UNIVERSITY OF TURKU

A Two-Stage Framework for Identifying Robust Radiomic Biomarkers in Prostate MRI: Evidence from the PRODIF Study

Joint Repeatability and Reliability Analysis at the Lesion Level

Biomedical Imaging/Faculty of Medicine

Master's Thesis

Ahmad Arkan Ichsan

04.03.2026

Turku

Master's thesis

Subject: Biomedical Imaging

Author(s): Ahmad Arkan Ichsan

Title: A Two-Stage Framework for Identifying Robust Radiomic Biomarkers in Prostate MRI: Evidence from the PRODIF Study

Supervisor(s): Harri Merisaari

Number of pages: 100 pages

Date: 04.03.2026

Prostate cancer is the most prevailing non-cutaneous malignant cancer growths among the male population worldwide, which poses obstacles for early detection and clinical decision-making. Despite extensive adoption of multiparametric magnetic resonance imaging (mpMRI), the shortfalls in reliable and repeatable quantitative imaging biomarkers (QIB) persists as a major barrier to the clinical translation of radiomics-based prostate cancer management. Moreover, there is an increasing necessity for minimally invasive prognostic tools which may consolidate diagnostic workflows and assist clinicians in performing triage towards patients requiring either urgent or targeted treatment.

mpMRI has become central to prostate cancer diagnostics; however, variability in the imaging acquisition techniques, reconstruction methods, and quantitative analysis can affect diagnostic performance and reproducibility of mpMRI radiomics as a QIB. This thesis evaluates the classification performance, repeatability and reproducibility of multiple MRI modalities through radiomics-based analysis, with primary focus on diffusion basis spectrum imaging (DBSI) and oscillating gradient spin-echo (OGSE) imaging.

The study analyses diffusion MRI data from the PRODIF cohort study deriving 37 patients with histopathologically confirmed prostate cancer following prostatectomy. Imaging was performed using a 3T Ingenuity PET/MR with 32-channel cardiac coil. Statistical analysis was confined to DBSI and OGSE modalities to assess the repeatability and quantitative robustness of advanced diffusion parameterisations in prostate cancer imaging. Radiomic feature analysis included features from first-order moments, Gabor filters, fast Fourier transform features, edge- and corner-based features (2D/3D) and PyRadiomics-derived lesion metrics were conducted using R to assess reliability and repeatability across acquisitions through a two-stage Intraclass Correlations (ICC) and Area Under Curve (AUC) statistical analytical framework.

Initial results show that when paired with certain MRI modalities, edge- and corner-derived radiomics and PyRadiomics toolkits show potential as repeatable and reliable biomarkers. Secondary stage framework showed only Scharr and Hessian edge- and corner-derived radiomics persisted in both qib-level candidacy analysis in both repeatability, reliability and precision metrics. Through structured repeatability and reliability analysis of modality-specific MRI parameterisations and radiomic features, this thesis identifies structurally stable descriptors with potential translational relevance, while highlighting that broader multi-centre standardisation remains essential for full QIB validation.

Keywords: PCa (PCa), Multiparametric Magnetic Resonance Imaging (mpMRI), Diffusion Basis Spectrum Imaging (DBSI), Oscillating Gradient Spin-Echo (OGSE), Radiomics, Quantitative Imaging Biomarkers (QIB), Reliability, Repeatability, Reproducibility.

Pro gradu -tutkielma

Oppiaine: Biolääketieteellinen kuvantaminen

Tekijä: Ahmad Arkan Ichsan

Työn nimi: Kaksivaiheinen viitekehys radiomiikan biomarkkereiden tunnistamiseen eturauhasen magneettikuvauksessa: Näyttöä PRODIF-tutkimuksesta

Ohjaaja: Harri Merisaari

Sivumäärä: 100 sivua

Päivämäärä: 04.03.2026

Eturauhassyöpä (PCa) on maailmanlaajuisesti yleisin ei-ihoperäinen syöpä miehillä ja muodostaa merkittävän haasteen varhaisdiagnostiikalle ja kliiniselle hoidolle. Vaikka moniparametrinen magneettikuvaus (mpMRI) on vakiinnuttanut asemansa eturauhassyövän diagnostiikassa, luotettavien ja toistettavien kvantitatiivisten kuvantamisbiomarkkereiden (QIB) puute rajoittaa edelleen radiomiikkaan perustuvien menetelmien kliinistä käyttöönottoa. Lisäksi kliinisessä päätöksenteossa on kasvava tarve minimaalisesti invasiivisille prognostisille työkaluille, jotka tehostavat diagnostisia prosesseja ja tukevat kiireellistä tai kohdennettua hoitoa vaativien potilaiden tunnistamista.

