.475
	GLDZM_INN	-4.736	2.987	-.299	-1.586
	NGTDM_busyness	3.305E-6	.000	.018	.165
	NGTDM_contrast	-.323	1.202	-.039	-.269
	IH_skewness	.534	.410	.201	1.301
	IH_kurtosis	-.193	.178	-.155	-1.088
	IH_minGradI	-.153	.116	-.229	-1.318
	Shape_leastaxislength	-.663	.601	-.146	-1.102
	Shape_surface	-.062	.117	-.188	-.532
	Shape_areaDensityBB	1.103	2.337	.070	.472
	Shape_elongation	-1.496	1.052	-.200	-1.421

Table 6.5. Radiomics-only model parameters for prediction of PIMO score on ADC maps.

PIMO Score Radiomics-Clinical Model					
Model Coefficients		Unstandardised Coefficients		Standardised Coefficients	t
		B	Std. Error	Beta	
	(Constant)	2.669	1.140		2.341
	GLDZM_INN	-2.764	1.733	-.175	-1.594
	Shape_areaDensityBB	-.982	1.779	-.062	-.552
	GLRLM_RLV	4.157	1.725	.267	2.410

Table 6.6. Radiomics and clinical parameter model parameters for prediction of PIMO score on ADC maps.

Note that the final model includes radiomics features only

6.3.4.2 PREDICTION OF GLEASON SCORE AT BIOPSY

There was a statistically significant, moderate negative correlation between ADC value and Gleason score at biopsy, $r(91) = .37$, $p = 0.002$ (Figure 6.6). No radiomics feature was found to statistically significantly predict for Gleason score at biopsy for either DW-MRI or ADC datasets.

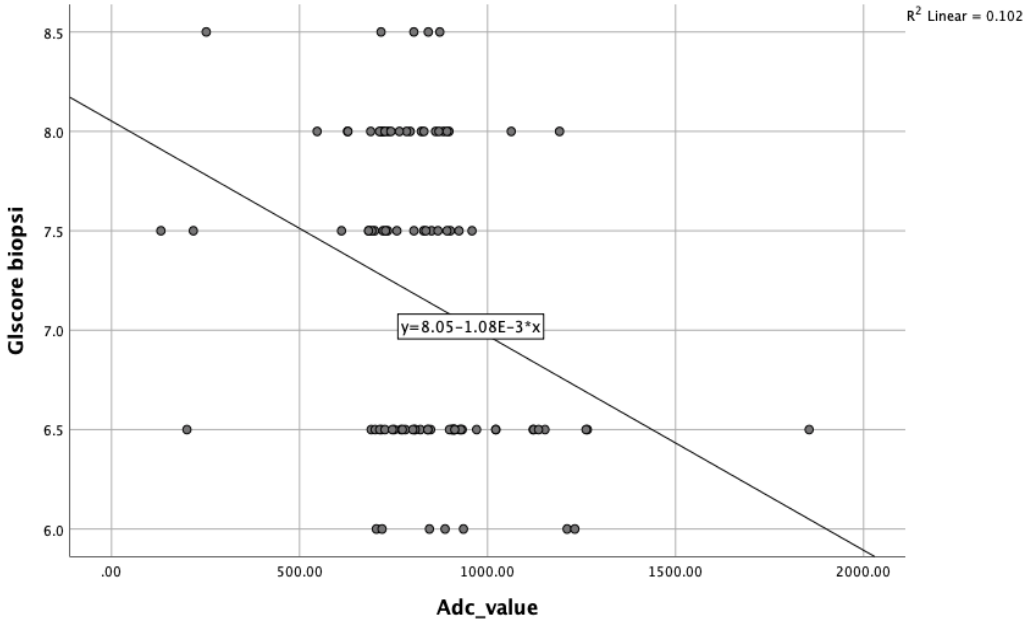


Figure 6.6 Scatterplot of Gleason score at biopsy versus ADC value.

6.3.4.3 PREDICTION OF CLINICAL TUMOUR STAGE

No radiomics model was found to statistically significantly predict for clinical T stage regression on DW-MRI or on ADC datasets.

6.3.4.4 PREDICTION OF D'AMICO RISK STRATIFICATION SCORE

A radiomics model significantly predicted for D'Amico risk stratification on DW-MRI and included the following radiomics features: GLRLM_RLV, GLSZM_INN, Shape_leastaxislength, GLSZM_IV, GLSZM_LAE, NGTDM_busyness, NGTDM_contrast, IH_skewness, IH_kurtosis, IH_minGradl, Shape_surface, Shape_areaDensityBB and Shape_elongation $F(13,103)=2.005$, $p = 0.029$ accounting for 22.5% of the variation in risk stratification, with adjusted R^2 = 11.3%. GLRLM_RLV had the highest β coefficient in the model (9.814 p=0.129), followed by Shape_area_Density BB (β=-1.328, p=0.119) (Table 6.7). The area

under the ROC curve for the model was 0.519 (Figure 6.7). For ADC, Shape_areaDensityBB only predicted for risk stratification score $F(1,93)= 4.468$, $p = 0.037$ accounting for 4.6% of the variation in risk stratification, with adjusted $R^2 = 3.6\%$ (Table 6.8). The model yielded an AUC of 0.638 (Figure 6.8).

D'amico Risk Stratification Score					
Model		Unstandardised Coefficients		Standardised Coefficients	t
		B	Std. Error	Beta	
	(Constant)	2.899	.535		5.419
	GLRLM_RLV	9.814	6.400	.313	1.533
	GLSZM_INN	-1.193	1.548	-.141	-.770
	Shape_leastaxislength	-.316	.226	-.160	-1.397
	GLSZM_IV	.001	.001	.102	.757
	GLSZM_LAE	-.072	.032	-.405	-2.296
	NGTDM_busyness	-.095	.094	-.117	-1.008
	NGTDM_contrast	-.005	.006	-.099	-.827
	IH_skewness	.223	.129	.198	1.728
	IH_kurtosis	-.033	.078	-.047	-.426
	IH_minGradI	.001	.008	.013	.101
	Shape_surface	.015	.022	.098	.646
	Shape_areaDensityBB	-1.328	.843	-.180	-1.576
	Shape_elongation	.884	.373	.263	2.370

Table 6.7. Radiomics model parameters for prediction of D'amico risk stratification score on DW-MRI.

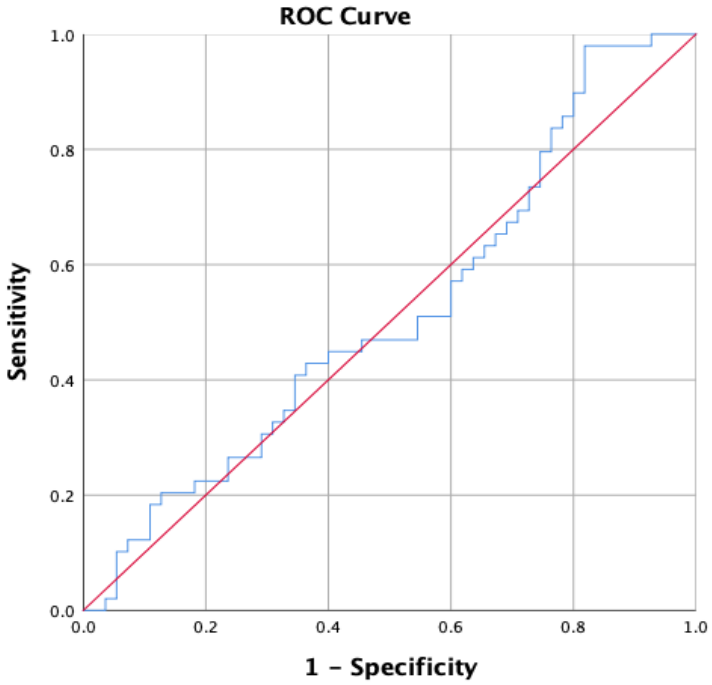
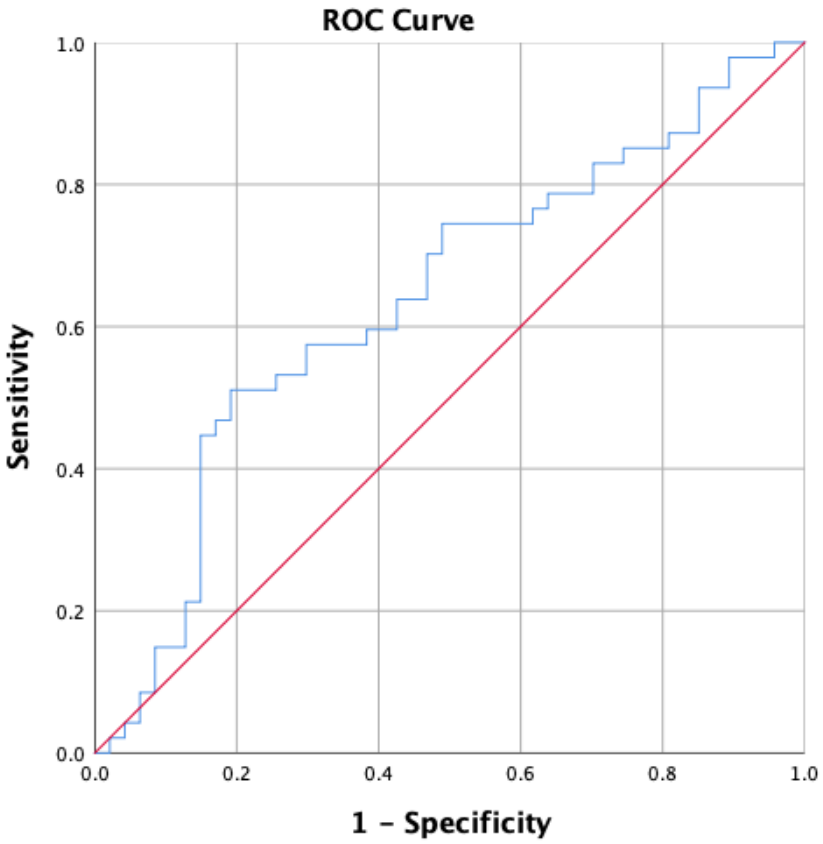


Figure 6.7. ROC curve for prediction of D'amico risk stratification score by radiomics model on DW-MRI.

D'amico Risk Stratification Score					
Model		**Unstandardised Coefficients**		**Standardised Coefficients**	**t**
		B	**Std. Error**	**Beta**	
	(Constant)	1.659	.387		4.289
	Shape_areaDensityBB	1.438	.680	.215	2.114

Table 6.8. Radiomics model parameters for prediction of D'amico risk stratification score on ADC maps.



AUC= 0.638

Figure 6.8. ROC curve for prediction of PIMO score by radiomics model on ADC maps.

6.4 DISCUSSION

Functional MRI techniques image intra-tumoral heterogeneity, but not the entire tumour microenvironment. Currently the analysis of these images and associated ADC maps is under-exploited using traditional visual evaluation methods. Therefore, the aim of this chapter was to establish whether radiomics features from DW-MRI and Apparent Diffusion Coefficient (ADC) maps could predict PIMO score, Gleason score at biopsy, clinical T stage and D'amico risk stratification score. First, the ADC values across the patient population were analysed.

6.4.1 ADC VALUE ANALYSIS

The mean ADC value for the cohort was 0.829 ± 0.225 X 10^{-3} mm^2/sec, which is line with previous reports [59, 224]. Linear regression analysis did not result in a significant prediction of PIMO score by ADC value in the current study. However, other studies utilising different markers of hypoxia have reported positive results. In a study of 39 patients with prostate cancer, it was found that the tumour ROI delineated on ADC maps had a negative correlation with the expression of HIF-1α [239]. Others reported in ex vivo mouse models moderate correlations between hypoxic level, as also given by the expression of HIF-1α, and ADC values (ρ=0.56, p<0.00001) [238].

ADC values have the potential to be important prognostic disease markers in identifying tumour aggressiveness. Currently, this is conducted using Gleason score on a biopsy specimen only, which has the associated limitations of sampling errors [192], as well as being an invasive procedure [62]. The principle behind using ADC maps in this manner is that altered gland formations such as medullary or solid patterns are likely to show impeded water diffusion and be associated with an increased histological Gleason grade [223, 240]. In this study, a statistically significant moderate negative correlation between ADC value and Gleason score was found, which supports this principle. ADC values for low-grade tumours have been reported as 1 ±0.23 mm^2/s, 0.7 ±0.17 mm^2/s for intermediate grade tumours, and approximately 0.5±0.13 mm^2/s for high grade tumours [221]. Similarly, Turkbey et al [43] reported that low ADC values were associated with higher grade tumours and therefore higher Gleason scores.

6.4.2. RADIOMICS PREDICTION OF PIMO SCORE

Imaging hypoxic tumour regions for adaption of radiation therapy for prostate cancer is currently challenging. Blood Oxygen level-dependent MRI (BOLD-MRI) is one non-invasive technique for mapping prostate tumour hypoxia where the patient is imaged on a conventional MRI scanner and the contrast is endogenous, paramagnetic deoxyhaemoglobin [76]. This technique has the drawback that the patient is not imaged in the treatment position nor at the time of radiation therapy delivery. As a commercially available MRI-linear accelerator has recently received 510(k) premarket notification from the U.S. Food and Drug Administration (FDA) [241], the potential to discriminate between tumours that are responding to radiation therapy during treatment and those that are not and require treatment adaption based not on geometrical differences as is the current practice but on tumour biology is fast becoming a clinical reality.

It has been demonstrated in mouse models that hypoxic fractions are proportionate to PIMO binding [78]. PIMO score was significantly correlated with pathologic Gleason score ($p<0.05$), largest tumour extent ($p<0.00005$) and pathologic tumour grade ($p<0.00005$) for the patient cohorts included in the current analysis [120], indicating that increased hypoxia and increased tumour aggressiveness are linked. The current analysis found that a 14-feature radiomics model from the DW-MRI dataset and a 13-feature model from the ADC dataset significantly predicted for PIMO score. The DW-MRI radiomics-only model yielded the highest prediction with an AUC of 0.582. The textural feature GLRLM_RLV featured prominently in both of these radiomics models as well as in the combined radiomics and clinical parameter models. The GLRLM feature family assesses the distribution of discretised grey levels in an image and specifically run lengths and GLRLM_RLV estimates the variance in runs over the run lengths [86]. GLRLM features from CT images have previously been found to predict for survival in 203 patients with inoperable non-small cell lung cancer (NSCLC) [242].

6.3.4.2. PREDICTION OF GLEASON SCORE AT BIOPSY

No radiomics feature on DW-MRI or on ADC could predict for Gleason score at biopsy, which is in contrast to the literature. Gnep et al [96] found that the first

order feature ADC median was significantly correlated with Gleason score in a 3T analysis of 74 patients. The ADC analysis was performed on ROIs that were propagated from those delineated on T2-weighted imaging. In a study of 33 prostate cancer patients using machine learning methods, ADC radiomics models were found to be predictive of Gleason score with a mean AUC of 0.739 [205]. Others have reported similar results with an AUC of 0.888 in an ADC radiomics model differentiating Gleason score $\leq$ 6 versus Gleason score$\geq$ 7 in 182 prostate cancer patients. The ADC radiomics model included features Correlation, GLCM_Entropy, Cluster_Prominence, Inertia (n=2), Cluster_Shade, Uniformity, Kurtosis and Variance. The same study reported that PI-RADS v2.0 yielded an AUC for 0.763 in the same Gleason score analysis [207]. Negative correlations between ADC textural radiomics features and Gleason score have also been reported [4]. It is likely that the difference in the results of the current study compared to the literature are first due to differences in the delineation of the ROI utilised for radiomics feature extraction. In the current study, the index lesion was delineated directly onto the ADC maps and diffusion-weighted images utilising the prostate specimen and on one axial slice only. Neither ADC thresholding nor propagation of ROIs from T2-weighted images were utilised. Second, it is likely that contrasting *b* values were utilised in this study compared with others. This study utilised diffusion-weighted images taken with a *b* value of 1500 s/mm^2 and ADC maps manually generated from *b* values of 200 to 800 s/mm^2, whereas in all other studies, the ADC maps were automatically generated at the MRI at time of scanning and their values are not always stated.

6.3.4.3. PREDICTION OF CLINICAL T STAGE

No radiomics model from either DW-MRI or ADC datasets predicted for clinical T stage in this analysis. However, ADC models in the machine learning study of Abdollahi et al [205] were predictive of overall stage in prostate cancer with an AUC of 0.675. In bladder cancer, a 19-feature model (4 features from T2 weighted image analysis and 15 from DWI analysis) could differentiate non-muscle-invasive from muscle-invasive carcinomas, which directly influences T stage [243].

6.3.4.4 PREDICTION OF D'AMICO RISK STRATIFICATION SCORE

The radiomics feature Shape_AreaDensityBB from the ADC dataset analysis was found to predict for D'amico risk stratification score with an AUC of 0.638. For the DW-MRI analysis, a 13 feature model with GLRLM_RLV and Shape_AreaDensityBB as the two most prominent features yielded an AUC of 0.519. Shape features derived from ADC maps have been reported to predict clinically significant from non-clinically significant prostate cancer in 75 patients. When ADC maps and T2-weighted images were combined, Shape features resulted in an AUC of 0.78 for the detection of clinically significant prostate cancer [244]. Others have found correlations between D'amico risk stratification and textural features in a small cohort of 26 patients extracted from T2-weighted imaging, but not from ADC map analysis [4]. Three radiomic features from ADC maps (Laws and the first order statistics of Gradient and Sobel) have discriminated clinically significant prostate cancer better than PI-RADS v2.0 alone [47]. Two radiomics features from ADC maps have been reported as being able to discriminate prostate cancer from benign tissue in the peripheral zone and up to six radiomics features to discriminate in the transition zone. These features were distinct from one another, indicating the presence of zone-specific features. It should be noted however that more features were present in the model from T2-weighted images than from ADC maps in this study [41].

6.3.4.5 CONCLUSION

The aim of this chapter was to establish whether radiomics features from DW-MRI and Apparent Diffusion Coefficient (ADC) maps could predict PIMO score, Gleason score at biopsy, clinical T stage and D'amico risk stratification score. The textural feature GLRLM_RLV was present in 4 models predicting for PIMO score, with the DW-MRI radiomics-only model yielding the best predictive potential with an AUC of 0.582. D'amico risk stratification score was predicted by models which included the morphological feature Shape_AreaDensityBB, with AUCs of 0.519 and 0.638 for DW-MRI and ADC maps, respectively. Unlike the T2-weighted analysis, no radiomics model from DW-MRI or ADC map was predictive of clinical T stage or Gleason score. Similar to Study 3, the analysis is limited by the small sample size as well as the lack of a validation cohort.

CHAPTER 7. DISCUSSION

The aims of this thesis were, first, to determine if radiomics features from multiparametric MRI could predict specified clinical endpoints in prostate cancer and, second, to determine if the radiomics models identified could be considered reproducible under different imaging and pre-processing conditions.

This aim was achieved through four studies. In Study 1, an MRI-compatible 3D printed pelvic phantom was created. This was used to measure the impact of geometric distortion on the size of simulated ROIs in the prostate. The phantom was also used to determine which radiomics features were reproducible when matrix size, FOV and slice thickness are changed from baseline on T2-weighted MRI.

In Study 2, the impact of image pre-processing was considered in patient prostate cancer datasets across three MRI sequences (T2-weighted imaging, DW-MRI and ADC maps) used in the detection of prostate cancer. The pre-processing parameters analysed were image normalisation, voxel resampling and interpolation algorithm.

The results of Studies 1 and 2 informed Studies 3 and 4 in that the most reproducible imaging conditions and pre-processing parameters were adhered to before determining the predictive potential of radiomics features from T2-weighted imaging in Study 3 and DW-MRI and ADC maps in Study 4 for clinical features for prostate cancer. In both studies, these features were PIMO score (as a surrogate for hypoxia), Gleason score at biopsy, Clinical T stage and D'amico risk stratification score.

7.1. STUDY 1: REPRODUCIBILITY OF RADIOMICS FEATURES - A PHANTOM ANALYSIS.

In Study 1, the impact of imaging parameters on the reproducibility of radiomics features extracted from T2-weighted MRI was evaluated. This was achieved using a custom-made 3D printed pelvic phantom. 3D printing offers the potential to print various geometries at a low cost [148]. To date, most MRI phantom utilisation in radiation therapy has related to dosimetry, quality assurance and IGRT [127, 137, 148, 245]. A systematic review of the repeatability and

reproducibility of radiomic features did not include any phantom study of repeatable and/or reproducible features from MRI [246], making Study 1 novel in the field. One of the main challenges in producing an MRI-compatible phantom is utilising materials that possess an MR signal [151]. The phantom in Study 1 was modelled on previous irrevocably anonymised prostate cancer patient datasets. The phantom was printed using ABS, which is considered a 'negative mould', meaning that it must be filled with substances that have an abundance of water hydrogen protons in order to elicit an MR signal. Silicone was utilised to simulate the pelvic tissues, distilled water for the prostate gland and prostate lesions and a combination of silicone and 'chalk' (gypsum) to mimic the cortical bone of the femoral heads. Others have also utilised gypsum to simulate bony structures and silicone to mimic soft tissues [150].

7.1.1. GEOMETRIC DISTORTION

First, geometric distortion was considered. The impact of geometric distortion on the delineation process in MRI radiomics must be considered as the ROI delineated impacts downstream on the entire radiomics process. If radiomics analysis is to become part of the radiation therapy adaptive MR workflow, analysis of the impact of geometric distortion on delineation is pertinent, given that accurate geometry is a requirement for radiation therapy treatment planning [247].