mpMRI on keskeinen osa eturauhassyövän diagnostiikkaa. Kuvantamisprotokollien, rekonstruktiomenetelmien ja kvantitatiivisten analyysien välinen vaihtelu voi kuitenkin heikentää diagnostista suorituskykyä sekä radiomiikkapiirteiden toistettavuutta ja uusittavuutta QIB-markkereina. Tässä tutkielmassa arvioidaan useiden magneettikuvausmodaliteettien luokittelukykyä, toistettavuutta ja uusittavuutta radiomiikkaan perustuvan analyysin avulla. Pääpaino on *diffusion basis spectrum imaging* (DBSI) - ja *oscillating gradient spin-echo* (OGSE) -tekniikoissa.

Tutkimus perustuu PRODIF-kohorttiin, joka koostuu 37 potilaasta, joilla eturauhassyöpä varmistettiin histopatologisesti prostatektomian jälkeen. Kuvantaminen suoritettiin 3 teslan Ingenuity PET/MR -laitteella käyttäen 32-kanavaista sydänkelaa. Vaikka aineistoon sisältyi useita kuvantamissekvenssejä, tilastollinen analyysi kohdistettiin DBSI- ja OGSE-modaliteetteihin edistyneiden diffuusioparametrisaatioiden toistettavuuden ja kvantitatiivisen robustisuuden arvioimiseksi eturauhassyövän kuvantamisessa. Biomarkkerien erottamimen toteutettiin R-ohjelmointikielellä. Analyysiin sisältyivät tilastolliset momentit, Gabor-suodattimiin perustuvat piirteet, nopean Fourier-muunnoksen (FFT) deskriptorit, reuna- ja kulmaperusteiset 2D- ja 3D-piirteet sekä PyRadiomics-kirjastolla johdetut leesio-ominaisuudet. Luokittelukykyä arvioitiin pinta-alalla ROC-käyrän alla (AUC), ja piirteiden stabiliteettia arvioitiin intraluokkaisen korrelaatiokertoimen (ICC) avulla.

Tulokset osoittavat, että erityisesti reuna- ja kulmaperusteiset piirteet sekä PyRadiomics-metriikat tuottavat toistettavia ja luotettavia biomarkkereita. Tietyt kuvantamisparametrit paransivat suorituskykyä analysoitaessa ennalta määriteltyjä Gleason-luokituksia vasten. Tutkimus tarkasteli modaliteettikohtaista mpMRI ja radiomiikan biomarkkereita toistettavuus- ja luotettavuusanalyysin kautta, osoittaen valittujen piirteiden korkean mittaustabiliteetin. Tulokset tukevat niiden mahdollista kliinistä hyödyntämistä, samalla korostaen laajemman standardoinnin tarvetta kvantitatiivisessa eturauhasen magneettikuvauksessa.

Avainsanat: Eturauhassyöpä (prostate cancer), moniparametrinen magneettikuvaus (mpMRI), diffusion basis spectrum imaging (DBSI), oscillating gradient spin-echo (OGSE), radiomiikka, kvantitatiiviset kuvantamisbiomarkkerit (QIB), luotettavuus, toistettavuus.

List of Abbreviations

ADC – Apparent Diffusion Coefficient

AD – Axial Diffusivity

AI – Artificial Intelligence

AUC – Area Under the Curve

BG - Background

BPH – Benign Prostatic Hyperplasia

CI – Confidence Interval

CNR – Contrast-to-Noise Ratio

csPCa – Clinically Significant Prostate Cancer

CV – Coefficient of Variation

DBSI – Diffusion Basis Spectrum Imaging

DCE – Dynamic Contrast-Enhanced Imaging

DRE – Digital Rectal Examination

DWI – Diffusion-Weighted Imaging

EC2D3D – Edges- and Corners-Based Two- and Three-Dimensional Features

FFT2D – Two-Dimensional Fast Fourier Transform

FWHM – Full Width at Half Maximum

GGG – Gleason Grade Group

GLCM – Gray-Level Co-occurrence Matrix

GLDM – Gray-Level Dependence Matrix

GLRLM – Gray-Level Run Length Matrix

GLSZM – Gray-Level Size Zone Matrix

HR-DWI – High-Resolution Diffusion-Weighted Imaging

IBSI – Image Biomarker Standardisation Initiative

ICC – Intraclass Correlation Coefficient

ISUP – International Society of Urological Pathology

IQR – Interquartile Range

LoG – Laplacian of Gaussian

mpMRI – Multiparametric Magnetic Resonance Imaging

MRCRadiomics – MRI Research Center Radiomics

MRI – Magnetic Resonance Imaging

NGTDM – Neighbouring Gray-Tone Difference Matrix

OGSE – Oscillating Gradient Spin-Echo

PCa – Prostate Cancer

PI-RADS – Prostate Imaging Reporting and Data System

PRODIF – Prostate Diffusion Imaging Study (test–retest cohort)