Distortions in MRI are typically divided into system-related field distortions (non-linearity of the applied gradient magnetic fields and inhomogeneity of the static background magnetic fields, known as B_0 inhomogeneity) and patient-related field distortions (chemical shift and magnetic susceptibility) of the imaged tissues [158]. Modern MRI designs have focused on improving patient comfort, especially for those with claustrophobia, and therefore have wider and shorter bores. While patient-centred machine design is always to be welcomed, this innovation brings with it increased difficulty in the optimisation of the scanner, increasing geometric distortion as the slew rate increases [134]. B_0 inhomogeneity can be addressed by increasing bandwidth, but this reduces signal-to-noise ratio by using more localised volumetric shimming and by B_0 correction [247]. However, due to shimming coils in superconducting MRI

systems, static field inhomogeneity is expected to be small relative to the distortion from the gradient field non-linearity [134].

The American Association of Physicists in Medicine (AAPM) report 100 for MR Quality Assurance highly recommends that the geometric accuracy of clinical MRI scanners be determined [248]. Geometric distortion for MRI scanners is typically measured at the time of calibration using 3D phantoms with control point positions defined along all three orthogonal axes. The geometric distortion is then mapped by correlating the control point positions between the distorted image and their 'true' positions, defined by the phantom geometry. From the data at these control points, the geometric distortion at any other point in the phantom can be interpolated [134]. The current analysis was therefore a crude measurement of the impact of geometric distortion as in-depth methodology of geometric distortion measurement was beyond the scope of this study. Nevertheless, it was found that simulated prostate ROIs at the periphery of the field of view and at the furthest distance from the isocentre had the highest standard deviations when measured in their transverse diameter. For the simulated lesion at the isocentre, the range of measurements was 4.67-4.96 mm, (SD:0.075) and ranged from 4.1 to 6.36 mm for other simulated lesions, depending on their location in the phantom and their position in the field of view (Figure 7.1). The highest standard deviation for any one lesion was 0.49 mm. This is an interesting result as all the lesions in this study were simulated with 1cc of de-ionised water and yielded this large variation in diameter due to geometric distortion. The majority of prostate cancer tumour foci are between 0.5-1.5 cc [249] and it has been reported that 0.2 cc is the limit for tumour visibility on MRI [41], which may lead us to question the 'true' size of a lesion that is appreciable on MRI when the effect of geometric distortion is considered. Following routine calibration of MRI scanners, displacements of <0.1 mm near the isocentre of the magnet have been reported [250]. Therefore, similar to the findings of this study, lesions close to the isocentre suffer least from geometric distortion.

Distortions of up to 25 mm over a 24 cm field for 1.5T MRI have previously been reported [251, 252]. It has been reported that for targets with diameters up to 3cm, as are seen in prostate focal lesion boosts [253], D_{95} could be affected by spatial distortions of approximately 1.5 mm [159]. This is important if using an MR-only radiation therapy workflow as magnet and gradient specifications are

likely to be worse than on a diagnostic MRI scanner [160]. As demonstrated in this study, distortion is largest at the edge of the field of view. While it can be argued that prostate tumours are unlikely to be located at the edge of the field of view, it has been reported that even when the largest geometric distortion occurs at the body surface, internal organs still have some distortion and that this needs to be considered for GTV contouring [155]. It should also be recognised that the patient external contour will certainly be in this field of view and this can impact on the treatment planning process [156], particularly when modulated techniques are utilised and path-lengths vary [254].

This phantom did not consider the bladder and rectum, which, as deformable organs, can have enclosed air pockets. Such air pockets can change the amplitude and extent of susceptibility-induced geometric distortions on the prostate [255], but these were not simulated in this phantom, which can be considered a limitation.

In summary, geometric distortion uncertainties and their impact on delineation should be considered along with interobserver variation in delineation for future radiomics studies. However, based on the results of this study and the literature, ROIs at distance from the isocentre must be delineated with care, due to residual geometric distortion. 3D arrays of control points as tools for the development of new or improved geometric distortion correction algorithms have already been proposed [136]. Improved geometric correction algorithms will impact positively on MR-radiomics workflow in radiation therapy as ROI delineation will be more spatially accurate. This is turn will result in increased accuracy in radiomics feature extraction.

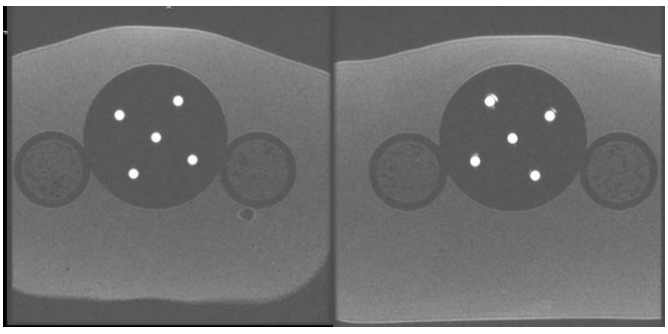


Figure 7.1. Geometric distortion as observed on phantom analysis. Note the difference in lesion size from centre of the field (left) to the edge of the FOV (right)

7.1.2. REPRODUCIBILITY OF RADIOMICS FEATURES

Image resolution and variations due to different reconstruction algorithms by different vendors lead to extensive variation in the data produced at image acquisition [38]. Partial volume effect, pitch, signal-to-noise ratio, resampling size and image quantification parameters can also vary significantly [256]. Artefacts can also influence model fit [257]. It has been proposed that such differences can lead to complexity in comparing and evaluating radiomics results gathered across institutions with different scanners and patient populations, particularly in relation to textural features [258]. Therefore, in Study 1 the aim was to determine which radiomics features were reproducible when matrix size, field of view and slice thickness were changed from baseline scanning parameters on T2-weighted MRI.

The phantom was scanned on a 1.5T GE Signa H-DXT MR Unit. The baseline scan parameters were TR: 5980 ms, TE: 120 ms, matrix: 256 X 192, FOV: 17 cm^2, slice thickness: 4 mm and slice gap: 0.5 mm. The scan was repeated three times, first changing the matrix to 320 X 256. Increasing the matrix size increases spatial resolution but decreases signal. Next, the FOV was changed from 17 cm^2 to 24 cm^2. Increasing the FOV decreases spatial resolution and increases voxel size. Finally, the slice thickness was changed from 4 mm to 7mm. Increasing slice thickness increases coverage, decreases spatial resolution and increases the signal-to-noise ratio but also increases the partial volume effect. Each change was made separately for the purpose of this analysis, but all three parameters are inextricably linked when considering spatial resolution of an MR image [123].

The coefficient of variation (COV) was used to compare the result of each radiomics feature extracted when matrix size, FOV or slice thickness was changed relative to that feature when imaged with the baseline parameters. The COV is frequently used for such phantom analyses in radiomics research [133, 163, 165]. The concordance correlation coefficient (CCC) is an alternate statistical test to assess whether feature values are reproducible under varying conditions and could also have been utilised in this analysis [175]. CCC ranges from 0 to 1, with 1

being perfect agreement. Features with a CCC ≥ 0.9 are considered reproducible [183], but features have been considered reproducible with a CCC as low as ≥ 0.75 in another study [115].

The Shape feature family was found to be robust to changes in imaging parameters, with the COV remaining <10% whether the scanning parameter change was in matrix size, FOV or slice thickness. This is an important result for radiation therapy, as Shape features are dependent on the ROI delineated. It does, however, highlight the consequence of accurate delineation, and this is well known to be the 'weakest link' in the radiation therapy chain despite extensive use of consensus guidelines for each anatomic site [90]. It has previously been reported that interobserver variability in delineation using CT images for NSCLC, head and neck cancer and malignant pleural mesothelioma radiomics feature extraction resulted in stability of radiomics features of 90%, 59% and 36%, respectively [259]. Given the superior soft tissue contrast of MR relative to CT, autodelineation of ROIs on MRI has the potential to replace manual delineation methods [156]; atlas delineation for MRI-based IMRT treatment planning has already been reported [260]. In CT radiomics, semiautomatic delineation methods have been reported as being closer to tumour surgical resection sizes than manual delineation methods [261] and radiomics features were reported to be more stable when derived from semiautomatic delineation methods than from manual methods [262]. Fractal average was also found to be a reproducible feature. Fractal-based texture analysis examines the difference between pixels at different length scales [208].

Changing the FOV and matrix size in this analysis did not have a negative impact on the reproducibility of features, with 27/31 (87%) of features having a COV<10% following a change in FOV and 29/31 (93.5%) of features having high reproducibility, even when the matrix size was increased. Therefore, the decrease in signal that comes with increasing matrix size did not impact on the reproducibility of the features extracted from the image datasets. Conversely, in a positron emission tomography (PET) radiomics study, matrix size impacted significantly on the reproducibility of features, with 56% of 100 features extracted being very sensitive to matrix changes (COV ≥ 20%) [133]. In the current study, the increased FOV decreased spatial resolution and increased voxel size but this also did not have an impact on the reproducibility of the majority of features.

This is an interesting result as the voxels were deliberately not resampled prior to radiomics feature extraction. A CT phantom analysis also indicated that when FOV was increased from the range 25-55 cm^2 to the range 45-55 cm^2, the number of features having a COV <10% increased from 125/177 to 151/177 and for a COV <15% increased from 138/177 to 160/177 [165]. It has also been reported that texture features are sensitive to variations in MRI acquisition parameters that increase spatial resolution but if the spatial resolution is sufficiently high anyhow, variation in acquisition parameters have little or no effect on textural analysis [164], which explains the results of the current study. We can therefore conclude that increases in matrix size and FOV overall did not affect feature reproducibility in Study 1 as spatial resolution remained significantly high.

Increasing the slice thickness from 4 mm to 7 mm did have a negative effect on the reproducibility of features. Only 12/31 (38.7%) features were considered reproducible with a COV< 10%. The features most highly impacted upon were the textural features GLCM_ClusShade, GLSZM_SZN, NGTDM_busyness and the intensity feature IH_skewness. Zhao et al [263] reported similar results in a CT phantom analysis comparing the effect of changing from 1.25 mm to 5 mm slice thickness on 14 features, with textural features being the most sensitive. Similarly, in another CT phantom analysis, when slice thickness range was changed from 2-8 mm to 3-5 mm, the 10% threshold COV of radiomics features improved from 112/177 to 119/177 and the 15% threshold improved from 132/177 to 136/177 features [165]. It has been demonstrated in lung cancer that retaining the same reconstruction algorithm but varying the slice thickness from 1.25mm to 2.5 mm yielded the best agreement in investigations of the concordance of combinations of smooth and sharp reconstruction algorithms together with varying slice thicknesses from 1.25mm to 5mm. Unsurprisingly, the worst results were observed when comparing the thinnest slice thickness of 1.25mm with the sharp reconstruction algorithm to the largest slice thickness of 5mm with the smooth reconstruction algorithm [169].

In conclusion, repeatability and reproducibility of radiomics features under various imaging conditions is fundamental to their routine implementation in the clinical setting [183]. This phantom analysis established that increased matrix size and FOV did not affect the reproducibility of radiomics features on T2-weighted MRI but increased slice thickness had a negative impact on the

reproducibility of more than 60% of the 32 features analysed. A phantom analysis, while useful at providing a 'ground truth' analysis, does not however consider the effect of biological and physical parameters inherent in heterogeneous tumours on radiomics feature extraction [133]. Study 2 therefore moved from phantom to patient reproducibility analysis.

7.2. STUDY 2: IMPACT OF IMAGE PRE-PROCESSING ON REPRODUCIBILITY OF RADIOMICS FEATURES -A PATIENT ANALYSIS

The aims of Study 2 were to calculate the reproducibility of radiomics features in patient cohorts across MR sequences when image pre-processing parameters were varied and, based on these results, to make recommendations about which features were most robust to changes in image pre-processing. The MR sequences analysed were T2-weighted imaging acquired with the PROPELLER technique, DW-MRI acquired at a *b* value of 1500 s/mm^2 and ADC maps created using *b* values from 200 to 800 s/mm^2. As the DW-MRI dataset available for analysis had been acquired at a high *b* value and the ADC dataset over a lower range of *b* values, both were included in this study and subsequently in Study 4, for the same reason.

The ROI utilised for radiomics feature extraction on all sequences was the prostate index lesion, delineated by a radiologist on a single axial slice with reference to the gross prostatectomy specimen and histopathologic reports, making it as close to 'ground truth' as possible. It is well understood that prostate cancer is generally multifocal and consists of a dominant region, this being the index lesion, as well as one or more secondary smaller foci of tumour. Currently, the entire prostate gland is routinely treated in radiation therapy, which can lead to chronic toxicities for patients that may affect their quality of life [264]. It has been proposed that smaller foci of disease could be considered as non-malignant, as there is no robust evidence to suggest that they meet the criteria of the six hallmarks of cancer [69], and that they could even be termed 'indolent lesions of epithelial origin (IDLE) [69] as non-index, secondary lesions rarely exhibit Gleason score 4 or higher [265]. Epidemiological evidence also supports this as a third of men have prostate cancer that will not adversely affect them during their lifetimes [266]. Similarly, it has been suggested that the index

lesion is the lesion most responsible for tumour progression [267] and it may be the lower grade regions inside the index lesion that are responsible for the lethality of prostate cancer [268]. Therefore, focusing the radiomics features extraction on the index lesion instead of the entire prostate gland in this analysis was considered legitimate.

As in Study 1, 33 radiomics features were analysed in Study 2 from the following feature families: Fractal (n=2), GLCM (n=4), GLDZM (n=3), GLRLM (n=4), GLSZM (n=4), IH (n=4), NGLDM (n=3), NGTDM (n=1), Shape (n=4), Stats (n=4). First, all three MR sequence datasets were pre-processed for the 'baseline' radiomics results, to which the results from the changed pre-processing parameters were compared. Images were normalised to zero mean unit variance, resampled to 3 mm isotropic voxels with slice spacing 3 mm, bin width was 0.1 and the linear interpolation algorithm was utilised. Computation of texture features requires binning of the image intensities into a limited number of grey levels [122]. A small bin width (0.1) is typically required to calculate textural features. If a large bin width (e.g. 25) is selected, the textural features are lost as the ROI is completely or almost completely observed as homogenous and is discretised in one or two bins only [117]. A fixed bin width rather than a fixed bin count was therefore utilised, as has been recommended in the literature [122, 132].

Maintaining consistent isotropic spacing is considered important for reproducibility in radiomics research; however there is no clear indication as to whether up-sampling or down-sampling is preferable. Down-sampling can introduce artificial information while up-sampling may result in a loss of information as well as the introduction of image aliasing artifacts [86]. Isotropic resampling is also preferable to non-isotropic resampling for the calculation of textural features [132]. As the variability in feature values is lower when the voxel size is equal for all images, isotropic resampling was conducted in this study for the baseline analysis [168].

Linear interpolation utilises the intensities of the closest 8 voxels in the original grid to calculate a new interpolated intensity [86] Other interpolation algorithms that may be utilised are tricubic and nearest neighbour. As no consensus exists on the most suitable interpolation algorithm and linear interpolation is seen as a conservative option, it was utilised for the baseline radiomics feature extraction.

Intraclass correlation coefficients (ICC) were calculated to determine the level of concordance in feature values following each pre-processing change [41]. The ICC was first introduced as a modification of the Pearson correlation coefficient, but here is calculated using mean squares obtained through analysis of variance (ANOVA). As the three MR sequences analysed were from the same patient dataset and ICC values are known to give a good estimate of comparison between features and between changes in image processing parameters within the same dataset [171], they were considered optimal for analysis in Study 2. Similarly, ICC is invariant to linear scaling and shifting, which is important in radiomics analysis where features operate in different units and scale spaces and their 95% confidence interval provide clarity on the size of the heterogeneity between features [122].

When normalisation of all MR sequences was not performed prior to radiomics feature extraction, the reproducibility of features relative to baseline was extremely poor. On ADC maps, no feature reached even moderate reproducibility. On DW-MRI, fractal average reached moderate reproducibility on ICC value alone (ICC=0.538), but as its 95% confidence interval illustrates significant heterogeneity, it cannot be deemed reproducible. On T2-weighted imaging, IH_ kurtosis (ICC=0.882) and IH_skewness (ICC=0.860) had excellent reproducibility. Shape_elongation, Shape-flatness, Shape_volume_density_BB and Stats_skewness had perfect reproducibility with ICCs of 1. It has been reported that Shape_flatness cannot accurately be assessed based on a single slice 2D radiomics extraction, however [122]. Therefore its result in Study 2 is not considered reliable. Intensity-based features describe how grey levels within the ROI are distributed [86] and Shape-based features illustrate its morphological characteristics, describing the intensity distribution within the ROI without considering the spatial relationship between neighbouring voxels [269]. These results highlight the absolute necessity of performing normalisation prior to radiomics feature extraction on MR as MR signal intensities are relative and cannot be compared between images [116]. It has been found in CT radiomics that images that are not normalised prior to feature extraction do not convey information about texture or other intervoxel relationships [163] and this was observed in the results of Study 2, where essentially only Shape features and intensity-based features were reproducible, without image normalisation. The reproducibility of Shape features on T2-weighted MRI without image

normalisation is expected as Shape features rely heavily on delineation and are expected to be more robust against voxel level intensity variations between images [119, 171]. An interesting finding is that these Shape features were reproducible on T2-weighted MRI and not on DW-MRI and ADC maps. As T2-weighted MRI provides morphologic information on the prostate, it can be postulated that its delineation, based on the prostate specimen and histopathologic report, was more easily transferable to an image that depicts the prostate morphology rather than its function. Shape features on T2-weighted imaging in a cervical cancer patient cohort have been reported as highly repeatable and reproducible under test-retest conditions when scanned on different MR units and when contoured by different observers [116]. Furthermore, similar results for Shape features on test-retest conditions have also been reported in rectal cancer [170].

When resampling of voxels was not performed, overall reproducibility of features was poor across all three MR sequences, particularly for ADC maps. All three sequences had moderate to excellent reproducibility for first order Stats features. This result is not unexpected as first order features can be considered global rather than local statistical descriptors and have been reported as generally reproducible with respect to resampling in rectal cancer [172].

Others have previously reported that voxel size has a significant impact on repeatability of textural features on T2-weighted MRI [166]. Of note in Study 2 was the difference in results between ADC maps, which at *b* values between 200 and 800 m/s^2 did not have any reproducible textural features, while on high *b* value DW-MRI (1500 m/s^2), one textural feature was considered reproducible (GLSZM_IV).

When the nearest neighbour interpolation algorithm was utilised instead of linear interpolation, 26 features on ADC maps were reproducible, 19 on DW-MRI and 20 on T2-weighted imaging, a substantial improvement on the results seen from normalisation and resampling. All Stats features were considered reproducible on ADC maps and on T2-weighted MRI, but none on DW-MRI. NGLDM and NGTDM features were all considered reproducible on ADC maps but none on either DW-MRI or T2-weighed imaging. All Shape features were reproducible on ADC maps but only one each on DW-MRI and T2-weighted imaging. It can be speculated that the poor reproducibility of radiomics features on DW-MRI, when

interpolation algorithm was changed, may be due to the combination of variations in signal intensity from differences in radiofrequency coil sensitivity and placement between patients [173]. Certainly, the same issue was not appreciated for ADC maps calculated from lower *b* values, compared with the absolute values of signal intensity on the DW-MRI dataset. Based on these results it is likely that if this study were repeated using different *b* values for ADC map calculation, different results would be reported. This has been illustrated in rectal cancer where from 31 *b* value combinations, the optimal *b* values for feature repeatability were 0 and 1000s/mm^2 [181].

To summarise, image pre-processing is a critical aspect of radiomics workflow. If image pre-processing is not performed, radiomics feature values can be meaningless because the values are dominated by image noise. This was observed for all textural features in Study 2 when normalisation was not performed. Optimal pre-processing reduces image noise while maintaining the relative heterogeneity of the ROI for model building [270]. From the results of Study 2, it is recommended that all MR sequences be normalised prior to feature extraction to ensure greyscale values are within the same range. Voxel resampling is recommended to reduce noise, particularly in big datasets where different scanners and baseline scan parameters may have been utilised. Selection of interpolation algorithm does not impact as dramatically on feature reproducibility as normalisation and resampling; however, the results of this study would not indicate any reason to select nearest neighbour algorithm over linear.

7.3. STUDIES 3 AND 4: RADIOMICS FOR PREDICTION OF CLINICAL ENDPOINTS IN MULTIPARAMETRIC MRI FOR PROSTATE CANCER.

7.3.1. STRATIFICATION OF PATIENTS FOR RADIATION THERAPY

Personalised radiation therapy for prostate cancer involves, first, the stratification of patients for treatment based on staging and, second, tailoring radiation therapy to individual tumour characteristics. Identifying those who need intensified treatment and those whose disease is indolent avoids unnecessary treatment and associated toxicity. Chronic incontinence has been reported in 5%-20% of prostate cancer radiation therapy patients, sexual dysfunction in 30%-

70% and bowel toxicity in 5%-10% [31], illustrating the importance of correct patient selection for therapy. As the majority of tests for staging are invasive, Studies 3 and 4 of this thesis sought to establish whether radiomics features from multiparametric MRI could maximise information from routine diagnostic imaging that could assist in the stratification of patients. This is important particularly as inter-observer variability is known to limit accuracy in the interpretation of multiparametric MRI in prostate cancer detection [12] and biopsy is geographically dependent. This means that biopsy-based prognosis largely depends on the location that was biopsied within the tumour [271].

Under-sampling, using transrectal ultrasound (TRUS)-guided biopsy, has been reported to occur for a substantial proportion (20%-60%) of men with prostate cancer [272, 273]. A meta-analysis [232] reported that MRI-targeted biopsy had a higher rate of detection of significant prostate cancer compared to TRUS-guided biopsy, as well as a lower rate of detection of insignificant prostate cancer. Although performing multiple biopsies could provide a better understanding of tumours, this has the drawback of increased patient discomfort, increased potential for complications, as well as an increase in cost. This highlights the urgent need for alternate methods of stratifying and staging prostate cancer in order to utilise radiation therapy in a more efficient manner and to improve prostate treatment outcomes [274]. As radiomics presents the opportunity to analyse images objectively and quantitatively, this limitation could potentially be overcome [31]. Radiomics has the potential to define areas for targeted biopsy and to give an indication of tumour grade, which should correlate with targeted biopsies [275]. Therefore, the first aim of Studies 3 and 4 was to establish if radiomics features could predict clinical endpoints that are currently used in patient staging and stratification for prostate cancer. This has recently been advocated for in lung cancer to allow for interpretation of radiomics results in a clinical context [276]. The clinical features analysed were Gleason score at biopsy, D'amico risk stratification score and clinical T stage.

ADC has been reported as being able to detect alterations in compartmental volumes of cancerous tissues, including epithelium, stroma and lumen space, as well as cellularity [277]. A statistically significant, moderate negative correlation was found between ADC value and Gleason score at biopsy, $r(91) = -0.37$, $p = 0.002$, a very similar result to Oto et al [278], who also found a moderate

negative correlation between ADC values and Gleason score (r=-0.376) but no correlation with tumour angiogenesis parameters (VEGF expression) in 73 patients. Similarly no correlation between PIMO score and ADC value was found in Study 4.

In Study 3 (T2-weighted imaging), a model with Stats_Var and GLCM_inverseVar was identified to predict Gleason score, with Stats_Var the most prominent feature in the model. It is interesting to note that the second order textural feature GLCM_inverseVar was also reported in the PIMO prediction models, which would support the hypothesis that this feature is predictive of tumour aggressiveness in prostate cancer. When Gleason score was dichotomised into Gleason 7 (4+3) and above versus Gleason 7 (3+4) and below, the model yielded an AUC of 0.580 on the ROC curve. This dichotomy was performed as patients with Gleason (3+4) have known better outcomes than those with Gleason (4+3) [279]. An AUC of 0.66 has previously been reported by Cheung et al [280] for the ability of Gleason score to predict risk stratification in prostate cancer and an AUC of 0.98 in the discrimination of Gleason 3+4 from Gleason 4+3 in a series of 506 patients, based on features extracted from CT images [281]. The high AUC yielded from CT datasets relative to MR datasets would indicate a correlation between Gleason score and Hounsfield units, representing the electron densities of the tissues, which are absent in T2-weighted MRI. No radiomics feature was found to statistically significantly predict for Gleason score at biopsy for either DW-MRI or ADC datasets. Others have found that both T2-weighted imaging and ADC models can predict for Gleason score but that the T2-weighted imaging models were more predictive [205]. The *b v*alues used for ADC calculation were not specified and may be different from those in Study 4.

Despite its widespread clinical use, Gleason score as a grading system is not without its flaws. There are limitations of the Gleason score to determine significant high-grade tumours, as much variability in prostate tumour cellularity is not considered in the current grading system [216]. The issue of the optimal treatment of Gleason 3+4 disease remains a conundrum for most clinicians. One proposal to overcome this is that Gleason 6 and 7 prostate cancers without stromogenic carcinoma can be considered indolent whereas Gleason 6 and 7 with stromogenic pattern cannot [282]. While this is not considered a standard of care in current clinical guidelines, it is an interesting proposal.

A model consisting of Stats_Var,GLCM_inverseVar and Shape_areaDensityBB significantly predicted Clinical Tumour (T) stage. GLCM features have previously been found to correlate with T stage on PET imaging for oesophageal cancer [283]. ROC analysis was equivocal when clinical T stage was stratified according to a threshold of T3a or above (T3a indicating unilateral or bilateral extracapsular extension and impacting on treatment decision), with an AUC of 0.501. Others have reported an AUC of 0.66 for clinical T stage as a predictor of prostate cancer risk [280]. No radiomics model was found to statistically significantly predict for clinical T stage on DW-MRI or ADC datasets. Based on these results, identification of clinical tumour stage with radiomics features cannot be recommended ahead of the current standard, the invasive digital rectal examination procedure.

The D'amico risk stratification score, while considered by some to be heterogeneous in terms of recurrence risk [96], is a validated method to determine a patient's pre-treatment prostate cancer specific mortality [213]. It assists in the stratification for patients in that those with low and intermediate risk disease can be offered more conservative treatment approaches, such as active surveillance and those with high risk disease are recommended for definitive treatment. D'amico risk stratification score considers Gleason score, clinical stage and PSA level. A radiomics model including all 16 radiomics variables from feature reduction significantly predicted D'amico risk stratification score on T2-weighted imaging. On ROC analysis, the model had moderate prediction, with an AUC of 0.753 when dichotomised into low or intermediate risk versus high risk. A radiomics model significantly predicted for D'amico risk stratification on DW-MRI with GLRLM_RLV having the highest β coefficient in the model, followed by Shape_area_Density BB. However, the area under the ROC curve for the model was low at 0.519. For ADC, Shape_areaDensityBB only predicted for risk stratification score, with an AUC of 0.638. These results may be unexpected as D'amico risk stratification score is a whole gland score and not just a score based on the defined ROI. These results add strength to Studies 3 and 4 in that the radiomics models reported can predict for a whole gland validated scale such as D'amico risk stratification. If validated in an external dataset, these results show promise that radiomics features extracted from T2-weighted imaging and ADC maps could assist in stratifying patients into risk categories. This in turn would determine which patients could be monitored by active surveillance and which patients need intensified treatment. Adding validated

radiomics features to existing nomograms for risk stratification in a trial setting would validate this potential, as has already been reported for some imaging features [284].

There is a paucity of literature that investigates the relationship between radiomics and PSA level, most likely because of the considerable known limitations of PSA as a marker in sensitivity, specificity, accuracy and positive predictive value [207]. PSA is organ-specific but not prostate cancer-specific and PSA levels cannot distinguish between aggressive and indolent disease [285]. In fact, the National Comprehensive Cancer Network (NCCN) has reported that only about 18% of men with a PSA level in the range 4-10 ng/mL will subsequently be diagnosed with cancer upon biopsy [286]. PSA as a predictor for any prostate cancer has been cited as yielding an AUC of 0.53 on ROC analysis [287]. A recent small study of 26 patients did not find any correlation between radiomics texture parameters and serum PSA [4]. Where radiomics and PSA level has been investigated is in the ability of radiomics features to predict those patients who should undergo biopsy who have a PSA of between 4-10 ng/mL, as cited by the NCCN. This was the subject of analysis in a recent Chinese study [288] which found that incorporating radiomics features into a fusion prediction model, together with age, PSA density and the PI-RADS score yielded an AUC of 0.956 on a primary cohort of 133 patients and 0.933 on a validation cohort of 66 patients. The T2-weighted imaging features included in the model were intensity, difference entropy and textural features.

7.3.2. TAILORING RADIATION THERAPY TO INDIVIDUAL TUMOUR CHARACTERISTICS.

The second aim of personalised radiation therapy is to tailor treatment delivery to each tumour's characteristics. Tumours regularly exhibit broad biological, cellular and tissue heterogeneity, with local variations in angiogenesis and hypoxia, leading to changes in glucose metabolism. These features determine the aggressiveness of a tumour [289]. Current adaptive radiation therapy (ART) methods in prostate radiation therapy are based on observed anatomical

variations only, with no data on such changes in tumour biology readily available during a course of fractionated radiation therapy. Radiomics, together with the arrival of the MR-linear accelerator to radiation oncology departments, have the potential to overcome this limitation [290]. Therefore the second aim of Studies 3 and 4 was to identify if radiomics features could predict for PIMO score, as a surrogate for hypoxia, a known cause of radiation therapy failure.

For the prediction of PIMO score, the most prominent radiomics features in the 7 feature T2-weighted dataset were NGLDM_SDE, Shape_flatness and GLCM_inverseVar. A high correlation within prostate tumours between PIMO staining grids and those tumours with high transverse relaxation rate signal on BOLD-MRI has previously been reported [76]. GLCM texture features from T2-weighted imaging have previously been associated with prostate cancer aggressiveness from the index lesion [204]. On DW-MRI the highest ranked features in the models predictive of hypoxia were GLRLM_RLV and GLSZM_INN and Stats_min and for ADC maps the features were GLDZM_INN, GLRLM_RLV and Shape_Elongation. On ROC analysis, only the models from T2-weighted MRI and DW-MRI had AUCs > 0.5; for T2-weighted MRI, an AUC of 0.591 and for DW-MRI, an AUC of 0.582 were reported. The model features from the NGLDM, GLCM, GLDZM and GLRLM families are textural features. Textural features are measurements computed using a texture matrix, with texture being the spatial arrangement of a predefined number of voxels (in the ROI) that allow the extraction of complex image properties [182]. Texture features are sensitive to voxel level variations [171] and therefore resampling prior to feature extraction was conducted across all three sequences in Studies 3 and 4 to account for this. Stats_min was featured prominently in the T2-weighted model for PIMO score and statistical descriptors are known to be dependent on intensity [291], hence the need for normalisation as part of pre-processing, which was performed in these studies. Shape features were reported in the models for both T2-weighted imaging and ADC maps. Shape features are heavily reliant on the ROI delineated and in Study 2 were seen to be the most reproducible. However, the reliability of Shape_flatness as a feature based on a 2D ROI has been questioned [122]. Future radiomics studies should consider fully automated tools for delineation as many textural features were reported as prominent in predicting PIMO score in Studies 3 and 4 and interobserver delineation differences have been reported as amplified in textural features [246].

In summary, from the preliminary work in this thesis and subject to model validation, it is hypothesised that radiomics features could be utilised for the identification of the inter-fraction hypoxic region. This would permit daily biological adaptation, rather than geometric adaption, as is the current 'gold standard' in ART. Such hypoxic sub-volumes could be boosted and treatment intensified on a patient by patient basis, thereby meeting the goal of personalised radiation therapy. One such case in a canine patient has been previously described where the hypoxic sub volumes were identified using Eppendorf electrodes and these regions were 'painted' using inverse treatment planning to yield a tumour control probability (TCP) three times higher than the uniform dose to the rest of the target [292]. Biological adapted radiation therapy in head and neck cancer using PET has already been reported [293]. Preliminary investigation of intra-treatment diagnostic MRI in head and neck cancer has shown radiation therapy-dependent changes in oxygenation levels [294], as have histogram features of ADC values of MRIs acquired during therapy [295]. While not performed on an MR-linear accelerator and without radiomics analysis, these preliminary data indicate that extension to MR radiomics analysis on the linear accelerator has potential. One of the limiting features of the MR-linear accelerator for radiomics analysis will be the decreased signal-to-noise ratio in image intensity due to the lower energy for image acquisition compared with diagnostic MRI. This should be the subject of future radiomics analyses.

7.3.3. STATISTICAL ANALYSIS IN RADIOMICS RESEARCH

Statistical analysis of radiomics features is complex, owing to the large number of features extracted and the high potential for multicollinearity across features [296]. For example, textural feature families of GLSZM features and GLDZM features are interconnected in their measure of shared grey level values and GLRLM features are also highly correlated with GLSZM and may not provide any additional information [182]. Similarly, the NGTDM feature family and the NGLDM feature family were introduced as alternatives to the GLCM feature family. Therefore, a feature reduction statistical technique is required to take account of this issue of multicollinearity. It also reduces overfitting of models and therefore increases their generalisability [297]. The feature reduction technique utilised in this thesis was principal components analysis (PCA).

Principal components analysis aims to simplify complexity in high-dimensional data while retaining trends and patterns. It reduces data by geometrically projecting data onto lower dimensions or scores called principal components (PCs), which describe the single value decomposition of the features [103]. When performing PCA it is important to remember that the aim is not to find clusters of similar data but to maximise variance, known as the maximum variance theory [298]. It can be argued that PCA may not give the best features for a radiomics analysis, but it does give a reduced subset of features using an unbiased selection towards the clinical features [299] and for this reason, it was utilised in Studies 3 and 4.

Other options in feature reduction are maximum relevance, minimum redundancy with Wilcoxon's rank-sum test as a feature selection criterion. This is particularly useful for machine learning algorithms [171] and has been reported for training quadratic discriminant analysis classifiers [31]. Maximum relevance ensures that the selected features have the highest correlated information with the endpoint [300] and minimum redundancy eliminates the redundant features [45]. It has been shown to have high prognostic performance and stability across head and neck and NSCLC [97]. Machine learning is the use of computational methods or models using data to improve performance or to make accurate predictions [301]. These algorithms 'learn' from the data and can automate the prediction process. In general in machine learning methods, a clustering unsupervised approach is used for associating radiomic features with histological subtypes. However, if labelled data are available, supervised approaches are preferable as they result in higher prediction accuracies [302].

7.3.4 THE FUTURE OF RADIOMICS RESEARCH - ARTIFICIAL INTELLIGENCE

While the radiomics feature extraction and statistical analysis in this thesis were conducted in the traditional manner, radiomics is now moving towards artificial intelligence (AI) frameworks, such as convolution neural networks and decision forests, with a blending of traditional radiomics approaches with these AI frameworks [303]. Renewed interest in AI, the branch of computer science that tries to mimic human intelligence in machines using algorithms without direct human input to specific tasks, occurred at the turn of the century, when the

financial constraints of computational power became less significant [304]. The nomenclature of AI can be challenging but in summary, convolutional neural networks are a subset of deep learning, which is turn is a subset of machine learning and machine learning comes under the umbrella term of AI [304].

In radiomics research, the ability of AI methods in automation of segmentation tasks [92] is a major step forward from observer-defined ROIs in traditional radiomics. It has been postulated that delineation, which is currently considered a critical skill of radiation oncologists and radiation therapists, will be replaced by artificial intelligence methods [303]. The machine learning methods that have been specifically used in auto segmentation are artificial neural networks (ANN), support vector machine (SVM) and deep learning [93]. Deep learning is essentially a variant of ANN that consists of a much higher number of hidden layers or multi-layered processes and uses many levels and iterations of abstraction to ascertain correlations from large amounts of data [305]. In radiomics analysis, an unsupervised deep neural network called 'Auto-Encoder' has recently been developed to generate texture descriptors [300]. Principal component analysis, as utilised in Studies 3 and 4 of this thesis, is considered a traditional unsupervised learning algorithm [306]. Other machine learning algorithms that have been reported in oncology include decision trees, where simple algorithms create mutually exclusive classes by answering questions in a previously defined order, *k*-nearest neighbours where a feature is classified based on its nearest neighbour in a dataset and support vector machine, where a model that is already trained can categorise new data [307]. Big datasets are required for supervised artificial intelligence-based feature extraction, which will require an agreed ontology for the naming of ROIs to reduce variability in format [308]. One such initiative has already been reported for prostate and head and neck cancers, where a tool-supported approach called 'Stature' successfully relabelled radiation therapy ROIs and dose volume histogram (DVH) metrics with agreed nomenclature by the AAPM task group 263 (TG-263), allowing for the automatic extraction of these data from DVHs [309]. Others have cited that because of the difficulties in labelling data, a drive towards unsupervised machine learning techniques is foreseen as vast amounts of unlabelled data already exist in medical imaging departments [310].

These techniques do not require the input of pre-processed ROIs for feature selection, as conducted in this thesis, but rather these algorithms learn these features from whole image volumes. However, their inability to explain in a comprehensible format to humans how they operate and come to a decision is problematic [311]. This 'black box' nature of AI models, as well as the lack of standardised longitudinal tumour control and toxicity profile data, means that adoption into routine clinical radiation oncology is some time away [312]. Accountability in radiation oncology is important and a prediction system that is not able to explain itself in a method understandable to humans has both ethical and legal consequences [313]. Coupled with this is the competing need for strict privacy protection of patient data [303], exacerbated for European radiation oncology departments by the recent introduction of the General Data Protection Regulation (GDPR).

7.4 LIMITATIONS AND STRENGTHS OF STUDIES 1-4- RADIOMICS QUALITY SCORE

A major limitation of both Studies 3 and 4 is the relatively small sample size for a radiomics analysis as well as the lack of a validation cohort. In radiomics analysis, the radiomics quality score (RQS) scores each component of the radiomics workflow, as well as considering cost-effectiveness, prospective versus retrospective study design and validation of data [314]. The highest weighted components are study design (prospective versus retrospective) and validation. The studies in this thesis were scored as given in Appendix 4 and achieved a result of 6 points out of a total of 38, an RQS score of 15.8%. This is a low score as no validation was performed, the data and radiomics package used were not open source and the study was retrospective in nature. To put this score in context, however, a recent systematic review on radiomics quality score reported that out of 41 included studies, 9 had an RQS of 0%, and, overall, the majority of studies were scored below 50% due to a lack of robust segmentation, external validation and cut-off analysis reporting, as well as a lack of cost-effectiveness monitoring and retrospective design [40]. More recently it has been found that the mean RQS score was 26.1% in 77 radiomics articles analysed [315]. The studies in this thesis scored well on criteria such as image protocol utilised, phantom study performed, statistical feature reduction performed and results were compared to the current 'gold standard' in the stratification of prostate cancer (PSA, Gleason

Score and D'amico risk stratification score - no 'gold standard' for PIMO score exists at present). Despite the retrospective nature of the study, the imaging protocol utilised was in line with ESUR recommendations [63]. The phantom study analysed the impact of different scan parameters on the reproducibility of features and principal component analysis was utilised as the feature reduction method. According to the RQS criteria, the ROI delineations in this study were not considered robust and were given a score of zero. However, the ROI delineations were specifically based on prostatic specimens as well as histopathologic reports, yielding as close to a 'ground truth' as is possible. It is postulated that the RQS score for this criterion is specifically for studies where the prostatic specimen are not available.

Studies 3 and 4 in this thesis were not validated using an external dataset and this is a recognised limitation of this work and of many radiomics studies. It has been reported in a review of radiomics studies that a validation dataset was used in only 8 out of 27 studies [316]. Acquiring a validation dataset that utilised prostatectomy specimens and whose patients had been administered PIMO was not feasible in this instance. Other considerations in the acquisition of validation datasets include issues between institutions and differences in data collection practices [317]. Previous radiomic studies have divided their own datasets into training and validation cohorts, but this was specifically not performed in this analysis in order to ensure a reasonable number of patients for each feature extracted, otherwise the models reported would have been invalid. Cross-validation, leave-one out and bootstrapping are methods to overcome these issues. In cross-validation, the data are randomly split into a fixed number (*k*) parts of equal size. All but one of these parts is used to train the data and one to validate it. This is repeated *k* times, each time changing which parts are used to train and validate until the model has been validated on every part once. Leave-one put is a specific variety of cross-validation, in which the number of parts is the same as the number of data points. During each iteration, the model is then trained on all but one data point and validated on the one left out. Boot-strapping artificially increases data by creating new data sets of equal size, randomly sampled from the original data [300].

A single axial 2D delineation was utilised as the ROI for all datasets in Studies 2-4 of this thesis, as do most clinical correlation studies. It cannot be assumed that

the heterogeneity of a tumour can be correctly characterised by a single 2D image; however partial volume effects in radiomics analysis are exacerbated if the ROI is three-dimensional [115].

The ADC datasets in used in Studies 2-4 in this thesis were generated from *b* values of 200-800 s/mm^2. The updated PI-RADS version 2, which was published after these images were acquired, recommends that one *b*-value, preferably at 50-100s/mm^2, one intermediate *b* value set at 800-1000s/mm^2, along with a mandatory high *b* value $\geq$ 1400s/mm^2 be acquired for ADC map calculation [64]. It is likely that the results reported in this thesis from the ADC datasets would differ if the ADC maps had been acquired at these values.

7.5 FUTURE DIRECTIONS

There are several studies that can subsequently be conducted using the datasets for Studies 2-4 in this thesis. The first is the validation of the current results. This is likely to be conducted using one of the cross-validation methods or boot-strapping methods previously described if a suitable dataset with PIMO score cannot be sourced. Second, Studies 2-4 can be repeated with 3D delineations of the ROIs, based on automatic or semi-automatic delineation methods. This will indicate whether feature extraction based on a single axial 2D slice is representative of that from a 3D volume and will indicate if, as the literature suggests, semi-automatic or automatic delineation methods result in radiomics features with higher prediction potential than manual delineation methods. Third, the phantom analysis in Study 1 can be repeated on a higher energy MR at 3 Tesla to ascertain if the higher signal-to-noise ratio is impacted to a lesser or greater degree than 1.5 Tesla when image parameters such as slice thickness, field of view and matrix are altered.

Outside of the current study, the potential for radiomics analysis in the identification of hypoxic regions of prostate tumours on the daily images from the MR linear accelerator warrants further investigation. The ability to adapt radiation therapy based on the biology of the tumour rather than on geometric features will assist in the true 'personalisation' of radiation therapy for prostate cancer patients. Exploiting AI for this analysis will expedite the process and enable the true biological adaption of prostate radiation therapy patients to become a clinical reality.

CHAPTER 8. REFERENCES

1. Grau, C., et al., *Radiotherapy equipment and departments in the European countries: Final results from the ESTRO-HERO survey.* Radiother Oncol, 2014. **112**(2): p. 155-64.

2. Bruinsma, S.M., et al., *Active surveillance for prostate cancer: a narrative review of clinical guidelines.* Nature Reviews Urology, 2016. **13**(1759-4820 (Electronic)): p. 151-67.

3. Neal, D.E., et al., *Ten-year Mortality, Disease Progression, and Treatment-related Side Effects in Men with Localised Prostate Cancer from the ProtecT Randomised Controlled Trial According to Treatment Received.* European Urology, 2020. **77**(3).