PSA – Prostate-Specific Antigen

QIB – Quantitative Imaging Biomarker

RD – Radial Diffusivity

ROC – Receiver Operating Characteristic

ROI – Region of Interest

SD – Standard Deviation

SEM – Standard Error of Measurement

SPAIR – Spectral Attenuated Inversion Recovery

T1w – T1-weighted Imaging

T2w – T2-weighted Imaging

TE – Echo Time

TR – Repetition Time

TRUS – Transrectal Ultrasound Sonography

Table of contents

1 Literature Overview

1.1 Global Clinical Context of Prostate Caner

1.1.1 Epidemiology and Disease Burden

Prostate cancer (PCa) is the most commonly diagnosed non-cutaneous malignant growths among the male population in many parts of the globe. It remains a leading cause of cancer-related mortalities, in 2020 alone, an approximate amount of 1.4 million men were recently diagnosed and over 375,000 deaths were reported to be caused by PCaglobally.[1–5] Current demographic forecasts suggest that by the year 2040, the global burden of PCa will escalate to relatively 2.4 million cases and 712,000 deaths, largely driven by aging and growing global population.[6–8] Oncogenetic predisposition, scanning protocols, environmental exposures, and healthcare infrastructure are identified as causes for a wide difference of incidence and mortality and PCa across geographic regions.[2,5,9]

In spite of its increasing incidence, heterogeneity is a trait commonly encountered in PCa pathogenesis. Tumour indolence within prolonged periods is a trait observable in a sizable amount of localised tumours, the prevalence of aggressive cancer phenotypes may result in infrastructure burden caused by early metastatic potential and resistance to therapy.[2,3,8] Thus, the principal clinical challenge to innovate beyond means of detection, whereupon a necessity to reliably discriminate clinically significant prostate cancer (csPCa) from indolence.[10,11]

Cornford et al[11] underlines in the European Association of Urology that prostate-specific antigen (PSA) testing, digital rectal examination (DRE), and systematic transrectal ultrasound-guided biopsy are among the methods in traditional diagnostic pathways for PCa identification. However, it was emphasized that PSA lacks disease specificity to elimitate differential diagnosis,[11–22] and systematic biopsy is inherently sampling-dependent.[11,15,23–26] Tombal et al[27] and Gulati et al[28] argues that the underdiagnosis of clinically significant tumours lead to inaccuracy of care, while Dushimova et al[29] further underpins that the overdiagnosis of non-significant PCa would lead to the overtreatment of clinically insignificant cases.

These technical limitations underscore the need for more reliable and less invasive risk stratification strategies, in this context, multiparametric magnetic resonance imaging has emerged as a promising tool for improving lesion localisation and risk stratification.

1.1.2 Clinical Heterogeneity and the Need for Risk Stratification

Multiparametric magnetic resonance imaging (mpMRI) has emerged as a principal component of PCa assessment.[11,25,26,30–33] The integrative approach of mpMRI enables the illustration of tumour-associated structural distortion and diffusion restriction in PCa.[32–35]

Various randomised trials have demonstrated that MRI-guided biopsy detects a higher proportion of clinically significant cancers compared with systematic ultrasound-guided approaches, while reducing detection of indolent disease.[25,26,36–38]

Cross-institutional interobserver consistency is a key factor undertook by the prostate Imaging Reporting and Data System (PI-RADS), which helps contain inter-observer variability during PCa imaging.[11,39–43]

Advances in acquisition and reconstruction protocols have also facilitated quantitative mapping of tissue characteristics. For example, diffusion-weighted based imaging modalities catechizes tissue microstructure by mapping water mobility and gauging water restriction commonly noted in malignant transformation.[44–46] These developments place mpMRI as a potential touchstone for PCa quantitative biomarker development.[47–50]

1.2 Prostatic Anatomy and Zonal Architecture

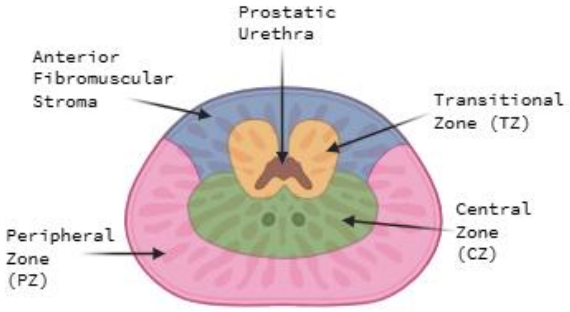
1.2.1 Zonal Anatomy and Imaging Correlates