4. Hameed, M., et al., *The clinical utility of prostate cancer heterogeneity using texture analysis of multiparametric MRI.* International Urology and Nephrology, 2019. **51**(5): p. 817-824.

5. Steiger, P. and H.C. Thoeny, *Prostate MRI based on PI-RADS version 2: how we review and report.* Cancer Imaging, 2016. **16**(1): p. 9.

6. *Mottet N, Bellmunt J, Briers E, van den Bergh RCN, Bolla M, van Casteren NJ, et al. Guidelines on Prostate Cancer 2015 v2. 2015. http://uroweb.org/wp-content/uploads/EAU-Guidelines-Prostate-Cancer-2015-v2.pdf. Accessed 1 Dec 2015.*

7. Franiel, T., et al., *Areas suspicious for prostate cancer: MR-guided biopsy in patients with at least one transrectal US-guided biopsy with a negative finding--multiparametric MR imaging for detection and biopsy planning.* 2011. **259**(1): p. 162-72.

8. Schick, U., et al., *MRI-derived radiomics: methodology and clinical applications in the field of pelvic oncology.* 2019. **92**(1104).

9. Ree, A.H. and K.R. Redalen, *Personalized radiotherapy: concepts, biomarkers and trial design.* BJR, 2015(1051).

10. Zaorsky, N.G., et al., *Evolution of advanced technologies in prostate cancer radiotherapy.* Nature Reviews Urology, 2013. **10**(10): p. 565-579.

11. Miller, K.D., et al., *Cancer treatment and survivorship statistics, 2016.* CA Cancer J Clin, 2016. **66**(4): p. 271-89.

12. Mohler, J., et al., *NCCN clinical practice guidelines in oncology: prostate cancer.* J Natl Compr Canc Netw, 2010. **8**(2): p. 162-200.

13. Ferlay, J., et al., *Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012.* Int J Cancer, 2015. **136**(5): p. E359-86.

14. Ferlay, J., et al., *Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012.* Eur J Cancer, 2013. **49**(6): p. 1374-403.

15. Heidenreich, A., et al., *EAU guidelines on prostate cancer. part 1: screening, diagnosis, and local treatment with curative intent-update 2013.* Eur Urol, 2014. **65**(1): p. 124-37.

16. Cooperberg, M.R., J.M. Broering, and P.R. Carroll, *Time trends and local variation in primary treatment of localized prostate cancer.* J Clin Oncol, 2010. **28**(7): p. 1117-23.

17. Hou, Z., S. Li G Fau - Bai, and S. Bai, *High dose versus conventional dose in external beam radiotherapy of prostate cancer: a meta-analysis of long-term follow-up.* Journal of Cancer Research and Clinical Oncology, 2015. **141**(6): p. 1063-1071.

18. Hartford, A.C., et al., *American College of Radiology (ACR) and American Society for Radiation Oncology (ASTRO) Practice Guideline for Intensity-modulated Radiation Therapy (IMRT).* American Journal of Clinical Oncology, 2012. **35**(6): p. 612-617.

19. Hummel, S., et al., *Intensity-modulated radiotherapy for the treatment of prostate cancer: a systematic review and economic evaluation.* Health Technology Assessment, 2010. **14**(47).

20. Bolla, M., *Adjuvant hormonal treatment with radiotherapy for locally advanced prostate cancer.* European Urology, 1999. **35**(Suppl 1): p. 23-26.

21. Denham, J.W., et al., *Short-term androgen suppression and radiotherapy versus intermediate-term androgen suppression and radiotherapy, with or without zoledronic acid, in men with locally advanced prostate cancer (TROG 03.04 RADAR): 10-year results from a randomised, phase 3, factorial trial.* Lancet Oncology, 2019. **20**(2): p. 267-281.

22. Lee, W.R., et al., *Randomized Phase III Noninferiority Study Comparing Two Radiotherapy Fractionation Schedules in Patients With Low-Risk Prostate Cancer.* Journal of Clinical Oncology, 2016. **34**(20): p. 2325-32.

23. Catton, C.N., et al., *Randomized Trial of a Hypofractionated Radiation Regimen for the Treatment of Localized Prostate Cancer.* Journal of Clinical Oncology, 2017. **35**(17): p. 1884-1890.

24. Dearnaley, D., et al., *Conventional versus hypofractionated high-dose intensity-modulated radiotherapy for prostate cancer: 5-year outcomes of the randomised, non-inferiority, phase 3 CHHiP trial.* Lancet Oncology, 2016. **17**(8).

25. Incrocci, L., et al., *Hypofractionated versus conventionally fractionated radiotherapy for patients with localised prostate cancer (HYPRO): final efficacy results from a randomised, multicentre, open-label, phase 3 trial.* Lancet Oncology, 2016. **17**(8): p. 1061-1069.

26. Yin, Z., et al., *Moderate hypofractionated radiotherapy vs conventional fractionated radiotherapy in localized prostate cancer: a systemic review and meta-analysis from Phase III randomized trials.* OncoTargets and Therapy, 2019. **12**: p. 1259-68.

27. Vogelius, I.R. and S.M. Bentzen, *Diminishing returns from ultra-hypofractionated radiation therapy for prostate cancer. LID - S0360-3016(20)30069-9 [pii] LID - 10.1016/j.ijrobp.2020.01.010 [doi].* Int J Radiat Oncol Biol Phys, 2020(2): p. 299-304.

28. Sandler, H.M., et al., *A phase III protocol of androgen suppression (AS) and 3DCRT/IMRT versus AS and 3DCRT/IMRT followed by chemotherapy (CT) with docetaxel and prednisone for localized, high-risk prostate cancer (RTOG 0521).* Journal of Clinical Oncology, 2015. **33**(18).

29. Kishan, A.U., et al., *Clinical Outcomes for Patients with Gleason Score 9-10 Prostate Adenocarcinoma Treated With Radiotherapy or Radical Prostatectomy: A Multi-institutional Comparative Analysis.* Eur Urol, 2017. **71**(5): p. 766-773.

30. Monninkhof, E.M., et al., *Standard whole prostate gland radiotherapy with and without lesion boost in prostate cancer: Toxicity in the FLAME randomized controlled trial. LID - S0167-8140(17)32777-9 [pii] LID - 10.1016/j.radonc.2017.12.022 [doi].* Radiotherapy and Oncology, 2018. **127**(1): p. 74-80.

31. Shiradkar, R., et al., *Radiomics based targeted radiotherapy planning (Rad-TRaP): a computational framework for prostate cancer treatment planning with MRI.* Radiat Oncol, 2016. **11**(1): p. 148.

32. Jackson, W.C., et al., *Stereotactic Body Radiation Therapy for Localized Prostate Cancer: A Systematic Review and Meta-Analysis of Over 6,000 Patients Treated On Prospective Studies.* Int J Radiat Oncol Biol Phys, 2019. **104**(4): p. 778-789.

33. Corradini, S., et al., *MR-guidance in clinical reality: current treatment challenges and future perspectives.* Radiation Oncology, 2019. **14**(1): p. 92.

34. Bruynzeel, A.M.E., et al., *A Prospective Single-Arm Phase 2 Study of Stereotactic Magnetic Resonance Guided Adaptive Radiation Therapy for Prostate Cancer: Early Toxicity Results.* International Journal of Radiation Oncology*Biology*Physics, 2019. **105**(5): p. 1086-1094.

35. Yaromina, A., M. Krause, and M. Baumann, *Individualisation of cancer treatment from radiotherapy perspective.* Molecular Oncology, 2012. **6**: p. 211-221.

36. Aerts, H.J., et al., *Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach.* Nat Commun, 2014. **5**: p. 4006.

37. Yip, S.S. and H.J. Aerts, *Applications and limitations of radiomics.* Phys Med Biol, 2016. **61**(13): p. R150-66.

38. Kumar, V., et al., *Radiomics: the process and the challenges.* Magn Reson Imaging, 2012. **30**(9): p. 1234-48.

39. Lambin, P., et al., *Radiomics: extracting more information from medical images using advanced feature analysis.* Eur J Cancer, 2012. **48**(4): p. 441-6.

40. Sanduleanu, S., et al., *Tracking tumor biology with radiomics: A systematic review utilizing a radiomics quality score.* Radiotherapy and Oncology, 2018. **127**(3): p. 349-60.

41. Ginsburg, S.B., et al., *Radiomic features for prostate cancer detection on MRI differ between the transition and peripheral zones: Preliminary findings from a multi-institutional study.* J Magn Reson Imaging, 2016. **46**(1): p. 184-93.

42. Gandaglia, G., et al., *The Problem Is Not What to Do with Indolent and Harmless Prostate Cancer-The Problem Is How to Avoid Finding These Cancers.* European Urology, 2016(4): p. 547-8.

43. Turkbey, B., H. Mani, and S. V, *Multiparametric 3T prostate magnetic resonance imaging to detect cancer: histopathological correlation using prostatectomy specimens processed in customized magnetic resonance imaging based molds.* Journal of Urology, 2011. **186**: p. 1818-1824.

44. Le, J., N. Tan, and E. Shkolyar, *Multifocality and prostate cancer detection by multiparametric magnetic resonance imaging: correlation with whole-mount histopathology.* European Urology, 2015. **67**(3): p. 569-76.

45. Khalvati, F., A. Wong, and M.A. Haider, *Automated prostate cancer detection via comprehensive multi-parametric magnetic resonance imaging texture feature models.* BMC Med Imaging, 2015. **15**: p. 27.

46. Cameron, A., et al., *MAPS: A Quantitative Radiomics Approach for Prostate Cancer Detection.* IEEE Trans Biomed Eng, 2016. **63**(6): p. 1145-56.

47. Algohary, A.A.-O.h.o.o., et al., *Radiomic features on MRI enable risk categorization of prostate cancer patients on active surveillance: Preliminary findings. LID - 10.1002/jmri.25983 [doi].* Journal of Magnetic Resonance Imaging, 2018. **48**(3): p. 818-828.

48. Alyaev, Y.G., et al., *Prostate cancer: opportunities for magnetic resonance imaging.* Journal of Men's Health, 2008. **5**(1): p. 11-17.

49. Barucci, A., et al., *Prostate cancer radiomics using multi parametric MR imaging.*, in *IFAC, Technical, Scientific and Research Reports.* 2017. p. 1-51.

50. Park, B.K., et al., *Prospective evaluation of 3-T MRI performed before initial transrectal ultrasound-guided prostate biopsy in patients with high prostate-specific antigen and no previous biopsy.* American Journal of Roentgenology, 2011. **197**: p. W876-W881.

51. Ikonen, S., et al., *Optimal timing of post-biopsy MR imaging of the prostate.* Acta Radiologica, 2001. **42**(1): p. 70-73.

52. Hricak, H., et al., *Imaging prostate cancer: a multidisciplinary perspective.* Radiology, 2007. **243**(1): p. 28-53.

53. White, S., et al., *Prostate cancer: effect of postbiopsy hemorrhage on interpretation of MR images.* Radiology, 1995. **195**(2): p. 385-90.

54. Qayyum, A., et al., *Organ-confined prostate cancer: effect of prior transrectal biopsy on endorectal MRI and MR spectroscopic imaging.* American Journal of Roentgenology, 2004. **183**(4): p. 1079-83.

55. Ahmed, H.U., et al., *Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study.* Lancet, 2017. **389**(10071): p. 815-822.

56. Soodana-Prakash, N., et al., *Entering an era of radiogenomics in prostate cancer risk stratification.* Translational Andrology and Urology, 2018. **7**(Suppl 4): p. S443-S452.

57. Kwee, T.C., et al., *Whole-body diffusion-weighted magnetic resonance imaging.* European Journal of Radiology, 2009. **70**(3): p. 409-417.

58. Neill, J., *Diffusion Imaging Concepts for Clinicians.* Journal of Magnetic Resonance Imaging, 2008. **27**: p. 1-7.

59. Afaq, A., et al., *Clinical utility of diffusion-weighted magnetic resonance imaging in prostate cancer.* BJU International, 2011. **108**(11): p. 1716-1722.

60. Hosseinzadeh, K. and S. Schwarz, *Endorectal Diffusion-Weighted Imaging in Prostate Cancer to differentiate Malignant and Benign Peripheral Zone Tissue.* Journal of Magnetic Resonance Imaging, 2004. **20**: p. 654-661.

61. Malayeri, A., et al., *Principles and applications of diffusion-weighted imaging in cancer detection, staging and treatment follow-up.* Radiographics, 2011. **31**: p. 1773-1791.

62. Jie, C., T. Rongbo L Fau - Ping, and T. Ping, *The value of diffusion-weighted imaging in the detection of prostate cancer: a meta-analysis.* European Radiology, 2014. **24**(8): p. 1929-1941.

63. Barentsz, J.O., et al., *ESUR prostate MR guidelines 2012.* Eur Radiol, 2012. **22**(4): p. 746-57.

64. Turkbey, B., et al., *Prostate Imaging Reporting and Data System Version 2.1: 2019 Update of Prostate Imaging Reporting and Data System Version 2.* European Urology, 2019. **75**(3): p. 385-396.

65. Hamoen, E., et al., *Use of the Prostate Imaging, Reporting and Data System (PI-RADS) for prostate cancer detection with multi parametric magnetic resonance imaging: a diagnostic meta-analysis.* Eur Urol, 2015. **67**: p. 1112-1121.

66. Kasivisvanathan, V., et al., *MRI-Targeted or Standard Biopsy for Prostate-Cancer Diagnosis.* New England Journal of Medicine, 2018. **378**(19): p. 1767-1777.

67. Milosevic, M., et al., *Tumor hypoxia predicts biochemical failure following radiotherapy for clinically localized prostate cancer.* Clinical Cancer Research, 2012. **18**(7): p. 2108-2114.

68. Grimes, D.R., D.R. Warren, and S. Warren, *Hypoxia imaging and radiotherapy: bridging the resolution gap.* British Journal of Radiology, 2017. **90**(1076).

69. Ahmed, H.U., et al., *Do low-grade and low-volume prostate cancers bear the hallmarks of malignancy?* Lancet Oncology, 2012. **13**(11): p. e509-17.

70. Fraga, A., et al., *Hypoxia and Prostate Cancer Aggressiveness: A Tale With Many Endings.* Clinical Genitourinary Cancer, 2015. **13**(4): p. 295-301.

71. Walsh, J.C., et al., *The clinical importance of assessing tumor hypoxia: relationship of tumor hypoxia to prognosis and therapeutic opportunities.* Antioxidants and Redox Signaling, 2014. **21**(10): p. 1516-1554.

72. Hoskin, P.J., *Hypoxia dose painting in prostate and cervix cancer.* Acta Oncologica, 2015. **54**(9): p. 1259-62.

73. Walsh, S., et al., *Decision Support Systems in Oncology.* JCO clinical cancer informatics, 2019. **3**.

74. Cao, X., et al., *Tissue pO2 distributions in xenograft tumors dynamically imaged by Cherenkov-excited phosphorescence during fractionated radiation therapy.* Nature Communications, 2020. **11**(1): p. 573.

75. Ragnum, H.B., et al., *The tumour hypoxia marker pimonidazole reflects a transcriptional programme associated with aggressive prostate cancer.* British Journal of Cancer, 2015. **112**(2): p. 382-390.

76. Hoskin, P.J., et al., *Hypoxia in prostate cancer: correlation of BOLD-MRI with pimonidazole immunohistochemistry-initial observations.* Int J Radiat Oncol Biol Phys, 2007. **68**(4): p. 1065-71.

77. Kaanders, J.H., et al., *Pimonidazole binding and tumor vascularity predict for treatment outcome in head and neck cancer.* Cancer Research, 2002. **62**(23): p. 7066-74.

78. Zhao, D., et al., *Tumor oxygen dynamics: correlation of in vivo MRI with histological findings.* Neoplasia, 2003. **5**(4): p. 308-18.

79. Knight, S.P., et al., *A novel anthropomorphic flow phantom for the quantitative evaluation of prostate DCE-MRI acquisition techniques.* Physics in Medicine and Biology, 2016. **61**(20): p. 7466-7483.

80. Marcu, L.G., J.C. Forster, and E. Bezak, *The Potential Role of Radiomics and Radiogenomics in Patient Stratification by Tumor Hypoxia Status.* Journal of the American College of Radiology, 2019. **16**(9, Part B): p. 1329-1337.

81. Overgaard, J., et al., *Five compared with six fractions per week of conventional radiotherapy of squamous-cell carcinoma of head and neck: DAHANCA 6 and 7 randomised controlled trial.* The Lancet, 2003. **362**(9388): p. 933-40.

82. Stoyanova, R., et al., *Prostate cancer radionics and the promise of radiogenomics.* Translational Cancer Research, 2016. **5**(4): p. 432-447.

83. Gillies, R.J., P.E. Kinahan, and H. Hricak, *Radiomics: Images Are More than Pictures, They Are Data.* Radiology, 2016. **278**(2): p. 563-77.

84. Fave, X., et al., *Delta-radiomics features for the prediction of patient outcomes in non-small cell lung cancer.* Sci Rep, 2017. **7**(1): p. 588.

85. O'Connor, J.P., et al., *Imaging biomarker roadmap for cancer studies.* Nat Rev Clin Oncol, 2017. **14**(3): p. 169-186.

86. Zwanenburg, A., et al., *The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping.* Radiology, 2020. **295**(2): p. 328-338.

87. You, X., et al., *Denoising of MR images with Rician noise using a wider neural network and noise range division.* Magnetic Resonance Imaging, 2019. **64**: p. 154-59.

88. Yang, F., et al., *Evaluation of radiomic texture feature error due to MRI acquisition and reconstruction: A simulation study utilizing ground truth.* Physics Medica, 2018. **50**: p. 26-36.

89. Scalco, E., et al., *T2w-MRI signal normalization affects radiomics features reproducibility. LID - 10.1002/mp.14038 [doi].* Medical Physics, 2020. **47**(4): p. 1680-91.

90. Vinod, S.K., et al., *Uncertainties in volume delineation in radiation oncology: A systematic review and recommendations for future studies.* Radiotherapy and Oncology, 2016. **121**(2): p. 169-179.

91. Cardenas Ce Fau - Yang, J., et al., *Advances in Auto-Segmentation.* Seminars in Radiation Oncology, 2019. **29**(3): p. 185-197.

92. Qin, W., et al., *Superpixel-based and boundary-sensitive convolutional neural network for automated liver segmentation.* Physics in Medicine and Biology, 2018. **63**(9).

93. Boon, I.S., T.P.T. Au Yong, and C.S. Boon, *Assessing the Role of Artificial Intelligence (AI) in Clinical Oncology: Utility of Machine Learning in Radiotherapy Target Volume Delineation. LID - E131 [pii] LID - 10.3390/medicines5040131 [doi].* Medicines (Basel), 2018. **5**(4).

94. Aerts, H.J., et al., *Defining a Radiomic Response Phenotype: A Pilot Study using targeted therapy in NSCLC.* Sci Rep, 2016. **6**: p. 33860.

95. Kumar, S., et al., *Cardiac tamponade and suprerior vena cava thrombosis: A dilemma.* Journal of the American College of Cardiology, 2018. **71**(11).

96. Gnep, K., et al., *Haralick textural features on T2 -weighted MRI are associated with biochemical recurrence following radiotherapy for peripheral zone prostate cancer.* J Magn Reson Imaging, 2017. **45**(1): p. 103-117.

97. Parmar, C., et al., *Radiomic Machine-Learning Classifiers for Prognostic Biomarkers of Head and Neck Cancer.* Front Oncol, 2015. **5**: p. 272.

98. Balagurunathan, Y., et al., *Test-retest reproducibility analysis of lung CT image features.* J Digit Imaging, 2014. **27**(6): p. 805-23.

99. Parmar, C., et al., *Radiomic feature clusters and prognostic signatures specific for Lung and Head & Neck cancer.* Sci Rep, 2015. **5**: p. 11044.

100. Aerts, H., et al., *Radiomics: extracting advanced features from medical imaging.* Radiotherapy and Oncology, 2012. **103**(S1): p. S70-S71.

101. van der Schaaf, A., et al., *Multivariate modelling of complications with data driven variable selection: guarding against overfitting and effects of data set size.* Radiotherapy and Oncology, 2012. **105**(1): p. 115-121.

102. Wibmer, A., et al., *Haralick texture analysis of prostate MRI: utility for differentiating non-cancerous prostate from prostate cancer and differentiating prostate cancers with different Gleason scores.* Eur Radiol, 2015. **25**: p. 2840-2850.

103. Huynh, E., et al., *CT-based radiomic analysis of stereotactic body radiation therapy patients with lung cancer.* Radiother Oncol, 2016. **120**(2): p. 258-66.

104. Mattonen, S.A., et al., *Detection of Local Cancer Recurrence After Stereotactic Ablative Radiation Therapy for Lung Cancer: Physician Performance Versus Radiomic Assessment.* Int J Radiat Oncol Biol Phys, 2016. **94**(5): p. 1121-8.

105. Coroller, T.P., et al., *CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma.* Radiother Oncol, 2015. **114**(3): p. 345-50.

106. Hesketh, R.L., A.X. Zhu, and R. Oklu, *Radiomics and circulating tumor cells: personalized care in hepatocellular carcinoma?* Diagn Interv Radiol, 2015. **21**(1): p. 78-84.

107. Hambrock, T., et al., *Relationship between apparent diffusion coefficients at 3.0-T MR imaging and Gleason grade in peripheral zone prostate cancer.* Radiology, 2011. **259**(2): p. 453-61.

108. Freedland, S.J., et al., *Prognostic utility of cell cycle progression score in men with prostate cancer after primary external beam radiation therapy.* Int J Radiat Oncol Biol Phys, 2013. **86**(5): p. 848-53.

109. Bristow, R.G., A. Berlin, and A. Dal Pra, *An arranged marriage for precision medicine: hypoxia and genomic assays in localized prostate cancer radiotherapy.* Br J Radiol, 2014. **87**(1035): p. 20130753.

110. Lalonde, E., et al., *Tumour genomic and microenvironmental heterogeneity for integrated prediction of 5-year biochemical recurrence of prostate cancer: a retrospective cohort study.* Lancet Oncol, 2014. **15**(13): p. 1521-32.

111. Cojoc, M., et al., *Aldehyde Dehydrogenase Is Regulated by beta-Catenin/TCF and Promotes Radioresistance in Prostate Cancer Progenitor Cells.* Cancer Res, 2015. **75**(7): p. 1482-94.

112. Abdel-Wahab, M., et al., *Influence of number of CAG repeats on local control in the RTOG 86-10 protocol.* Am J Clin Oncol, 2006. **29**(1): p. 14-20.

113. Panth, K.M., et al., *Is there a causal relationship between genetic changes and radiomics-based image features? An in vivo preclinical experiment with doxycycline inducible GADD34 tumor cells.* Radiother Oncol, 2015. **116**(3): p. 462-6.

114. Stoyanova, R., et al., *Association of multiparametric MRI quantitative imaging features with prostate cancer gene expression in MRI-targeted prostate biopsies.* Oncotarget, 2016. **7**(33): p. 53362-53376.

115. Zhao, B., et al., *Reproducibility of radiomics for deciphering tumor phenotype with imaging.* Sci Rep, 2016. **6**: p. 23428.

116. Fiset, S., et al., *Repeatability and reproducibility of MRI-based radiomic features in cervical cancer.* Radiotherapy and Oncology, 2019. **135**: p. 107-114.

117. Leijenaar, R.T., et al., *The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis.* Sci Rep, 2015. **5**: p. 11075.

118. van Velden, F.H., et al., *Repeatability of Radiomic Features in Non-Small-Cell Lung Cancer [(18)F]FDG-PET/CT Studies: Impact of Reconstruction and Delineation.* Mol Imaging Biol, 2016. **18**(5): p. 788-95.

119. Traverso, A., et al., *Sensitivity of radiomic features to inter-observer variability and image pre-processing in Apparent Diffusion Coefficient (ADC) maps of cervix cancer patients. LID - S0167-8140(19)33047-6 [pii] LID - 10.1016/j.radonc.2019.08.008 [doi].* Radiotherapy and Oncology, 2020. **143**: p. 88-94.

120. Hompland, T., et al., *Combined MR Imaging of Oxygen Consumption and Supply Reveals Tumor Hypoxia and Aggressiveness in Prostate Cancer Patients.* Cancer Research, 2018. **78**(16): p. 4774-4785.

121. van der Kwast, T.H., et al., *International Society of Urological Pathology (ISUP) Consensus Conference on Handling and Staging of Radical Prostatectomy Specimens. Working group 2: T2 substaging and prostate cancer volume.* Modern Pathology, 2011. **24**: p. 16-25.

122. Schwier, M., et al., *Repeatability of Multiparametric Prostate MRI Radiomics Features.* Scientific Reports, 2019. **9**.

123. McRobbie, D., et al., *MRI from picture to proton*. Third ed. 2017, Cambridge: Cambridge Medicine.

124. Waugh, S.A., et al., *The influence of field strength and different clinical breast MRI protocols on the outcome of texture analysis using foam phantoms.* Medical Physics, 2011. **38**(9): p. 5058-5066.

125. Cusumano, D., et al., *Fractal-based radiomic approach to predict complete pathological response after chemo-radiotherapy in rectal cancer.* La Radiologia Medica, 2018. **123**: p. 286-295.

126. Thwaites, D., et al., *An Australian mining boom: Developing an Australian network for datamining routine radiotherapy clinical and radiomics data for clinical decision support.* Australasian Physical and Engineering Sciences in Medicine, 2016. **39**(1): p. 322-323.

127. Tino, R.A.-O., et al., *A Systematic Review on 3D-Printed Imaging and Dosimetry Phantoms in Radiation Therapy.* Technology in Cancer Research and Treatment, 2019. **18**: p. 1-14.

128. Oikonomou, A.A.-O.X., et al., *Radiomics analysis at PET/CT contributes to prognosis of recurrence and survival in lung cancer treated with stereotactic body radiotherapy.* Scientific Reports, 2018. **8**(1).

129. Kaiser, F., *The application of electronic computers to factor analysis.* Educational and Psychology Measurement, 1960. **20**: p. 141-151.

130. Cattell, R., *The scree test for the number of factors.* Multivariate Behavioural Research, 1966. **1**: p. 245-276.

131. Hanley Ja Fau - McNeil, B.J. and B.J. McNeil, *The meaning and use of the area under a receiver operating characteristic (ROC) curve.* Radiology, 1982. **143**(1).

132. Bologna, M., V. Corino, and L. Mainardi, *Technical Note: Virtual phantom analyses for preprocessing evaluation and detection of a robust feature set for MRI-radiomics of the brain.* Medical Physics, 2019. **46**(11): p. 5116-5123.

133. Shiri, I., et al., *The impact of image reconstruction settings on 18F-FDG PET radiomic features: multi-scanner phantom and patient studies.* European Radiology, 2017. **27**(11): p. 4498-4509.

134. Wang, D., et al., *Geometric distortion in clinical MRI systems Part I: evaluation using a 3D phantom.* Magnetic Resonance Imaging, 2004. **22**(9): p. 1211-1221.

135. Mallozzi, R., *Geometric Distortion in MRI*. 2015, The Phantom Laboratory, Incorporated.

136. Torfeh, T., et al., *Characterization of 3D geometric distortion of magnetic resonance imaging scanners commissioned for radiation therapy planning.* Magnetic Resonance Imaging, 2016. **34**(5): p. 645-653.

137. Bieniosek, M.F., B.J. Lee, and C.S. Levin, *Technical Note: Characterization of custom 3D printed multimodality imaging phantoms.* Medical Physics, 2015. **42**(10): p. 5913-8.

138. Surry, K., et al., *Poly (vinyl alcohol) cryogel phantoms for use in ultrasound and MR imaging.* Phys. Med. Biol., 2004. **49**: p. 5529-5546.

139. Niebuhr, N.I., et al., *The ADAM-pelvis phantom-an anthropomorphic, deformable and multimodal phantom for MRgRT.* Physics in Medicine and Biology, 2019. **64**(4).

140. Sun, J., et al., *MRI simulation:end-to-end testing for prostate radiation therapy using geometric pelvic MRI phantoms.* Phys. Med. Biol., 2015. **60**: p. 3097-3109.

141. Chu, K. and B. Rutt, *Polyvinyl alcohol cryogel: An ideal phantom material for MR studies of arterial flow and elasticity.* Magnetic Resonance in Medicine, 1997. **37**: p. 314-319.

142. Cournane, S., et al., *Assessment of the accuracy of an ultrasound elastography liver scanning system using a PVA-cryogel phantom with optimal acoustic and mechanical properties.* Phys. Med. Biol., 2010. **55**: p. 5965-5983.

143. Lindner, U., et al., *Construction and evaluation of an anatomically correct multi-image modality compatible phantom for prostate cancer focal ablation.* The Journal of Urology, 2010. **184**: p. 352-357.

144. Hattori, K., et al., *Development of MRI phantom equivalent to human tissues for 3.0-T MRI.* Medical Physics, 2013. **40**(3).

145. D'Souza, W., et al., *Tissue mimicking materials for a multi-modality prostate phantom.* Medical Physics, 2001. **28**(4): p. 688-700.

146. Jerome, N.P., et al., *Development of a temperature-controlled phantom for magnetic resonance quality assurance of diffusion, dynamic, and relaxometry measurements.* Medical Physics, 2016. **43**(6): p. 2998-3007.

147. George, E., et al., *Measuring and Establishing the Accuracy and Reproducibility of 3D Printed Medical Models.* Radiographics, 2017. **37**(5): p. 1424-1450.

148. Madamesila, J., et al., *Characterizing 3D printing in the fabrication of variable density phantoms for quality assurance of radiotherapy.* Physics Medica, 2016. **32**(1): p. 242-247.

149. Menikou, G., et al., *MRI compatible head phantom for ultrasound surgery.* 57, 2015: p. 144-52.

150. Hazelaar, C., et al., *Using 3D printing techniques to create an anthropomorphic thorax phantom for medical imaging purposes.* Medical Physics, 2018. **45**(1): p. 92-100.

151. Mitsouras, D., et al., *Three-dimensional printing of MRI-visible phantoms and MR image-guided therapy simulation.* Magnetic Resonance in Medicine, 2017. **77**(2): p. 613-622.

152. Jia, G., et al., *Assessing prostate volume by magnetic resonance imaging: a comparison of different measurement approaches for organ volume analysis.* Investigative Radiology, 2005. **40**(4): p. 243-248.

153. Hu, P., et al., *Reproducibility with repeat CT in radiomics study for rectal cancer.* Oncotarget, 2016. **7**(44): p. 71440-71446.

154. Fütterer, J.J., S.W.T.P.J. Barentsz Jo Fau - Heijmink, and S.W. Heijmink, *Value of 3-T magnetic resonance imaging in local staging of prostate cancer.* Topics in Magnetic Resonance Imaging, 2008. **19**(6).

155. Han, S., F.F. Yin, and J. Cai, *Evaluation of dosimetric uncertainty caused by MR geometric distortion in MRI-based liver SBRT treatment planning.* Journal of Applied Clinical Medical Physics, 2019. **20**(2): p. 43-50.

156. Baldwin, L.N., B.G. Wachowicz K Fau - Fallone, and B.G. Fallone, *A two-step scheme for distortion rectification of magnetic resonance images.* Medical Physics, 2009. **36**(9): p. 3917-26.

157. Jovicich, J., et al., *Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data.* Neuroimage, 2006. **30**(2): p. 436-43.

158. Pappas, E.P., et al., *MRI-Related Geometric Distortions in Stereotactic Radiotherapy Treatment Planning: Evaluation and Dosimetric Impact.* Technology in Cancer Research and Treatment, 2017. **16**(6): p. 1120-29.

159. Agarwal, S., et al., *An analysis of immediate postmastectomy breast reconstruction frequency using the surveillance, epidemiology, and end results database.* Breast Journal, 2011. **17**(4): p. 352-358.

160. Walker, A., et al., *MRI geometric distortion: Impact on tangential whole-breast IMRT.* Journal of Applied Clinical Medical Physics, 2016. **17**(5): p. 7-19.

161. Cloonan, A.J., et al., *3D-Printed Tissue-Mimicking Phantoms for Medical Imaging and Computational Validation Applications.* 3D Printing and Additive Manufacturing, 2014. **1**(1): p. 14-23.

162. Yan, J., et al., *Impact of Image Reconstruction Settings on Texture Features in 18F-FDG PET.* The Journal of Nuclear Medicine, 2015. **56**(11): p. 1667-1673.

163. Shafiq-Ul-Hassan, M., et al., *Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels.* Medical Physics, 2017. **44**(3): p. 1050-1062.

164. Mayerhoefer, M.E., et al., *Effects of MRI acquisition parameter variations and protocol heterogeneity on the results of texture analysis and pattern discrimination: an application-oriented study.* Medical Physics, 2009. **36**(4): p. 1236-43.

165. Berenguer, R., et al., *Radiomics of CT Features May Be Nonreproducible and Redundant: Influence of CT Acquisition Parameters.* Radiology, 2018. **288**(2): p. 407-415.

166. Lam, T.B.L., et al., *EAU-EANM-ESTRO-ESUR-SIOG Prostate Cancer Guideline Panel Consensus Statements for Deferred Treatment with Curative Intent for Localised Prostate Cancer from an International Collaborative Study (DETECTIVE Study).* European Urology, 2019. **76**(6): p. 790-813.

167. Leijenaar, R.T., et al., *Stability of FDG-PET Radiomics features: an integrated analysis of test-retest and inter-observer variability.* Acta Oncol, 2013. **52**(7): p. 1391-7.

168. Larue, R., et al., *Influence of gray level discretization on radiomic feature stability for different CT scanners, tube currents and slice thicknesses: a comprehensive phantom study.* Acta Oncologica, 2017. **56**(11): p. 1544-1553.

169. Lu, L., et al., *Assessing Agreement between Radiomic Features Computed for Multiple CT Imaging Settings.* PLoS One, 2016. **11**(12): p. e0166550.

170. van Timmeren, J.E., et al., *Test-Retest Data for Radiomics Feature Stability Analysis: Generalizable or Study-Specific?* Tomography, 2016. **2**(4): p. 361-365.

171. Merisaari, H.A.-O.h.o.o., et al., *Repeatability of radiomics and machine learning for DWI: Short-term repeatability study of 112 patients with prostate cancer. LID - 10.1002/mrm.28058 [doi].* Magnetic Resonance in Medicine, 2020. **83**(6): p. 2293-2309.

172. Traverso, A., et al., *Stability of radiomic features of apparent diffusion coefficient (ADC) maps for locally advanced rectal cancer in response to image pre-processing.* Physics Medica, 2019. **61**: p. 44-51.

173. Peerlings, J., et al., *Stability of radiomics features in apparent diffusion coefficient maps from a multi-centre test-retest trial.* Scientific Reports, 2019. **9**(1).

174. Whybra, P.A.-O.h.o.o., et al., *Assessing radiomic feature robustness to interpolation in (18)F-FDG PET imaging.* Scientific Reports, 2019. **9**(1).

175. Oliver, J.A., et al., *Sensitivity of Image Features to Noise in Conventional and Respiratory-Gated PET/CT Images of Lung Cancer: Uncorrelated Noise Effects.* Technol Cancer Res Treat, 2017. **16**(5): p. 595-608.

176. Meier-Schroers, M., et al., *Revised PROPELLER for T2-weighted imaging of the prostate at 3 Tesla: impact on lesion detection and PI-RADS classification.* European Radiology, 2018. **28**(1): p. 24-30.

177. Malyarenko, D.I., et al., *Correction of Gradient Nonlinearity Bias in Quantitative Diffusion Parameters of Renal Tissue with Intra Voxel Incoherent Motion.* Tomography, 2015. **1**(2): p. 145-151.

178. Koo, T.K. and M.Y. Li, *A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research.* Journal of Chiropractic Medicine, 2016. **15**: p. 155-163.

179. Shrout, P.E. and J.L. Fleiss, *Intraclass correlations: uses in assessing rater reliability.* Psychological Bulletin, 1979. **86**(2): p. 420-428.

180. Demetrashvili, N. and E.R. Van den Heuvel, *Confidence intervals for intraclass correlation coefficients in a nonlinear dose-response meta-analysis.* Biometrics, 2015. **71**(2): p. 548-55.

181. Chen, L., et al., *Diffusion-weighted imaging of rectal cancer on repeatability and cancer characterization: an effect of b-value distribution study.* Cancer Imaging, 2018. **18**(1).

182. Tixier, F., et al., *Reproducibility of tumor uptake heterogeneity characterization through textural feature analysis in 18F-FDG PET.* Journal of Nuclear Medicine, 2012. **53**(5): p. 693-700.

183. Midya, A.A.-O.h.o.o., et al., *Influence of CT acquisition and reconstruction parameters on radiomic feature reproducibility.* Journal of Medical Imaging, 2018. **5**(1).

184. Le Jd Fau - Huang, J., L.S. Huang J Fau - Marks, and L.S. Marks, *Targeted prostate biopsy: value of multiparametric magnetic resonance imaging in detection of localized cancer.* Asian Journal of Andrology, 2014. **16**(4): p. 522-29.

185. Bottomley Pa Fau - Foster, T.H., et al., *A review of normal tissue hydrogen NMR relaxation times and relaxation mechanisms from 1-100 MHz: dependence on tissue type, NMR frequency, temperature, species, excision, and age.* Medical Physics, 1984. **11**(4): p. 425-48.

186. Langer, D., et al., *Prostate cancer detection with multi-parametric MRI: Logistic regression analysis of quantitative T2, diffusion-weighted imaging and dynamic contrast-enhanced MRI.* Journal of Magnetic Resonance Imaging, 2009. **30**: p. 327-334.

187. Puech, P., et al., *Detecting prostate cancer with MRI - why and how.* Diagnostic and Interventional Imaging, 2012. **93**(4): p. 268-78.

188. Gibbs, P., et al., *Comparison of quantitative T2 mapping and diffusion-weighted imaging in the normal and pathologic prostate.* Magnetic Resonance in Medicine, 2001. **46**(6): p. 1054-58.

189. Barrett, T., B. Turkbey, and P.L. Choyke, *PI-RADS version 2: what you need to know.* Clin Radiol, 2015. **70**(11): p. 1165-76.

190. Rodrigues, G., et al., *Consensus statement on palliative lung radiotherapy: third international consensus workshop on palliative radiotherapy and symptom control.* Clinical Lung Cancer, 2012. **13**(1): p. 1-5.

191. Turkbey, B. and P.L. Choyke, *PIRADS 2.0: what is new?* Diagnostic and Interventional Radiology, 2015. **21**: p. 382-384.

192. Heijmink, S.W., et al., *A comparison of the diagnostic performance of systematic versus ultrasound-guided biopsies of prostate cancer.* European Radiology, 2006. **16**(4): p. 927-38.

193. Cohen, J., *Statistical power analysis for the behavioural sciences* 2nd ed. 1988, Hillsdale:NJ: Lawrence Erlbaum Associates.

194. Marcu, L.G., L. Moghaddasi, and E. Bezak, *Imaging of Tumor Characteristics and Molecular Pathways With PET: Developments Over the Last Decade Toward Personalized Cancer Therapy.* Int J Radiat Oncol Biol Phys, 2018. **102**(4): p. 1165-1182.

195. Fjeldbo, C.S., et al., *Integrative Analysis of DCE-MRI and Gene Expression Profiles in Construction of a Gene Classifier for Assessment of Hypoxia-Related Risk of Chemoradiotherapy Failure in Cervical Cancer.* Clinical Cancer Research, 2016. **22**(16): p. 4067-76.

196. Beig, N., et al., *Radiogenomic analysis of hypoxia pathway is predictive of overall survival in Glioblastoma.* Scientific Reports, 2018. **8**(1).

197. Ganeshan, B., et al., *Non-small cell lung cancer: histopathologic correlates for texture parameters at CT.* Radiology, 2013. **260**(1): p. 326-36.

198. Sun, Y., et al., *Association analysis between quantitative MRI features and hypoxia-related genetic profiles in prostate cancer: a pilot study.* British Journal of Radiology, 2019. **92**(1104).

199. Yu, T.T., et al., *Pretreatment Prediction of Adaptive Radiation Therapy Eligibility Using MRI-Based Radiomics for Advanced Nasopharyngeal Carcinoma Patients.* Frontiers in Oncology, 2019. **16**(9).

200. Fave, X., et al., *Can radiomics features be reproducibly measured from CBCT images for patients with non-small cell lung cancer?* Med Phys, 2015. **42**(12): p. 6784-97.

201. Shi, L., et al., *Cone-beam computed tomography-based delta-radiomics for early response assessment in radiotherapy for locally advanced lung cancer. LID - 10.1088/1361-6560/ab3247 [doi].* Physics in Medicine and Biology, 2020. **65**(1).

202. Sorensen, A.A.-O.h.o.o., et al., *Textural features of hypoxia PET predict survival in head and neck cancer during chemoradiotherapy. LID - 10.1007/s00259-019-04609-9 [doi].* European Journal of Nuclear Medicine and Molecular Imaging, 2020. **47**: p. 1056-1064.

203. Fehr, D., et al., *Automatic classification of prostate cancer Gleason scores from multiparametric magnetic resonance images.* Proceedings of the National Academy of Sciences of the United States of America, 2015. **112**(46): p. E6265-E6273.

204. Vignati, A., et al., *Texture features on T2-weighted magnetic resonance imaging: new potential biomarkers for prostate cancer aggressiveness.* Physics in Medicine and Biology, 2015. **60**(7): p. 2685-2701.

205. Abdollahi, H.A.-O.h.o.o., et al., *Machine learning-based radiomic models to predict intensity-modulated radiation therapy response, Gleason score and stage in prostate cancer.* La radiologia medica, 2019. **124**: p. 555-567.

206. Chaddad, A., et al., *Predicting Gleason Score of Prostate Cancer Patients Using Radiomic Analysis.* Frontiers in Oncology, 2018. **8**(630).

207. Chen, T., et al., *Prostate Cancer Differentiation and Aggressiveness: Assessment With a Radiomic-Based Model vs. PI-RADS v2.* Journal of Magnetic Resonance Imaging, 2019. **49**(3): p. 875-884.

208. Stoyanova, R., et al., *Prostate cancer radiomics and the promise of radiogenomics.* Transl Cancer Res, 2016. **5**(4): p. 432-447.

209. Ma, S., et al., *Preoperative Prediction of Extracapsular Extension: Radiomics Signature Based on Magnetic Resonance Imaging to Stage Prostate Cancer. LID - 10.1007/s11307-019-01405-7 [doi].* Molecular Imaging and Biology, 2020. **22**: p. 711-721.

210. Ma, X., et al., *MRI-based radiomics of rectal cancer: preoperative assessment of the pathological features.* BMC Medical Imaging, 2019. **19**.

211. Sun, Y., et al., *Radiomic features of pretreatment MRI could identify T stage in patients with rectal cancer: Preliminary findings. LID - 10.1002/jmri.25969 [doi].* Journal of Magnetic Resonance Imaging, 2018. **48**(3): p. 615-21.

212. Leger, S., et al., *A comparative study of machine learning methods for time-to-event survival data for radiomics risk modelling.* Sci Rep, 2017. **7**(1): p. 13206.