The prostate is an exocrine gland located inferior to the urinary bladder, anterior to the rectum, and posterior to the pubic symphysis.[51–53] This anatomical location permits clinical assessment through digital rectal examination and facilitates high-resolution pelvic MRI.[11,19,26,53] Posterior lesions, particularly within the peripheral zone, are generally more accessible to both biopsy and imaging evaluation than anteriorly situated tumours.[41,54]

In adult males, the prostate typically weighs 20–25 g and measures approximately 4 × 3 × 2 cm.[51–54] Although historically described using a lobar classification, contemporary clinical and imaging frameworks emphasise zonal anatomy, which more accurately reflects patterns of tumour origin and corresponding imaging characteristics.[40,54]

From an imaging perspective, the zonal division of the prostate constitutes the most diagnostically meaningful anatomical framework, illustrated in Figure 1.1.

Figure 1.1 Prostate Zone Anatomy



- Peripheral Zone (PZ): Accounting for approximately 70% of glandular volume, the PZ is located posteriorly and laterally beneath the prostatic capsule.[54] It represents the site of origin of 70–80% of PCa,[11,30,53] and therefore constitutes the primary focus of diffusion-weighted assessment within the PI-RADS framework.[39–42] The PZ is characterised by abundant luminal spaces and relatively low stromal density, histological features that subscribe to its homogeneous hyperintensity depiction on T2-weighted imaging protocols.[43,55,56]

- Transition Zone (TZ): Surrounding the proximal urethra,[51–53] the TZ enlarges with age and is commonly associated with benign prostatic hyperplasia.[53,57] Approximately 10–20% of PCa arise in this zone,[11] where tumour detection is complicated by overlapping stromal and hyperplastic changes that reduce lesion conspicuity in MRI.[11,41–43,58,59]

- Central Zone (CZ): Comprising approximately 20% of the prostate, the CZ surrounds the ejaculatory ducts.[43] Cornford et al[11] and McNeal et al[54] highlighted that some aggressive pathological features such as early seminal vesicle invasion are associated in tumours originating in prostatic CZ.

- Anterior Fibromuscular Stroma: The non-glandular region of the prostate which consist fibrous and muscular tissue, providing structural support to the prostate. Malignancies originating from this region are reported to be rare.[51–54]

It is therefore imperative to distinguish prostatic zonal anatomy to differentiate baseline signal characteristics across regions. These signals may influence lesion conspicuity and overall interpretive reliability.[34,40,43,55,56,60–62]

1.2.2 Histological Basis of MRI Signal Characteristics

Presentation of PZ portrays homogenous hyperintensity under T2-weighted imaging protocols due to its abundant luminal spaces and relatively low stromal intensity.[43,54–56]

Lee et al[63] and Cheng et al[64] emphasizes that upon oncogenic progression, the prostatic structure is disrupted which causes increased cellular density and reduced luminal volume. These perturbations are further illustrated by Tamada et al[46] and Hricak et al[56] where under T2-weighted imaging, there is reduced signal intensity and diffusion restriction on Diffusion Weighted Imaging (DWI), corresponding to reductions in apparent diffusion coefficient (ADC) values.

This histo-anatomical relationship establishes the biological rationale for the utilization of diffusion-based weighted imaging in PCa diagnostics. As neoplastic progression amplifies cellular density and curtails extracellular space, the manoeuvrability of water molecules becomes successively dampened.[44,47,63,64] These microstructural alterations elucidates straight into quantifiable imaging contrast, acquiescing diffusion metrics such as ADC to serve as quantitative surrogates of tissue architecture.[44,46,47,56]

From a radiomic perspective, zonal heterogeneity introduces both fortuitous opportunities and methodological complexity.[65,66] Touchstone signal behaviour differs substantially between the PZ and the TZ of the prostate,[43,55] and these underlying characteristics may influence feature extraction and stability.[33,67–69] As a result, quantitative imaging biomarkers must manifest robustness as a *sine qua non* across anatomically specific tissue micro-environments to ensure dependable repeatability, and clinical reliability.[47,48,50]

1.3 PCa Biology and Risk Stratification

1.3.1 Tumour Progression and Clinical Staging

PCa most commonly presents as an adenocarcinoma rooted from the glandular epithelium of the prostate.[63,70] Since majority of glandular tissue is vested within the PZ,[54] consequently approximately 70-75% of tumours originate in this particular region.[30,60]