213. Rastinehad, A.R., et al., *D'Amico risk stratification correlates with degree of suspicion of prostate cancer on multiparametric magnetic resonance imaging.* Journal of Urology, 2011. **185**(3): p. 815-820.

214. Algarra, R., et al., *Optimizing D'Amico risk groups in radical prostatectomy through the addition of magnetic resonance imaging data.* Acta Urologia Esapnolas, 2014. **38**(9): p. 594-599.

215. D'Amico, A.V., et al., *Biochemical outcome after radical prostatectomy or external beam radiation therapy for patients with clinically localized prostate carcinoma in the prostate specific antigen era.* JAMA, 1998. **280**(11): p. 969-974.

216. Langer, D.L., et al., *Intermixed normal tissue within prostate cancer: effect on MR imaging measurements of apparent diffusion coefficient and T2--sparse versus dense cancers.* Radiology, 2008. **249**(3): p. 900-908.

217. Keetch, D.W. and W.J. Catalona, *Prostatic transition zone biopsies in men with previous negative biopsies and persistently elevated serum prostate specific antigen values.* Journal of Urology, 1995. **154**(5): p. 1795-97.

218. Catalona, W.J., et al., *Measurement of prostate-specific antigen in serum as a screening test for prostate cancer.* New England Journal of Medicine, 1991(17): p. 1156-61.

219. Catalona Wj Fau - Smith, D.S. and D.S. Smith, *Comparison of different serum prostate specific antigen measures for early prostate cancer detection.* Cancer, 1994. **74**(5): p. 1516-1518.

220. Schröder, F.H. and M.J. Roobol, *ERSPC and PLCO Prostate Cancer Screening Studies: What Are the Differences?* European Urology, 2010. **58**(1): p. 46-52.

221. Panebianco, V., et al., *Multiparametric magnetic resonance imaging vs. standard care in men being evaluated for prostate cancer: a randomized study.* Urologic Oncology, 2015. **33**(1): p. e1-e17.e7.

222. Haider, M., Y. Amoozadeh, and K. Jhaveri, *DW-MRI for disease characterisation in the pelvis.*, in *Diffusion-weighted MR imaging*, D. Koh and H. then, Editors. 2010.

223. Yoshimitsu, K., et al., *Usefulness of apparent diffusion coefficient map in diagnosing prostate carcinoma: correlation with stepwise histopathology.* Journal of Magnetic Resonance Imaging, 2008. **27**(1): p. 132-9.

224. AbdelMaboud, N.M., H.H. Elsaid, and E.A. Aboubeih, *The role of diffusion – Weighted MRI in evaluation of prostate cancer.* The Egyptian Journal of Radiology and Nuclear Medicine, 2014. **45**(1): p. 231-236.

225. Chan, I., et al., *Detection of prostate cancer by integration of line-scan diffusion, T2-mapping and T2-weighted magnetic resonance imaging; a multichannel statistical classifier.* Med Phys, 2003. **30**(9): p. 2390-8.

226. Morgan, V.A., et al., *Evaluation of the potential of diffusion-weighted imaging in prostate cancer detection.* Acta Radiologica, 2007. **48**(6): p. 695-703.

227. Carlsson, J., et al., *Differences in microRNA expression during tumor development in the transition and peripheral zones of the prostate.* BMC Cancer, 2013. **29**(13).

228. McNeal, J.E., et al., *Zonal distribution of prostatic adenocarcinoma. Correlation with histologic pattern and direction of spread.* American Journal of Surgical Pathology, 1988. **12**(12): p. 897-906.

229. Sato, C., et al., *Differentiation of noncancerous tissue and cancer lesions by apparent diffusion coefficient values in transition and peripheral zones of the prostate.* Journal of Magnetic Resonance Imaging, 2005. **21**(3): p. 258-62.

230. Yoshizako, T., et al., *Usefulness of diffusion-weighted imaging and dynamic contrast-enhanced magnetic resonance imaging in the diagnosis of prostate transition-zone cancer.* Acta Radiologica., 2008. **49**(10): p. 1207-13.

231. Arsov, C., et al., *Prospective randomized trial comparing magnetic resonance imaging (MRI)-guided in-bore biopsy to MRI-ultrasound fusion and transrectal ultrasound-guided prostate biopsy in patients with prior negative biopsies.* European Urology, 2015. **68**(4): p. 713-20.

232. Schoots, I.G., et al., *Magnetic resonance imaging-targeted biopsy may enhance the diagnostic accuracy of significant prostate cancer detection compared to standard transrectal ultrasound-guided biopsy: a systematic review and meta-analysis.* Eur Urol, 2015. **68**(3): p. 438-50.

233. Willis, S.R., et al., *Multiparametric MRI followed by targeted prostate biopsy for men with suspected prostate cancer: a clinical decision analysis.* BMJ Open, 2014. **4**(6).

234. Soyka, J.D., et al., *Clinical impact of 18F-choline PET/CT in patients with recurrent prostate cancer.* European Journal of Nuclear Medicine and Molecular Imaging, 2012. **39**(6): p. 936-943.

235. Wang, H.J., et al., *Value of normalized apparent diffusion coefficient for estimating histological grade of vesical urothelial carcinoma.* Clinical Radiology, 2014. **69**(7): p. 727-31.

236. Groenendaal, G., et al., *Validation of functional imaging with pathology for tumour delineation in the prostate.* Radiotherapy and Oncology, 2010. **94**: p. 145-150.

237. Wang, J., et al., *Machine learning-based analysis of MR radiomics can help to improve the diagnostic performance of PI-RADS v2 in clinically relevant prostate cancer.* Eur Radiol, 2017. **27**(10): p. 4082-4090.

238. Sersa, I., et al., *Multiparametric High-Resolution MRI as a Tool for Mapping of Hypoxic Level in Tumors.* Technology in Cancer Research and Treatment, 2018. **17**(1-8).

239. Ma, T., et al., *Apparent diffusion coefficients in prostate cancer: correlation with molecular markers Ki-67, HIF-1alpha and VEGF. LID - 10.1002/nbm.3884 [doi].* NMR in Biomedicine, 2018. **31**(3).

240. Zelhof, B., et al., *Correlation of diffusion-weighted magnetic resonance data with cellularity in prostate cancer.* BJU International, 2009. **103**(7): p. 883-888.

241. Elekta. *https://www.elekta.com/meta/press-all/?id=0657A60BD7964523*. [cited 2020 January 22nd].

242. Bianconi, F., et al., *Evaluation of Shape and Textural Features from CT as Prognostic Biomarkers in Non-small Cell Lung Cancer.* Anticancer Research, 2018. **38**(4): p. 2155-2160.

243. Xu, X., et al., *Quantitative Identification of Nonmuscle-Invasive and Muscle-Invasive Bladder Carcinomas: A Multiparametric MRI Radiomics Analysis.* Journal of Magnetic Resonance Imaging, 2019. **49**(5): p. 1489-98.

244. Cuocolo, R., et al., *Clinically significant prostate cancer detection on MRI: A radiomic shape features study.* European Journal of Radiology, 2019. **116**: p. 144-149.

245. Ehler, E.D., et al., *Patient specific 3D printed phantom for IMRT quality assurance.* Physics in Medicine and Biology, 2014. **59**(19): p. 5763-73.

246. Traverso, A., et al., *Repeatability and Reproducibility of Radiomic Features: A Systematic Review.* Int J Radiation Oncol Biol Phys, 2018. **102**(4): p. 1143-1158.

247. Walker, A., et al., *Continuous table acquisition MRI for radiotherapy treatment planning: distortion assessment with a new extended 3D volumetric phantom.* Medical Physics, 2015. **42**(4): p. 1982-91.

248. 1, A.M.S.T.G., *Acceptance testing and quality assurance procedures for magnetic resonance imaging facilities*. 2010. p. 1-32.

249. Hollman, B., et al., *Gross tumor volume and clinical target volume in prostate cancer: How do satellites relate to the index lesion.* Radiotherapy and Oncology, 2015. **115**: p. 96-100.

250. O'Callaghan, J., et al., *Is your system calibrated? MRI gradient system calibration for pre-clinical, high-resolution imaging.* PLoS One, 2014. **9**(5).

251. Wang, D., G. Doddrell Dm Fau - Cowin, and G. Cowin, *A novel phantom and method for comprehensive 3-dimensional measurement and correction of geometric distortion in magnetic resonance imaging.* Magnetic Resonance Imaging, 2004. **22**(4): p. 529-42.

252. Doran, S.J., et al., *A complete distortion correction for MR images: I. Gradient warp correction.* Physics in Medicine and Biology, 2005. **50**(7): p. 1343-61.

253. Lips, I.M., et al., *Single blind randomized phase III trial to investigate the benefit of a focal lesion ablative microboost in prostate cancer (FLAME-trial): study protocol for a randomized controlled trial.* Trials, 2011. **12**: p. 255.

254. Forde, E., J. Booth, and M. Leech, *A volumetric approach to path-length measurements is essential when treating radiotherapy with modulated beams.* Medical Dosimetry, 2014. **39**(2): p. 194-96.

255. Stanescu, T., D.A. Wachowicz K Fau - Jaffray, and D.A. Jaffray, *Characterization of tissue magnetic susceptibility-induced distortions for MRIgRT.* Medical Physics, 2012. **39**(12): p. 7185-93.

256. Naqa, I., *The role of quantitative PET in predicting cancer treatment outcomes.* Clin. Transl. Imaging, 2014. **2**(4): p. 305-320.

257. Leijenaar, R.T., et al., *External validation of a prognostic CT-based radiomic signature in oropharyngeal squamous cell carcinoma.* Acta Oncol, 2015. **54**(9): p. 1423-9.

258. Zhao, B., et al., *Exploring Variability in CT Characterization of Tumors: A Preliminary Phantom Study.* Translational Oncology, 2014. **7**(1): p. 88-93.

259. Pavic, M.A.-O.h.o.o., et al., *Influence of inter-observer delineation variability on radiomics stability in different tumor sites.* Acta Oncologica, 2018. **57**(8): p. 1070-74.

260. Chen, S., et al., *MR image-based synthetic CT for IMRT prostate treatment planning and CBCT image-guided localization.* Journal of Applied Clinical Medical Physics, 2016. **17**(3): p. 236-245.

261. Velazquez, E.R., et al., *Volumetric CT-based segmentation of NSCLC using 3D-Slicer.* Sci Rep, 2013. **3**: p. 3529.

262. Parmar, C., et al., *Robust Radiomics feature quantification using semiautomatic volumetric segmentation.* PLoS One, 2014. **9**(7): p. e102107.

263. Zhao, B., et al., *Exploring variability in CT characterisation of tumours: A preliminary phantom study.*, in *Trans. Oncol.* 2014. p. 88-93.

264. Bruner, D.W., et al., *Predictors of preferences and utilities in men treated with 3D-CRT for prostate cancer.* International Journal of Radiation Oncology, Biology, Physics, 2004. **58**(1): p. 34-42.

265. Karavitakis, M., et al., *Tumor focality in prostate cancer: implications for focal therapy.* Nature Reviews Clinical Oncology, 2011. **8**(1): p. 48-55.

266. Eggener, S.E., et al., *Predicting 15-year prostate cancer specific mortality after radical prostatectomy.* Journal of Urology, 2011. **185**(3): p. 869-75.

267. Baco, E., et al., *Magnetic resonance imaging-transectal ultrasound image-fusion biopsies accurately characterize the index tumor: correlation with step-sectioned radical prostatectomy specimens in 135 patients.* European Urology, 2015. **67**(4): p. 787-94.

268. Haffner Mc Fau - Mosbruger, T., et al., *Tracking the clonal origin of lethal prostate cancer.* The Journal of Clinical Investigation, 2013. **123**(11): p. 4918-4922.

269. Altazi, B.A., et al., *Reproducibility of F18-FDG PET radiomic features for different cervical tumor segmentation methods, gray-level discretization, and reconstruction algorithms.* Journal of Applied Clinical Medical Physics, 2017. **18**(6): p. 32-48.

270. Fave, X., et al., *Impact of image preprocessing on the volume dependence and prognostic potential of radiomics features in non-small cell lung cancer.* Translational Cancer Research, 2016. **5**(4): p. 349-363.

271. Gerlinger, M., et al., *Intratumor heterogeneity and branched evolution revealed by multiregion sequencing.* N Engl J Med, 2012. **366**(10): p. 883-92.

272. Jain, S., et al., *Gleason Upgrading with Time in a Large Prostate Cancer Active Surveillance Cohort.* J Urol, 2015. **194**(1): p. 79-84.

273. Klotz, L., et al., *Long-term follow-up of a large active surveillance cohort of patients with prostate cancer.* J Clin Oncol, 2015. **33**(3): p. 272-7.

274. Coates, J., L. Souhami, and I. El Naqa, *Big Data Analytics for Prostate Radiotherapy.* Front Oncol, 2016. **6**: p. 149.

275. Cook, G., et al., *Radiomics in PET: Principles and Applications.* Clin Transl Imaging, 2014. **2**(3): p. 269-276.

276. Davey, A., et al., *Is tumour sphericity an important prognostic factor in patients with lung cancer? LID - S0167-8140(19)33042-7 [pii] LID - 10.1016/j.radonc.2019.08.003 [doi].* Radiotherapy and Oncology, 2020. **143**: p. 73-80.

277. Chatterjee, A., et al., *Changes in Epithelium, Stroma, and Lumen Space Correlate More Strongly with Gleason Pattern and Are Stronger Predictors of Prostate ADC Changes than Cellularity Metrics.* Radiology, 2015. **277**(3): p. 751-62.

278. Oto, A., et al., *Diffusion-weighted and dynamic contrast-enhanced MRI of prostate cancer: correlation of quantitative MR parameters with Gleason score and tumor angiogenesis.* American Journal of Roentgenology, 2011. **197**(6): p. 1382-1390.

279. Wright, J.L., et al., *Prostate cancer specific mortality and Gleason 7 disease differences in prostate cancer outcomes between cases with Gleason 4 + 3 and Gleason 3 + 4 tumors in a population based cohort.* Journal of Urology, 2009. **182**(6): p. 2702-7.

280. Cheung, R., et al., *ROC optimization may improve risk stratification of prostate cancer patients.* Urology, 2001. **57**(2): p. 286-90.

281. Osman, S.O.S., et al., *Computed Tomography-based Radiomics for Risk Stratification in Prostate Cancer.* Int J Radiat Oncol Biol Physics, 2019. **105**(2): p. 448-456.

282. Miles B Fau - Ittmann, M., et al., *Moving Beyond Gleason Scoring.* Arch Pathol Lab Med, 2019. **143**(5): p. 565-70.

283. Dong, X., et al., *Three-dimensional positron emission tomography image texture analysis of esophageal squamous cell carcinoma: relationship between tumor 18F-fluorodeoxyglucose uptake heterogeneity, maximum standardized uptake value, and tumor stage.* Nuclear Medicine Communications, 2013. **34**(1): p. 40-6.

284. Watson, M.J., et al., *Risk stratification of prostate cancer: integrating multiparametric MRI, nomograms and biomarkers.* Future Oncology, 2016. **12**(21): p. 2417-2430.

285. Hatakeyama, S.A.-O.h.o.o., et al., *Recent progress and perspectives on prostate cancer biomarkers.* International Journal of Clinical Oncology, 2017. **22**(2): p. 214-221.

286. Gershman, B., et al., *Impact of Prostate-specific Antigen (PSA) Screening Trials and Revised PSA Screening Guidelines on Rates of Prostate Biopsy and Postbiopsy Complications.* European Urology, 2017. **71**(1): p. 55-65.

287. Jue, J.S., et al., *Re-examining Prostate-specific Antigen (PSA) Density: Defining the Optimal PSA Range and Patients for Using PSA Density to Predict Prostate Cancer Using Extended Template Biopsy.* Urology, 2017. **105**: p. 123-128.

288. Qi, Y.A.-O.h.o.o., et al., *Multiparametric MRI-Based Radiomics for Prostate Cancer Screening With PSA in 4-10 ng/mL to Reduce Unnecessary Biopsies. LID -*

10.1002/jmri.27008 [doi]. Journal of Magnetic Resonance Imaging, 2020. **51**(6): p. 1890-1899.

289. Bailly, C., et al., *Revisiting the Robustness of PET-Based Textural Features in the Context of Multi-Centric Trials.* PLoS One, 2016. **11**(7).

290. McPartlin, A.J., et al., *MRI-guided prostate adaptive radiotherapy - A systematic review.* Radiother Oncol, 2016. **119**(3): p. 371-80.

291. Toivonen, J.A.-O.h.o.o., et al., *Radiomics and machine learning of multisequence multiparametric prostate MRI: Towards improved non-invasive prostate cancer characterization.* PLoS One, 2019. **14**(7).

292. Malinen, E., et al., *Adapting radiotherapy to hypoxic tumours.* Physics in Medicine and Biology, 2006. **51**(19): p. 4903-21.

293. Hamming-Vrieze, O., et al., *Biological PET-guided adaptive radiotherapy for dose escalation in head and neck cancer: a systematic review.* The Quarterly Journal of Nuclear Medicine and Molecular Imaging, 2018. **62**(4): p. 349-68.

294. Wong, K.H., et al., *The emerging potential of magnetic resonance imaging in personalizing radiotherapy for head and neck cancer: an oncologist's perspective.* British Journal of Radiology, 2017. **90**(1071).

295. King, A.D., et al., *Head and neck squamous cell carcinoma: diagnostic performance of diffusion-weighted MR imaging for the prediction of treatment response.* Radiology, 2013. **266**(2): p. 531-8.

296. Welch, M.L., et al., *Vulnerabilities of radiomic signature development: The need for safeguards.* Radiotherapy and Oncology, 2019. **130**: p. 2-9.

297. Parmar, C., et al., *Machine Learning methods for Quantitative Radiomic Biomarkers.* Sci Rep, 2015. **5**: p. 13087.

298. Lever, J., M. Krzywinski, and N. Altman, *Principal component analysis.* Nature Methods, 2017. **14**(7): p. 641-642.

299. Yip, S.S., et al., *Associations between somatic mutations and metabolic imaging phenotypes in non-small cell lung cancer.* J Nucl Med, 2017. **58**(4): p. 569-576.

300. Peeken, J.C., et al., *Radiomics in radiooncology - Challenging the medical physicist.* Physics Medica, 2018. **48**: p. 27-36.

301. Mohri, M., A. Rostamizadeh, and A. Talwalkar, *Foundations of Machine Learning*. 2012, Cambridge, MA: MIT Press.

302. Wu, W., et al., *Exploratory Study to Identify Radiomics Classifiers for Lung Cancer Histology.* Front Oncol, 2016. **6**: p. 71.

303. Thompson, R.F., et al., *Artificial Intelligence in Radiation Oncology Imaging.* Int J Radiat Oncol Biol Phys, 2018. **102**(4): p. 1159-61.

304. Meyer, P., et al., *Survey on deep learning for radiotherapy.* Computers in Biology and Medicine, 2018. **98**(1879-0534 (Electronic)): p. 126-146.

305. LeCun, Y., Y. Bengio, and G. Hinton, *Deep learning.* Nature, 2015. **521**(7553): p. 436-444.

306. Litjens, G., et al., *A survey on deep learning in medical image analysis.* Medical Image Analysis, 2017. **42**: p. 60-88.

307. Bibault, J.E., P. Giraud, and A. Burgun, *Big Data and machine learning in radiation oncology: State of the art and future prospects.* Cancer Letters, 2016. **382**(1): p. 110-117.

308. Mayo, C.S., et al., *The big data effort in radiation oncology: Data mining or data farming?* Advances in Radiation Oncology, 2016. **1**(4): p. 260-271.

309. Schuler, T., et al., *Big Data Readiness in Radiation Oncology: An Efficient Approach for Relabeling Radiation Therapy Structures With Their TG-263 Standard Name in Real-World Data Sets.* Advances in Radiation Oncology, 2019. **4**(1): p. 191-200.

310. Hosny, A.A.-O.h.o.o.X., et al., *Artificial intelligence in radiology.* Nature Reviews Cancer, 2018. **18**(8): p. 500-510.

311. Kuo, R.J., et al., *Application of a two-stage fuzzy neural network to a prostate cancer prognosis system.* Artificial Intelligence in Medicine, 2015. **63**(2): p. 119-133.

312. Thompson, R.F., et al., *The Future of Artificial Intelligence in Radiation Oncology.* Int J Radiat Oncol Biol Phys, 2018. **102**(2): p. 247-8.

313. Nagel, K.N., et al., *Differentiation of Prostatitis and Prostate Cancer by Using Diffusion-weighted MR Imaging and MR-guided Biopsy at 3 T.* Radiology, 2013. **267**(1): p. 164-72.

314. Lambin, P., et al., *Radiomics: the bridge between medical imaging and personalized medicine.* Nature Reviews Clinical Oncology, 2017. **14**(12): p. 749-762.

315. Park, J.E., et al., *Quality of science and reporting of radiomics in oncologic studies: room for improvement according to radiomics quality score and TRIPOD statement.* European Radiology, 2020. **30**(1): p. 523-36.

316. Sala, E., et al., *Review: Unravelling tumour heterogeneity using next-generation imaging: radiomics, radiogenomics, and habitat imaging.* Clinical Radiology, 2017. **72**: p. 3-10.

317. Lambin, P., et al., *'Rapid Learning health care in oncology' - an approach towards decision support systems enabling customised radiotherapy'.* Radiotherapy and Oncology, 2013. **109**(1): p. 159-64.

Patient ID	Age	PreOp PSA	cT stage	GS at biopsy	D'amico risk	Pimo score
2	54	7.4	T3a	7 (4+3)	3	4
3	74	18	T2x	7 (3+4)	2	No PIMO
5	62	8.8	T2c	7 (3+4)	3	2
6	73	16	T1c	8	3	3.5
7	63	7.9	T2x	7(3+4)	2	2.5
8	64	29	T3x	7 (4+3)	3	2
9	70	2.5	T2a	7 (4+3)	2	3
11	70	11	T2a	7 (3+4)	2	No PIMO
12	65	10	T1c	7 (4+3)	2	2
13	61	10	T3x	7 (4+3)	3	No PIMO
14	74	42	T1c	7 (3+4)	3	No PIMO
15	67	18	T2a	7 (4+3)	2	5
17	56	7.6	T2x	7 (3+4)	2	1
18	70	13	T1c	8	3	1
19	61	8.8	T1c	8	3	No PIMO
21	64	6.2	T2a	8	3	3
22	62	8.3	T2x	7 (3+4)	2	2
23	66	16	T1c	8	3	1
24	71	14	T1c	7 (3+4)	2	2.5
25	68	6.3	T2b	7 (3+4)	2	No PIMO
26	64	12	T2a	8	3	No PIMO
28	73	11	T3a	8	3	3.5
30	67	38	T1c	7 (4+3)	3	2
31	66	9.4	T2a	7 (3+4)	2	3
32	56	15	Not stated	9	3	0
33	75	14	T2c	6	3	4

34	64	7.7	T2a	7 (3+4)	2	2
35	52	5.9	T1c	7 (3+4)	2	2.5
36	64	12	T1c	7 (4+3)	2	0.5
37	64	19	T2b	7 (3+4)	2	3
39	61	19	T2c	8	3	1
40	68	11	T3a	7 (4+3)	3	2
42	60	10	T1c	7 (3+4)	2	1.5
43	70	6.5	T3a	9	3	3.5
45	66	8	T2a	7 (4+3)	2	2.5
46	70	4.3	T2b	7 (4+3)	2	3
47	72	12	T1c	7 (3+4)	2	2
48	55	7.5	T1c	7 (3+4)	2	No PIMO
49	74	5.5	T3a	8	3	4
50	57	4.6	T1c	8	3	4
51	60	12	T3a	7 (3+4)	3	1.5
52	62	20	T1c	7 (4+3)	3	2.5
53	58	4.8	T2b	9	3	2.5
55	47	3.2	T2x	7 (3+4)	2	1
56	67	145	T1c	8	3	1
58	71	8.7	T3x	8	3	No PIMO
59	66	4.8	T1c	7 (3+4)	2	0.5
60	61	9.3	T1c	6	1	0
61	59	12	T1c	7 (3+4)	2	2.5
62	60	8.7	T2b	8	3	3
64	65	8.5	T1c	6	1	0.5
70	60	7.4	Not stated	8	3	2.5
71	64	14	T2x	8	3	2.5

72	50	3.2	T1c	7 (3+4)	2	0
73	67	6.9	Not stated	7 (3+4)	2	2.5
74	67	9.5	T1c	7 (3+4)	2	3
76	69	11	T1c	7 (4+3)	2	1.5
77	69	12	Not stated	8	3	0
78	68	19	T1c	7 (3+4)	2	0.5
79	50	8.9	T2x	6	2	0
80	55	11	T1c	8	3	3
81	67	2.2	T2x	7 (3+4)	2	3
82	74	4.7	T2x	7 (3+4)	2	3
83	52	4.2	T2x	6	2	3
84	68	6.8	T1c	7 (3+4)	2	0.5
86	61	4.5	T2c	8	3	0.5
87	53	2.2	T1c	7 (3+4)	2	0
89	66	11	T2x	7 (3+4)	2	2
90	68	9.6	T1c	7 (3+4)	2	0
91	69	4.6	T2b	7 (4+3)	2	No PIMO
92	56	38	T1c	7 (3+4)	3	2.5
93	52	7.9	Not stated	7 (3+4)	2	0
94	66	4.5	T2a	7 (3+4)	2	3.5
95	69	5	T2b	8	3	2
96	58	9.1	T2x	7 (3+4)	2	0
98	59	4.6	T1c	7 (3+4)	2	3
99	63	23	T2b	8	3	1.5
100	63	10	T2x	7 (4+3)	2	0
102	74	12	T1c	7 (4+3)	2	0
105	60	6.4	T1c	7 (3+4)	2	1

106	74	8.3	T1c	8	3	1.5
108	61	7.9	T3x	7 (3+4)	3	0
111	68	37	T3a	8	3	4
113	68	8.4	T2x	7 (4+3)	2	0
114	68	8.1	T1c	7 (3+4)	2	0.5
118	69	9	T3x	7 (3+4)	3	3.5
119	71	6.2	T1c	7 (4+3)	2	2
120	59	5.4	T1c	9	3	1.5

Appendix 1: Patient Characteristics for T2-weighted Radiomics Analysis.

Age is in years at inclusion, PSA is given in ng/mL, D'amico risk stratification : 1= low risk, 2= intermediate risk, 3= high risk.

Patient ID	Age	Pre-op PSA	cT stage	GS at biopsy	D'amico risk classification	PIMO score
2	54	7.4	T3a	7 (4+3)	3	4
3	74	18	T2x	7 (3+4)	2	No PIMO
4	66	19	T1c	6	2	3.5
5	62	8.8	T2c	7 (3+4)	3	2
6	73	16	T1c	8	3	3.5
7	63	7.9	T2x	7 (3+4)	2	2.5
8	64	29	Not stated	7 (4+3)	3	2
9	70	2.5	T2a	7 (4+3)	2	3
10	73	2	T1c	7 (3+4)	2	0.5
11	70	11	T1c	7 (3+4)	2	No PIMO
12	65	10	T1c	7 (4+3)	2	2
13	61	10	T3x	7 (4+3)	3	No PIMO
14	74	42	T1c	7 (3+4)	3	No PIMO
15	67	18	T2a	7 (4+3)	2	5
16	65	2.7	T1c	8	3	0.5
17	56	7.6	T2x	7 (3+4)	2	1
18	70	13	T1c	8	3	1
19	61	8.8	T1c	8	3	No PIMO
20	70	13	T1c	7 (3+4)	2	No PIMO
21	64	6.2	T2a	8	3	3
22	62	8.3	T2x	7 (3+4)	2	2
23	66	16	T1c	8	3	1
24	71	14	T1c	7 (3+4)	2	2.5
25	68	6.3	T2b	7 (3+4)	2	No PIMO
26	64	12	T2a	8	3	No PIMO

27	62	11	T3a	7 (3+4)	3	2
28	73	11	T3a	8	3	3.5
29	61	35	T2c	7 (4+3)	3	2
30	67	38	T1c	7 (4+3)	3	2
31	66	9.4	T2a	7 (3+4)	2	3
32	56	15	Not stated	9	3	0
33	75	14	T2c	6	3	4
34	64	7.7	T2a	7 (3+4)	2	2
35	52	5.9	T1c	7 (3+4)	2	2.5
36	64	12	T1c	7 (4+3)	2	0.5
37	64	19	T2b	7 (3+4)	2	3
38	61	7.9	Not stated	8	3	2.5
39	61	19	T2c	8	3	1
40	68	11	T3a	7 (4+3)	3	2
42	60	10	T1c	7 (3+4)	2	1.5
43	70	6.5	T3a	9	3	3.5
44	63	30	T3a	9	3	2.5
45	66	8	T2a	7 (4+3)	2	2.5
46	70	4.3	T2b	7 (4+3)	2	3
47	72	12	T1c	7 (3+4)	2	2
48	55	7.5	T1c	7 (3+4)	2	No PIMO
49	74	5.5	T3a	8	3	4
50	57	4.6	T1c	8	3	4
51	60	12	T3a	7 (3+4)	3	1.5
52	62	20	T1c	7 (4+3)	3	2.5
53	58	4.8	T2b	9	3	2.5
55	47	3.2	T2x	7 (3+4)	2	1
56	67	145	T1c	8	3	1
58	71	8.7	T3x	8	3	No PIMO
59	66	4.8	T1c	7 (3+4)	2	0.5

60	61	9.3	T1c	6	1	0
61	59	12	T1c	7 (3+4)	2	2.5
62	60	8.7	T2b	8	3	3
63	56	7	T2c	6	3	0
64	65	8.5	T1c	6	1	0.5
66	67	5.7	T1c	9	3	No PIMO
67	69	3.8	T2x	7 (4+3)	2	0.5
68	60	3.8	T1c	6	1	2.5
69	62	7.1	Not stated	7 (3+4)	2	0
70	60	7.4	Not stated	8	3	2.5
71	64	14	T2x	8	3	2.5
72	50	3.2	T1c	7 (3+4)	2	0
73	67	6.9	Not stated	7 (3+4)	2	2.5
74	67	9.5	T1c	7 (3+4)	2	3
75	66	8.2	T1c	8	3	2
76	69	11	T1c	7 (4+3)	2	1.5
77	69	12	Not given	8	3	0
78	68	19	T1c	7 (3+4)	2	0.5
79	50	8.9	T2x	6	2	0
80	55	11	T1c	8	3	3
81	67	2.2	T2x	7 (3+4)	2	3
82	74	4.7	T2x	7 (3+4)	2	3
83	52	4.2	T2x	6	2	3
84	68	6.8	T1c	7 (3+4)	2	0.5
85	76	10	T2c	7 (3+4)	3	4
86	61	4.5	T2c	8	3	0.5
87	53	2.2	T1c	7 (3+4)	2	0
89	66	11	T2x	7 (3+4)	2	2
90	68	9.6	T1c	7 (3+4)	2	0
91	69	4.6	T2b	7 (4+3)	2	No PIMO

93	52	7.9	Not stated	7 (4+3)	2	0
94	66	4.5	T2a	7 (4+3)	2	3.5
95	69	5	T2b	8	3	2
96	58	9.1	T2x	7 (3+4)	2	0
98	59	4.6	T1c	7 (3+4)	2	3
99	63	23	T2b	8	3	1.5
100	63	10	T2x	7 (4+3)	2	0
102	74	12	T1c	7 (4+3)	2	0
104	66	7.1	T1c	8	3	3
105	60	6.4	T1c	7 (3+4)	2	1
106	74	8.3	T1c	8	3	1.5
108	61	7.9	T3x	7 (3+4)	3	0
109	65	13	T3x	8	3	2.5
111	68	37	T3a	8	3	4
113	68	8.4	T2x	7 (4+3)	2	0
114	68	8.1	T1c	7 (3+4)	2	0.5
118	**69**	9	T3x	7 (3+4)	3	3.5
119	**71**	6.2	T1c	7 (4+3)	2	2
120	59	5.4	T1c	9	3	1.5

Appendix 2: Patient Characteristics for Diffusion-Weighted Radiomics Analysis.

Age is in years at inclusion, PSA is given in ng/mL, D'amico risk stratification : 1= low risk, 2= intermediate risk, 3= high risk.

Patient ID	Age	Pre-op PSA	cT stage	GS at biopsy	D'amico risk classification	PIMO score
2	54	7.4	T3a	7 (4+3)	3	4
3	74	18	T2x	7 (3+4)	2	No PIMO
4	66	19	T1c	6	2	3.5
5	62	8.8	T2c	7 (3+4)	3	2
6	73	16	T1c	8	3	3.5
7	63	7.9	T2x	7 (3+4)	2	2.5
8	64	29	Not stated	7 (4+3)	3	2
9	70	2.5	T2a	7 (4+3)	2	3
10	73	2	T1c	7 (3+4)	2	0.5
11	70	11	T1c	7 (3+4)	2	No PIMO
12	65	10	T1c	7 (4+3)	2	2
13	61	10	T3x	7 (4+3)	3	No PIMO
14	74	42	T1c	7 (3+4)	3	No PIMO
15	67	18	T2a	7 (4+3)	2	5
17	56	7.6	T2x	7 (3+4)	2	1
19	61	8.8	T1c	8	3	No PIMO
20	70	13	T1c	7 (3+4)	2	No PIMO
21	64	6.2	T2a	8	3	3
22	62	8.3	T2x	7 (3+4)	2	2
23	66	16	T1c	8	3	1
24	71	14	T1c	7 (3+4)	2	2.5
25	68	6.3	T2b	7 (3+4)	2	No PIMO

26	64	12	T2a	8	3	No PIMO
27	62	11	T3a	7 (3+4)	3	2
28	73	11	T3a	8	3	3.5
29	61	35	T2c	7 (4+3)	3	2
30	67	38	T1c	7 (4+3)	3	2
31	66	9.4	T2a	7 (3+4)	2	3
32	56	15	Not stated	9	3	0
33	75	14	T2c	6	3	4
34	64	7.7	T2a	7 (3+4)	2	2
35	52	5.9	T1c	7 (3+4)	2	2.5
36	64	12	T1c	7 (4+3)	2	0.5
37	64	19	T2b	7 (3+4)	2	3
38	61	7.9	Not stated	8	3	2.5
39	61	19	T2c	8	3	1
40	68	11	T3a	7 (4+3)	3	2
42	60	10	T1c	6 (3+4)	2	1.5
43	70	6.5	T3a	9	3	3.5
44	63	30	T3a	9	3	2.5
45	66	8	T2a	7 (4+3)	2	2.5
46	70	4.3	T2b	7 (4+3)	2	3
47	72	12	T1c	6 (3+4)	2	2
48	55	7.5	T1c	7 (3+4)	2	No PIMO
49	74	5.5	T3a	8	3	4
50	57	4.6	T1c	8	3	4
51	60	12	T3a	7 (3+4)	3	1.5
52	62	20	T1c	7 (4+3)	3	2.5
53	58	4.8	T2b	9	3	2.5
54	59	7.8	T2b	8	3	1.5
56	67	145	T1c	8	3	1
57	45	56	T2x	7 (4+3)	3	2.5

58	71	8.7	T3x	8	3	No PIMO
59	66	4.8	T1c	7 (3+4)	2	0.5
60	61	9.3	T1c	6	1	0
61	59	12	T1c	7 (3+4)	2	2.5
62	60	8.7	T2b	8	3	3
63	56	7	T2c	6	3	0
64	65	8.5	T1c	6	1	0.5
77	69	12	Not given	8	3	0
78	68	19	T1c	7 (3+4)	2	0.5
79	50	8.9	T2x	6	2	0
80	55	11	T1c	8	3	3
81	67	2.2	T2x	7 (3+4)	2	3
82	74	4.7	T2x	7 (3+4)	2	3
83	52	4.2	T2x	6	2	3
84	68	6.8	T1c	7 (3+4)	2	0.5
86	61	4.5	T2c	8	3	0.5
87	53	2.2	T1c	7 (3+4)	2	0
89	66	11	T2x	7 (3+4)	2	2
90	68	9.6	T1c	7 (3+4)	2	0
91	69	4.6	T2b	7 (4+3)	2	No PIMO
92	56	38	T1c	7(3+4)	3	2.5
93	52	7.9	Not stated	7 (4+3)	2	0
94	66	4.5	T2a	7 ($+3)	2	3.5
95	69	5	T2b	8	3	2
96	58	9.1	T2x	7 (3+4)	2	0
97	62	8.7	T2a	7 (4+3)	2	2
98	59	4.6	T1c	7 (3+4)	2	3
99	63	23	T2b	8	3	1.5
100	63	10	T2x	7 (4+3)	2	0
102	74	12	T1c	7 (4+3)	2	0

104	66	7.1	T1c	8	3	3
105	60	6.4	T1c	7 (3+4)	2	1
106	74	8.3	T1c	8	3	1.5
108	61	7.9	T3x	7 (3+4)	3	0
109	65	13	T3x	8	3	2.5
111	68	37	T3a	8	3	4
113	68	8.4	T2x	7 (4+3)	2	0
114	68	8.1	T1c	7 (3+4)	2	0.5
118	**69**	9	T3x	7 (3+4)	3	3.5
119	**71**	6.2	T1c	7 (4+3)	2	2
120	59	5.4	T1c	9	3	1.5

Appendix 3: Patient Characteristics for ADC Radiomics Analysis.

Age is in years at inclusion, PSA is given in ng/mL, D'amico risk stratification : 1= low risk, 2= intermediate risk, 3= high risk.

Appendix 4. Total Variance Explained Table from Principal Components Analysis for T2 weighted analysis (Chapter 5)

Feature Number	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	*% of Variance*	*Cumulative %*	*Total*	*% of Variance*	Cumulative %	*Total*	% of Variance	*Cumulative %*
1	64.959	39.369	39.369	64.959	39.369	39.369	50.670	30.709	30.709
2	35.060	21.248	60.618	35.060	21.248	60.618	20.937	12.689	43.398
3	10.686	6.477	67.094	10.686	6.477	67.094	18.518	11.223	54.621
4	8.303	5.032	72.127	8.303	5.032	72.127	16.386	9.931	64.552
5	6.979	4.229	76.356	6.979	4.229	76.356	8.133	4.929	69.481
6	5.258	3.187	79.543	5.258	3.187	79.543	7.816	4.737	74.218
7	4.380	2.655	82.197	4.380	2.655	82.197	7.722	4.680	78.897
8	3.917	2.374	84.571	3.917	2.374	84.571	4.451	2.697	81.595
9	3.024	1.833	86.404	3.024	1.833	86.404	3.679	2.230	83.825
10	2.578	1.563	87.967	2.578	1.563	87.967	2.779	1.684	85.509
11	1.865	1.130	89.097	1.865	1.130	89.097	2.604	1.578	87.087
12	1.801	1.091	90.188	1.801	1.091	90.188	2.581	1.564	88.651
13	1.544	.936	91.123	1.544	.936	91.123	2.383	1.444	90.095
14	1.338	.811	91.934	1.338	.811	91.934	1.971	1.195	91.290
15	1.204	.730	92.664	1.204	.730	92.664	1.831	1.110	92.400
16	1.125	.682	93.345	1.125	.682	93.345	1.561	.946	93.345
17	.982	.595	93.940						
18	.925	.560	94.501						
19	.772	.468	94.969						
20	.736	.446	95.415						

21	.641	.388	95.803						
22	.575	.348	96.151						
23	.508	.308	96.459						
24	.500	.303	96.762						
25	.398	.242	97.004						
26	.386	.234	97.238						
27	.359	.218	97.456						
28	.328	.199	97.654						
29	.314	.190	97.845						
30	.288	.174	98.019						
31	.270	.164	98.183						
32	.255	.155	98.338						
33	.225	.136	98.474						
34	.197	.120	98.594						
35	.189	.115	98.708						
36	.183	.111	98.819						
37	.164	.099	98.918						
38	.146	.088	99.007						
39	.132	.080	99.087						
40	.121	.073	99.160						
41	.112	.068	99.228						
42	.100	.061	99.289						
43	.094	.057	99.346						
44	.092	.056	99.402						
45	.085	.051	99.453						
46	.080	.049	99.502						

47	.071	.043	99.545						
48	.064	.039	99.583						
49	.061	.037	99.620						
50	.056	.034	99.654						
51	.051	.031	99.685						
52	.046	.028	99.713						
53	.042	.026	99.739						
54	.037	.022	99.761						
55	.035	.021	99.782						
56	.033	.020	99.802						
57	.028	.017	99.819						
58	.026	.016	99.835						
59	.025	.015	99.850						
60	.024	.014	99.864						
61	.022	.013	99.878						
62	.020	.012	99.890						
63	.019	.011	99.901						
64	.017	.010	99.911						
65	.015	.009	99.921						
66	.014	.008	99.929						
67	.013	.008	99.937						
68	.012	.008	99.944						
69	.011	.006	99.951						
70	.009	.005	99.956						
71	.008	.005	99.961						
72	.007	.005	99.965						

73	.007	.004	99.969						
74	.006	.004	99.973						
75	.006	.003	99.977						
76	.005	.003	99.979						
77	.004	.003	99.982						
78	.004	.002	99.985						
79	.004	.002	99.987						
80	.003	.002	99.989						
81	.003	.002	99.991						
82	.003	.002	99.992						
83	.002	.001	99.993						
84	.002	.001	99.995						
85	.002	.001	99.996						
86	.002	.001	99.996						
87	.001	.001	99.997						
88	.001	.001	99.998						
89	.001	.001	99.998						
90	.001	.000	99.999						
91	.001	.000	99.999						
92	.001	.000	100.000						
93	.000	.000	100.000						
94	.000	.000	100.000						
95	1.670E-14	1.012E-14	100.000						
96	1.589E-15	9.629E-16	100.000						
97	1.142E-15	6.921E-16	100.000						
98	1.036E-15	6.280E-16	100.000						

99	9.088E-16	5.508E-16	100.000						
100	8.742E-16	5.298E-16	100.000						
101	7.783E-16	4.717E-16	100.000						
102	7.052E-16	4.274E-16	100.000						
103	6.977E-16	4.228E-16	100.000						
104	6.678E-16	4.047E-16	100.000						
105	5.775E-16	3.500E-16	100.000						
106	5.344E-16	3.239E-16	100.000						
107	5.321E-16	3.225E-16	100.000						
108	4.837E-16	2.931E-16	100.000						
109	4.653E-16	2.820E-16	100.000						
110	3.972E-16	2.407E-16	100.000						
111	3.694E-16	2.239E-16	100.000						
112	3.290E-16	1.994E-16	100.000						
113	3.031E-16	1.837E-16	100.000						
114	2.519E-16	1.527E-16	100.000						
115	2.488E-16	1.508E-16	100.000						
116	2.216E-16	1.343E-16	100.000						
117	1.874E-16	1.136E-16	100.000						
118	1.710E-16	1.036E-16	100.000						
119	1.500E-16	9.093E-17	100.000						
120	1.356E-16	8.216E-17	100.000						
121	1.269E-16	7.689E-17	100.000						
122	1.067E-16	6.467E-17	100.000						
123	7.727E-17	4.683E-17	100.000						
124	4.874E-17	2.954E-17	100.000						