Early-phase disease is intermittently asymptomatic and typically detected incidentally through screening or assessment of elevated PSA levels.[30,70] A wide array of prostate tumours demonstrate indolent growth with negligible clinical impact, particularly in aging populations.[2,30] Nevertheless, a subset displays aggressive biological malignancy and develops to locally advanced or metastatic disease.[3,70] Late-stage malignancy may involve extracapsular extension, seminal vesicle invasion and dissemination to regional lymph nodes and bone.[64,70] Alongside this, skeletal metastases, which are characteristic of advanced-stage PCa, would predominantly cause bone pain, fatigue, weight loss, and in cases of spinal involvement, neurological compromise due to spinal cord compression.[70]

Clinical staging defines the anatomical extent of disease, from organ-confined tumours to locally advanced and metastatic states.[64] While stage guides treatment selection and overall risk stratification,[11,30,70] histological grade, particularly Gleason score and Grade Group, has emerged as the dominant predictor of biological aggressiveness and prognosis in localized PCa.[3,10,30,70]

1.3.2 Gleason Grading and Grade Group System

The Gleason grading system remains the cornerstone of pathological risk stratification in PCa.[25,30 63,64] underscore that the Greason grading system evaluates glandular architectural patterns in biopsy specimens, the system then assigns a composite score based on the sum of the primary and secondary tumour growth patterns. In clinical practice, higher biopsy-assigned Gleason reflect progressive cellular disorganisation.[3,30]

Further improvements to the grading systems are made by the International Society of Urological Pathology (ISUP), where they introduced the Grade Group system in 2014,[11,71,72] consolidating Gleason scores into five distinct prognostic categories:

- **Grade Group 1 (Gleason 3+3=6):** Well-differentiated, low-risk disease
- **Grade Group 2 (3+4=7):** Predominantly well-formed glands with a minor high-grade component
- **Grade Group 3 (4+3=7):** Predominantly high-grade architecture
- **Grade Group 4 (Gleason 8):** Poorly differentiated, high-risk disease
- **Grade Group 5 (Gleason 9–10):** Very high-risk carcinoma with substantial metastatic potential.

Increased risk of biochemical recurrence, metastatic progression and cancer progression is reflected in heightened ISUP Grade Group values.[11,72,73]

Histologically, advanced tumour grades impact cellular disorientation. These structural alterations directly influence magnetic resonance signal characteristics, contributing to T2 hypointensity and diffusion restriction observed on diffusion-weighted imaging.[41,46] Reliable differentiation between insignificant PCa and csPCa through grade group discrimination represents a crucial objective of quantitative prostate MRI research.[34,40,42]

1.3.3 Prostate -Specific Antigen and Diagnostic Limitations

Prostate-specific antigen (PSA) testing remains the most widely used screening tool for PCa diagnostics.[12,14] Lee et al[63] reports that PSA is produced by both benign and malignant prostatic epithelial cells, however its expression is not explicitly cancer related. These conclusions were emphasized by, Catalona et al[13] and Ilic et al[18] which underline that elevated PSA is seen in benign prostatic hyperplasia (BPH), prostatitis, recent ejaculation, or instrumentation. Meanwhile, Sarkar et al[74] reported that reduced PSA concentrations is observable upon administration of 5-alpha reductase inhibitors such as finasteride or dutasteride.

To prevent this, the inclusion of adjunctive biomarkers such as: free-to-total PSA ratio,[75] PSA density,[76] PSA velocity,[77] the Prostate Health Index,[78,79] and kallikrein-based panels,[79,80] aims to improve statistical discrimination in selected PCa populations.[17,21] Large randomized trials have shown that PSA screening increases cancer detection rates.[12,16]

Nonetheless, PSA approaches remain probabilistic and do not provide definitive diagnostic classification. It has been argued that PSA alone does not reliably distinguish indolent from clinically significant disease.[11,17,18] Biological overlap between benign and malignant conditions persists,[18,21,81,82] and histopathological confirmation remains the diagnostic standard.[11,14,23,83] This has led to dual concerns of PSA usage: overtreatment of low-risk tumours and incomplete detection of aggressive phenotypes.[17,20,26]

1.3.4 Genetic and Environmental Risk Factors

PCa aetiology is based on multiple factors relating to age, inherited susceptibility, and environmental exposure.[2,3,70]

Age is the main risk determinant of PCa, with incidence rising sharply after the sixth decade of life.[1,2] Post-mortem studies portrays a substantial proportion of elderly men harbour histologically detectable PCa, many of which remain clinically silent and biologically indolent.[30,70,84]