125	2.530E-17	1.533E-17	100.000						
126	1.110E-17	6.730E-18	100.000						
127	3.589E-20	2.175E-20	100.000						
128	-9.246E-18	-5.604E-18	100.000						
129	-1.958E-17	-1.187E-17	100.000						
130	-3.048E-17	-1.847E-17	100.000						
131	-5.526E-17	-3.349E-17	100.000						
132	-7.975E-17	-4.833E-17	100.000						
133	-1.091E-16	-6.611E-17	100.000						
134	-1.458E-16	-8.838E-17	100.000						
135	-1.788E-16	-1.083E-16	100.000						
136	-2.083E-16	-1.263E-16	100.000						
137	-2.166E-16	-1.313E-16	100.000						
138	-2.199E-16	-1.333E-16	100.000						

139	-2.537E-16	-1.538E-16	100.000						
140	-2.833E-16	-1.717E-16	100.000						
141	-3.017E-16	-1.828E-16	100.000						
142	-3.468E-16	-2.102E-16	100.000						
143	-3.784E-16	-2.294E-16	100.000						
144	-4.144E-16	-2.512E-16	100.000						
145	-4.357E-16	-2.641E-16	100.000						
146	-4.590E-16	-2.782E-16	100.000						
147	-5.006E-16	-3.034E-16	100.000						
148	-5.434E-16	-3.293E-16	100.000						
149	-5.760E-16	-3.491E-16	100.000						
150	-6.230E-16	-3.776E-16	100.000						
151	-6.820E-16	-4.133E-16	100.000						

152	-7.031E-16	-4.261E-16	100.000						
153	-7.489E-16	-4.539E-16	100.000						
154	-7.782E-16	-4.716E-16	100.000						
155	-8.013E-16	-4.856E-16	100.000						
156	-8.712E-16	-5.280E-16	100.000						
157	-9.083E-16	-5.505E-16	100.000						
158	-9.394E-16	-5.694E-16	100.000						
159	-1.011E-15	-6.125E-16	100.000						
160	-1.042E-15	-6.317E-16	100.000						
161	-1.155E-15	-7.001E-16	100.000						
162	-1.482E-15	-8.980E-16	100.000						
163	-3.332E-15	-2.019E-15	100.000						
164	-4.591E-15	-2.782E-15	100.000						

165	-1.048E-14	-6.349E-15	100.000						

Appendix 5. Total Variance Explained Table from Principal Components Analysis for DW-MRI weighted analysis (Chapter 6)

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	71.312	43.219	43.219	71.312	43.219	43.219	52.111	31.582	31.582
2	31.853	19.305	62.524	31.853	19.305	62.524	26.930	16.321	47.904
3	13.892	8.419	70.944	13.892	8.419	70.944	20.927	12.683	60.587
4	7.880	4.776	75.720	7.880	4.776	75.720	20.479	12.412	72.998
5	6.817	4.131	79.851	6.817	4.131	79.851	5.480	3.321	76.320
6	5.285	3.203	83.054	5.285	3.203	83.054	5.066	3.070	79.390
7	3.543	2.147	85.202	3.543	2.147	85.202	5.018	3.041	82.431
8	2.944	1.784	86.986	2.944	1.784	86.986	3.755	2.276	84.707
9	2.500	1.515	88.501	2.500	1.515	88.501	3.496	2.119	86.825
10	2.323	1.408	89.909	2.323	1.408	89.909	2.983	1.808	88.634
11	1.860	1.127	91.037	1.860	1.127	91.037	2.545	1.542	90.176
12	1.573	.953	91.990	1.573	.953	91.990	2.205	1.336	91.512
13	1.468	.890	92.880	1.468	.890	92.880	1.847	1.119	92.632
14	1.082	.656	93.536	1.082	.656	93.536	1.492	.904	93.536

15	.965	.585	94.120						
16	.883	.535	94.656						
17	.822	.498	95.154						
18	.708	.429	95.583						
19	.665	.403	95.986						
20	.605	.366	96.352						
21	.539	.327	96.678						
22	.476	.288	96.967						
23	.427	.259	97.226						
24	.418	.253	97.479						
25	.356	.216	97.695						
26	.344	.208	97.904						
27	.316	.192	98.095						
28	.292	.177	98.272						
29	.262	.159	98.430						
30	.231	.140	98.571						
31	.208	.126	98.697						
32	.205	.124	98.821						
33	.173	.105	98.926						
34	.169	.102	99.028						
35	.155	.094	99.122						
36	.141	.086	99.208						
37	.118	.071	99.279						
38	.107	.065	99.344						
39	.099	.060	99.404						
40	.090	.054	99.458						

41	.088	.053	99.512						
42	.072	.044	99.556						
43	.070	.042	99.598						
44	.058	.035	99.633						
45	.057	.034	99.667						
46	.048	.029	99.697						
47	.045	.027	99.724						
48	.042	.026	99.750						
49	.039	.023	99.773						
50	.034	.020	99.793						
51	.033	.020	99.813						
52	.028	.017	99.830						
53	.027	.016	99.846						
54	.025	.015	99.862						
55	.021	.013	99.874						
56	.021	.013	99.887						
57	.018	.011	99.897						
58	.016	.010	99.907						
59	.015	.009	99.916						
60	.013	.008	99.924						
61	.013	.008	99.932						
62	.012	.007	99.939						
63	.011	.007	99.945						
64	.010	.006	99.951						
65	.008	.005	99.956						
66	.007	.004	99.961						

67	.006	.004	99.964						
68	.006	.004	99.968						
69	.005	.003	99.971						
70	.005	.003	99.974						
71	.005	.003	99.977						
72	.004	.003	99.980						
73	.004	.003	99.982						
74	.004	.002	99.984						
75	.003	.002	99.986						
76	.003	.002	99.988						
77	.002	.001	99.989						
78	.002	.001	99.991						
79	.002	.001	99.992						
80	.002	.001	99.993						
81	.002	.001	99.994						
82	.001	.001	99.995						
83	.001	.001	99.995						
84	.001	.001	99.996						
85	.001	.001	99.996						
86	.001	.000	99.997						
87	.001	.000	99.997						
88	.001	.000	99.998						
89	.001	.000	99.998						
90	.000	.000	99.998						
91	.000	.000	99.999						
92	.000	.000	99.999						

93	.000	.000	99.999						
94	.000	.000	99.999						
95	.000	.000	99.999						
96	.000	.000	100.000						
97	.000	.000	100.000						
98	.000	8.461E-5	100.000						
99	.000	6.145E-5	100.000						
100	9.185E-5	5.567E-5	100.000						
101	5.714E-5	3.463E-5	100.000						
102	5.363E-5	3.250E-5	100.000						
103	4.116E-5	2.495E-5	100.000						
104	8.000E-15	4.848E-15	100.000						
105	2.132E-15	1.292E-15	100.000						
106	1.369E-15	8.296E-16	100.000						
107	1.096E-15	6.641E-16	100.000						
108	1.013E-15	6.141E-16	100.000						

109	9.883E-16	5.990E-16	100.000						
110	9.212E-16	5.583E-16	100.000						
111	8.170E-16	4.951E-16	100.000						
112	7.444E-16	4.511E-16	100.000						
113	7.115E-16	4.312E-16	100.000						
114	6.554E-16	3.972E-16	100.000						
115	5.917E-16	3.586E-16	100.000						
116	5.556E-16	3.367E-16	100.000						
117	4.878E-16	2.956E-16	100.000						
118	4.497E-16	2.725E-16	100.000						
119	4.319E-16	2.618E-16	100.000						
120	4.131E-16	2.503E-16	100.000						
121	3.682E-16	2.231E-16	100.000						

122	3.562E-16	2.159E-16	100.000						
123	3.139E-16	1.902E-16	100.000						
124	2.630E-16	1.594E-16	100.000						
125	2.436E-16	1.476E-16	100.000						
126	2.272E-16	1.377E-16	100.000						
127	2.136E-16	1.295E-16	100.000						
128	1.918E-16	1.163E-16	100.000						
129	1.580E-16	9.576E-17	100.000						
130	1.403E-16	8.501E-17	100.000						
131	1.082E-16	6.557E-17	100.000						
132	8.201E-17	4.971E-17	100.000						
133	4.715E-17	2.858E-17	100.000						
134	2.936E-17	1.779E-17	100.000						

135	2.705E-17	1.640E-17	100.000						
136	1.097E-17	6.649E-18	100.000						
137	-1.726E-17	-1.046E-17	100.000						
138	-5.682E-17	-3.444E-17	100.000						
139	-6.798E-17	-4.120E-17	100.000						
140	-8.332E-17	-5.050E-17	100.000						
141	-1.015E-16	-6.149E-17	100.000						
142	-1.343E-16	-8.138E-17	100.000						
143	-1.782E-16	-1.080E-16	100.000						
144	-2.198E-16	-1.332E-16	100.000						
145	-2.415E-16	-1.464E-16	100.000						
146	-3.029E-16	-1.836E-16	100.000						
147	-3.241E-16	-1.964E-16	100.000						

148	-3.385E-16	-2.052E-16	100.000						
149	-3.696E-16	-2.240E-16	100.000						
150	-4.626E-16	-2.804E-16	100.000						
151	-5.035E-16	-3.051E-16	100.000						
152	-5.478E-16	-3.320E-16	100.000						
153	-5.988E-16	-3.629E-16	100.000						
154	-6.722E-16	-4.074E-16	100.000						
155	-7.027E-16	-4.259E-16	100.000						
156	-7.452E-16	-4.517E-16	100.000						
157	-7.930E-16	-4.806E-16	100.000						
158	-8.720E-16	-5.285E-16	100.000						
159	-9.313E-16	-5.644E-16	100.000						
160	-1.079E-15	-6.541E-16	100.000						

161	-1.119E-15	-6.784E-16	100.000						
162	-1.785E-15	-1.082E-15	100.000						
163	-4.175E-15	-2.530E-15	100.000						
164	-8.761E-15	-5.310E-15	100.000						
165	-1.364E-14	-8.266E-15	100.000						

Appendix 6. Total Variance Explained Table from Principal Components Analysis for ADC analysis (Chapter 5)

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	68.732	41.656	41.656	68.732	41.656	41.656	54.754	33.184	33.184
2	34.153	20.699	62.355	34.153	20.699	62.355	25.914	15.706	48.890
3	10.739	6.509	68.864	10.739	6.509	68.864	18.269	11.072	59.962
4	8.056	4.882	73.746	8.056	4.882	73.746	15.097	9.150	69.112
5	5.771	3.497	77.243	5.771	3.497	77.243	5.972	3.619	72.731
6	4.981	3.019	80.262	4.981	3.019	80.262	5.039	3.054	75.785
7	4.026	2.440	82.702	4.026	2.440	82.702	4.632	2.808	78.592
8	3.369	2.042	84.744	3.369	2.042	84.744	4.234	2.566	81.158

9	2.957	1.792	86.536	2.957	1.792	86.536	3.580	2.169	83.328
10	2.489	1.509	88.045	2.489	1.509	88.045	3.303	2.002	85.330
11	2.260	1.370	89.415	2.260	1.370	89.415	2.932	1.777	87.107
12	1.954	1.184	90.599	1.954	1.184	90.599	2.744	1.663	88.770
13	1.464	.887	91.486	1.464	.887	91.486	2.051	1.243	90.013
14	1.277	.774	92.260	1.277	.774	92.260	1.641	.995	91.007
15	1.176	.713	92.973	1.176	.713	92.973	1.453	.881	91.888
16	1.081	.655	93.628	1.081	.655	93.628	1.441	.874	92.762
17	.922	.559	94.187	.922	.559	94.187	1.371	.831	93.593
18	.847	.513	94.700	.847	.513	94.700	1.297	.786	94.379
19	.714	.433	95.133	.714	.433	95.133	1.243	.754	95.133
20	.681	.413	95.545						
21	.655	.397	95.942						
22	.607	.368	96.310						
23	.535	.324	96.635						
24	.506	.307	96.941						

25	.438	.265	97.207						
26	.402	.244	97.450						
27	.360	.218	97.668						
28	.311	.188	97.857						
29	.298	.181	98.037						
30	.256	.155	98.193						
31	.232	.141	98.333						
32	.222	.135	98.468						
33	.214	.130	98.598						
34	.192	.116	98.714						
35	.175	.106	98.820						
36	.152	.092	98.912						
37	.143	.087	98.999						
38	.134	.081	99.080						
39	.126	.076	99.157						
40	.112	.068	99.224						

41	.107	.065	99.289						
42	.098	.059	99.349						
43	.091	.055	99.404						
44	.077	.046	99.450						
45	.074	.045	99.495						
46	.064	.039	99.534						
47	.061	.037	99.571						
48	.057	.035	99.605						
49	.053	.032	99.637						
50	.052	.031	99.669						
51	.047	.029	99.697						
52	.044	.026	99.724						
53	.042	.025	99.749						
54	.038	.023	99.772						
55	.036	.022	99.794						
56	.032	.020	99.814						

57	.030	.018	99.832						
58	.025	.015	99.848						
59	.023	.014	99.862						
60	.022	.013	99.875						
61	.021	.013	99.888						
62	.019	.012	99.899						
63	.017	.010	99.909						
64	.017	.010	99.919						
65	.014	.009	99.928						
66	.014	.008	99.936						
67	.013	.008	99.944						
68	.011	.007	99.951						
69	.009	.006	99.956						
70	.009	.005	99.962						
71	.007	.004	99.966						
72	.007	.004	99.970						

73	.006	.004	99.974						
74	.005	.003	99.977						
75	.005	.003	99.980						
76	.005	.003	99.982						
77	.004	.002	99.985						
78	.003	.002	99.987						
79	.003	.002	99.989						
80	.003	.002	99.990						
81	.003	.002	99.992						
82	.002	.001	99.993						
83	.002	.001	99.995						
84	.002	.001	99.996						
85	.001	.001	99.997						
86	.001	.001	99.997						
87	.001	.001	99.998						
88	.001	.001	99.998						

89	.001	.000	99.999						
90	.001	.000	99.999						
91	.001	.000	100.000						
92	.000	.000	100.000						
93	.000	.000	100.000						
94	1.123E-14	6.803E-15	100.000						
95	2.776E-15	1.682E-15	100.000						
96	1.324E-15	8.024E-16	100.000						
97	1.204E-15	7.296E-16	100.000						
98	1.012E-15	6.131E-16	100.000						
99	9.604E-16	5.821E-16	100.000						
100	8.944E-16	5.420E-16	100.000						

101	8.210E-16	4.975E-16	100.000						
102	7.512E-16	4.553E-16	100.000						
103	7.277E-16	4.411E-16	100.000						
104	7.105E-16	4.306E-16	100.000						
105	6.636E-16	4.022E-16	100.000						
106	6.350E-16	3.848E-16	100.000						
107	5.948E-16	3.605E-16	100.000						
108	5.904E-16	3.578E-16	100.000						
109	5.436E-16	3.294E-16	100.000						
110	5.063E-16	3.069E-16	100.000						

111	4.408E-16	2.671E-16	100.000						
112	4.264E-16	2.584E-16	100.000						
113	3.897E-16	2.362E-16	100.000						
114	3.564E-16	2.160E-16	100.000						
115	3.224E-16	1.954E-16	100.000						
116	2.913E-16	1.766E-16	100.000						
117	2.812E-16	1.704E-16	100.000						
118	2.592E-16	1.571E-16	100.000						
119	2.177E-16	1.319E-16	100.000						
120	2.051E-16	1.243E-16	100.000						

121	1.700E-16	1.030E-16	100.000						
122	1.490E-16	9.028E-17	100.000						
123	1.394E-16	8.447E-17	100.000						
124	7.942E-17	4.813E-17	100.000						
125	6.339E-17	3.842E-17	100.000						
126	5.650E-17	3.424E-17	100.000						
127	3.730E-17	2.260E-17	100.000						
128	1.643E-17	9.955E-18	100.000						
129	-1.447E-19	-8.769E-20	100.000						
130	-3.492E-18	-2.116E-18	100.000						

131	-1.116E-17	-6.765E-18	100.000						
132	-3.190E-17	-1.933E-17	100.000						
133	-4.705E-17	-2.852E-17	100.000						
134	-6.067E-17	-3.677E-17	100.000						
135	-6.710E-17	-4.067E-17	100.000						
136	-1.011E-16	-6.130E-17	100.000						
137	-1.187E-16	-7.194E-17	100.000						
138	-1.295E-16	-7.851E-17	100.000						
139	-1.448E-16	-8.776E-17	100.000						
140	-1.621E-16	-9.824E-17	100.000						

141	-2.179E-16	-1.321E-16	100.000						
142	-2.668E-16	-1.617E-16	100.000						
143	-2.922E-16	-1.771E-16	100.000						
144	-3.314E-16	-2.008E-16	100.000						
145	-3.796E-16	-2.300E-16	100.000						
146	-3.947E-16	-2.392E-16	100.000						
147	-4.293E-16	-2.602E-16	100.000						
148	-4.389E-16	-2.660E-16	100.000						
149	-4.648E-16	-2.817E-16	100.000						
150	-5.164E-16	-3.130E-16	100.000						

151	-5.507E-16	-3.338E-16	100.000						
152	-5.787E-16	-3.507E-16	100.000						
153	-6.283E-16	-3.808E-16	100.000						
154	-6.913E-16	-4.189E-16	100.000						
155	-7.454E-16	-4.518E-16	100.000						
156	-7.792E-16	-4.723E-16	100.000						
157	-8.088E-16	-4.902E-16	100.000						
158	-8.237E-16	-4.992E-16	100.000						
159	-9.367E-16	-5.677E-16	100.000						
160	-9.911E-16	-6.006E-16	100.000						

161	-1.002E-15	-6.070E-16	100.000						
162	-1.173E-15	-7.110E-16	100.000						
163	-1.331E-15	-8.064E-16	100.000						
164	-1.484E-15	-8.994E-16	100.000						
165	-3.589E-14	-2.175E-14	100.000						

Appendix 7. Radiomics quality scoring from Lambin et al [314]. *A maximum of 36 points is achievable. The higher the score, the higher the indication of the quality of research.*

Criterion	Points	Score of studies in this thesis
Image protocol quality-well documented image protocols (e.g. contrast, slice thickness, energy etc.) and/or usage of public image protocols allow reproducibility/replicability	+1 (if protocols are well-documented) +1 (if public protocol is used)	+1 +1
Multiple segmentations- possible actions are: segmentation by different physicians/algorithms/software, perturbing segmentations by (random) noise, segmentation at different breathing cycles. Analyse feature robustness to segmentation variabilities	+1	0
Phantom study on all scanners-detect inter-scanner differences and vendor-dependent features. Analyse feature robustness to these sources of variability	+1	+1
Imaging at multiple time points-collect individuals' images at additional time points. Analyse feature robustness to temporal variabilities (e.g. organ movement, organ expansion/shrinkage)	+1	0
Feature reduction or adjustment for multiple testing-decreases the risk of overfitting. Overfitting is inevitable if the number of features exceeds the number of samples. Consider robustness when selecting features	-3 (if neither measure is implemented) +3 (if either measure is implemented)	+3
Multivariable analysis with non-radiomic features (e.g. EGFR mutation)-is expected to provide a more holistic model. Permits correlating/inferencing between radiomics and non-radiomics features	+1	+1
Detect and discuss biological correlates-demonstration of phenotypic differences (possibly associated with underlying gene-protein expression patterns) deepens understanding of radiomics and biology	+1	0

Cut-off analyses-determine risk groups by either the median, a previously published cut-off or report a continuous risk variable. Reduces the risk of reporting overly optimistic results	+1	0
Discrimination statistics-report discrimination statistics (e.g. C-statistic, ROC curve, AUC) and their statistical significane (e.g. p-values, confidence intervals). One can also apply resampling method (e.g. bootstrapping, cross-validation)	+1 (if discrimination statistic and its statistical significance are reported) +1 (if also a resampling method technique is applied)	+1
Calibration statistics-report calibration statistics (e.g. Calibration-in-the-large/slope, calibration plots) and their statistical significance (e.g. p-values, confidence intervals). One can also apply resampling method (e.g. bootstrapping, cross-validation)	+1 (if a calibration statistic and its statistical significance are reported) +1 (if also a resampling method technique is applied)	0
Prospective study registered in a trial database-provides the highest level of evidence supporting the clinical validity and usefulness of the radiomics biomarker	+7 (for prospective validation of a radiomics signature in an appropriate trial)	0
Validation-the validation is performed without retraining and without adaptation of the cut-off value, provides crucial information with regards to credible clinical performance	-5 (if validation is missing) +2 (if validation is based on a dataset form the same institute) +3 (if validation is based on a dataset from another institute) +4 (if validation is based on two datasets from two distinct institutes) +4 (if the study validates a previously published study) +5 (if validation is based on three or more datasets from distinct institutes) *Datasets should be of comparable size and should have at least 10 events per model feature.	-5
Comparison to 'gold standard'-assess the extent to which the model agrees with/is superior to the current 'gold standard' method (e.g. TNM-staging for survival prediction). This comparison shows the added value of radiomics	+2	+2

Potential clinical utility-report on the current and potential application of the model in a clinical setting (e.g. decision curve analysis)	+2	+1
Cost-effectiveness analysis-report on the cost-effectiveness of the clinical application (e.g. quality adjusted life years generated)	+1	0
Open science and data-make code and data publicly available. Open science facilitates knowledge transfer and reproducibility of the study	+1 (if scans are open source) +1 (if region of interest segmentations are open source) +1 (if code is open source) +1 (if radiomics features are calculated on a set of representative RPOIs and the calculated features and representative ROIs are open source)	0