In multiple population-based analysis, men of African ancestry are reported to have substantially higher incidence rates and PCa-specific mortality compared with men of European ancestry.[1–3,7] Disparities in socioeconomic levels become a determinant in analysing outcome differences in between healthcare systems, which suggests inequities in PCa diagnostic capacity across nations.[2,3,70] Furthermore, a positive family history confers an approximately two- to three-fold increased risk of PCa among first-degree relatives.[2,9,30]

Germline mutations in DNA repair pathway genes, particularly BRCA2 and HOXB13, have been associated with earlier onset disease, higher grade tumours and increased risk of metastatic progression.[30,31,85–87] Beyond high-penetrance mutations, genome-wide association studies have identified numerous common susceptibility loci,

supporting a polygenic model of inherited risk in which multiple-low effect variants collectively influence disease probability.[2,3]

Metabolic and inflammatory conditions, including obesity and metabolic syndrome, have been associated with increased risk of higher-grade and more advanced PCa in several epidemiological analyses.[2,5,88] Proposed mechanisms include chronic low-grade inflammation, altered hormonal signalling, and metabolic dysregulation, although the magnitude and independence of these associations vary across cohorts and study designs.[30,88–90]

1.4 Contemporary Diagnostic Pathways

1.4.1 Biopsy Techniques and Limitations

Biopsy has persisted as gold standard for conclusive PCa diagnosis and remains the reference standard.[11,23] Systematic transrectal ultrasound-guided (TRUS) biopsy samples predefined regions of the prostate in a grid-based distribution.[91,92] Although widely implemented, this clinical strategy is limited by sampling variability and tumour heterogeneity. Reported false-negative rates range from approximately 15% to 46%.[91–93] In addition, pathological under grading relative to radical prostatectomy specimens occurs in a substantial proportion of cases.[92,94]

A technical fallacy in TRUS biopsy is primarily due to needle placement in systematic biopsy not directed toward radiologically visible lesions, causing clinically significant tumours may potentially remain unsampled, whereas small-volume insignificant presentations are incidentally detected.[47,91,93]

This diagnostic asymmetry contributes to both under recognition of aggressive disease and potential overtreatment of biologically low-risk tumours.[26,36,93,94] Furthermore, prostate biopsy is an invasive intervention associated with complications including infection, bleeding, pain and transient urinary dysfunction.[74,95] Recognition of these structural limitations has led to progressive integration of imaging-guided biopsy strategies, particularly MRI-targeted approaches, within contemporary diagnostic algorithms.[25,26,31,36–38]

1.5 Multiparametric Magnetic Resonance Imaging (mpMRI) and Parameter Modalities

Initially adopted for local staging, Multiparametric magnetic resonance imaging (mpMRI) sequences delineates distinct biological properties of the prostatic tissue as a functional and microstructural assessment.[26,40] By integrating T2-weighted imaging with functional techniques such DWI and dynamic contrast-enhanced imaging (DCE) sequences, mpMRI provides both anatomical and microstructural information that improves lesion detection and localisation.[40,41,43] Furthermore, this approach enables MRi guided biopsy of regions of interests, increasing detection of clinically significant presentations while reducing indolent tumor detection.[26,37,47] Altogether, oncologic application of mpMRI underscoring its role in PCa management pathways from detection, risk stratification, biopsy guidance and treatment planning.[11,32,35,96–98]

Standard protocols combine high resolution T2-weighted imaging with diffusion-weighted imaging (DWI), and dynamic contrast-enhanced imaging (DCE).[42] More recently, advanced diffusion techniques such as Oscillating-Gradient Spin-Echo (OGSE) and Diffusion Basis Spectrum Imaging (DBSI) have been introduced to interrogate tissue microstructure beyond conventional ADC mapping.[42,46] These approaches aim to increase sensitivity to cellular density, extracellular restriction, and complex diffusion environments. Since these modalities vary in biological sensitivity and technical variability, each warrants individual exploration within the mpMRI framework to further expound how diagnostic performance and measurement reproducibility interacts.[48,99–105]

1.5.1 T1-weighted

T1-weighted imaging serves a complementary role in prostate MRI. Its primary utility lies in detecting post-biopsy haemorrhage, which can mimic or obscure tumour signal on other sequences, and in evaluating extra prostatic findings such as lymph node enlargement or osseous metastases. Intraprostatic tumour foci themselves typically do not demonstrate distinctive T1 signal characteristics. Consequently, while T1-weighted imaging provides important contextual and staging information, it offers limited value for intraprostatic lesion detection or quantitative characterisation.[56,104,106,107]

1.5.2 T2-weighted Imaging (T2w)

T2-weighted imaging (T2w) provides high-resolution visualization of prostatic zonal anatomy and remains the anatomical backbone of MRI interpretation.[55,108] Its intrinsic soft-tissue contrast enables reliable differentiation between the PZ and TZ, facilitating lesion localisation and local staging assessment.

T2w shows hyperintensity when encountered with abundant luminal fluids within the well-formed glandular acini typical in PZ. Malignancy discombobulates the architecture through glandular crowding, reduced luminal volume and stromal invasion, these are shown as hypointensive presentations in T2.[60,63] For these reasons, the difference in imaging appearance directly reflects underlying histoarchitectural alteration. Unlike the PZ, TZ is more challenging to interpret on T2w. BPH presents nodular enlargement with variable stromal and glandular composition, which results in heterogenous T2 signal intensity and architectural distortion that mimic PCa.[62,109] Age-related volumetric expansion of the TZ further alters baseline signal characteristics and glandular morphology, increasing overlap between benign and malignant appearances.[61]

Furthermore, while T2w imaging reliably delineates capsular contour, neurovascular bundle involvement, and features reminiscent of extracapsular extension, its imaging specificity is constrained by signal overlap with benign hyperplasia, prostatitis, post-biopsy haemorrhage, and therapy-related changes. Posing challenges in differentiating between competing diagnostic possibilities and contributing to persistent diagnostic uncertainty.[58] Thus, although T2w imaging remains indispensable for anatomical mapping and local staging, accurate interzonal interpretation requires contextual integration with diffusion-weighted and other functional modality sequences.

1.5.3 Diffusion-Weighted Imaging (DWI)

Diffusion-weighted imaging (DWI) modality evaluates water molecule mobility, which is a concerted effort to map tissue microstructure, notably cellular density and extracellular composition.[44,46] In DWI-based parameterizations, prostatic malignancy is illustrated with cellular packing, reduced luminal volume, and stromal remodelling, all of which, restrict extracellular water diffusion. ADC maps would display decreased

values relative to benign PZ tissue.[63,110] Numerous clinical investigations have demonstrated inverse associations between ADC values in DWI and Gleason grade, supporting the concept that diffusion metrics reflect tumour aggressiveness and may function as quantitative biomarkers.[42,111,112]

The diagnostic performance and quantitative stability of DWI parameters are influenced by b-value selection. Higher b-values suppress background signal and enhance sensitivity to restricted diffusion, thereby improving lesion conspicuity, particularly in the PZ.[34,113] However, diffusion measurements are inherently sensitive to technical factors including echo time, gradient strength, signal-to-noise ratio, motion, distortion, and reconstruction methodology.[47] As a result, absolute ADC values are not directly interchangeable across scanners or acquisition protocols without harmonisation.[50,68,114]

1.5.4 High-Resolution Diffusion-Weighted Imaging (HR-DWI)

High-resolution diffusion-weighted imaging (HR-DWI) extends conventional DWI by increasing spatial resolution and reducing geometric distortion, with the aim of improving anatomical detail within the prostate. By reducing voxel size and utilising acquisition strategies such as reduced field-of-view or multi-shot techniques, HR-DWI enhances lesion margin definition and may improve visualization of subtle tumour features that are less apparent on standard protocols.[115,116]

In a comparative protocol study, higher spatial resolution HR-DWI demonstrated improved subjective image quality and internal architecture definition relative to conventional DWI, though at the expense of lower signal-to-noise ratio, a trade-off that can be mitigated by optimisation of acquisition parameters.[115] Similarly, in a prospective clinical cohort, HR-DWI outperformed standard DWI in predicting adverse biopsy histology based on Gleason scores by better detecting early changes in tumour size and ADC, suggesting potential sensitivity advantages in active surveillance contexts.[116]

Advanced HR-DWI techniques, including multiplexed sensitivity encoding (MUSE) methods, have shown further improvements in overall image quality and distortion reduction compared with both conventional and reduced field-of-view approaches, while maintaining comparable ADC quantification.[117]

However, increased spatial resolution introduces trade-offs. Smaller voxel sizes reduce signal-to-noise ratio, and multi-shot approaches are more susceptible to motion-related phase errors. Diffusion measurements remain sensitive to acquisition parameters, gradient performance, and reconstruction methods.[44,47]

1.5.5 Oscillating-Gradient Spin-Echo (OGSE)

Oscillating-gradient spin-echo (OGSE) imaging expounds upon conventional pulsed-gradient diffusion techniques by replacing monopolar gradient pulses with oscillatory waveforms.[118,119] Jones et al[44] and Does et al[120] underpins that the OGSE modality design aims to probe diffusion at shorter effective diffusion times, allowing water motion to be examined over smaller spatial scales than those captured by standard pulsed-gradient DWI. Because diffusion time determines the structural scale over which molecular displacement is measured, OGSE introduces sensitivity to cellular packing density, membrane restrictions, and subcellular organisation that may not be fully characterized at longer diffusion times.[118,121,122] Tamada et al[46] identifies that within PCa presentations, diffusion-time dependence provides a biologically plausible framework for detecting architectural disruption beyond conventional ADC maps.

However, the theoretical advantages of OGSE are not without its technical shortfalls. Quantitative measurements derived on gradient frequency, waveform configuration, lobe number, and scanner hardware performance, and achievable b-values along with signal-to-noise characteristics may vary with frequency selection. These imaging parameter configurations are not directly interchangeable across imaging platforms, introducing acquisition variability and complicating cross-centre standardization.[47,69]

From the perspective of quantitative imaging biomarker development, the central issue is not solely whether OGSE enhances microstructural sensitivity, but whether this sensitivity can be achieved with sufficient repeatability and cross-parameter stability. Diffusion MRI measurements are inherently dependent on acquisition design, including gradient timing, b-value selection, and waveform configuration, and these factors materially influence quantitative robustness.[44,123] Frequency-dependent diffusion measurements further introduce sensitivity to oscillation frequency and gradient

hardware performance, creating additional sources of variability that are not present in conventional pulsed-gradient acquisitions.[120,122]

Importantly, studies of diffusion model fitting and parameter optimisation have demonstrated that quantitative outputs may vary substantially depending on modelling strategy and acquisition scheme, even when biological tissue remains unchanged.[47,124] More recent harmonisation work in MRI radiomics has likewise shown that scanner-dependent and protocol-dependent variability can meaningfully alter quantitative feature behaviour.[68,69] Consequently, optimisation of diffusion frequency and rigorous reproducibility assessment remain essential prerequisites for translating OGSE-derived metrics into clinically reliable quantitative imaging biomarkers.

1.5.6 Diffusion Basis Spectrum Imaging (DBSI)

Diffusion Basis Spectrum Imaging (DBSI) represents a model-based extension of diffusion MRI that decomposes the measured diffusion signal into multiple anisotropic and isotropic components within each voxel.[100–102] In contrast to conventional ADC mapping, which assumes a single effective diffusion coefficient, DBSI estimates distinct restricted, hindered and free diffusion fractions, thereby attempting to separate coexisting microstructural environments.

This multi-compartment modelling framework offers theoretical advantages in tissues with complex architecture. By differentiating restricted cellular components from extracellular or luminal water, DBSI may increase biological specificity and reduce ambiguity arising from overlapping diffusion signatures. Quantitative maps such as restricted ratio (RR), radial diffusivity (RD), water ratio (WR), and axial diffusivity (AD) along with their respective first diffusion (F1) compartment types have therefore been proposed as surrogate markers of tissue composition and pathological burden.[100,102]

However, the strengths of model-based diffusion approaches are accompanied by methodological demands. Parameter estimation depends on acquisition design, b-value distribution, angular sampling, and signal-to-noise ratio.[44] The increased dimensionality of the fitting procedure introduces greater sensitivity to noise and modelling assumptions, and instability may propagate across derived parameters if

constraints are not carefully implemented.[47,124] As with other advanced diffusion techniques, reproducibility and cross-protocol robustness must be demonstrated before DBSI-derived metrics can be considered reliable quantitative imaging biomarkers.[47]

1.5.7 Spectral Attenuated Inversion Recovery (SPAIR)

Spectral Attenuated Inversion Recovery (SPAIR) is a frequency-selective fat-suppression technique commonly used in prostate MRI to reduce signal from periprostatic adipose tissue and improve lesion conspicuity.[41,108] SPAIR is a hybrid technique combining CHESS and STIR which provides effective fat suppression enhances anatomical boundary definition and reduces confounding signal from surrounding adipose tissue, thereby supporting more reliable visual interpretation.[125] Compared with basic fat-saturation MRI modalities, inversion-recovery methods such as SPAIR provide more homogeneous suppression, although their performance remains influenced by magnetic field non-uniformity and RF pulse characteristics.[126–128]

The performance of spectral fat suppression techniques depends on magnetic field homogeneity and sequence calibration. Imperfect field uniformity was designated as a cause of regional suppression failure or signal heterogeneity, which may influence perceived lesion margins and contrast.[43,104] Practical fat suppression by SPAIR amplifies the illustration of anatomical boundary and reduces confounding signal from surrounding fat, thereby promoting visual interpretation accuracy.[47,105,126,128]

1.5.8 Strengths and Limitations of mpMRI

mpMRI has considerably improve detection of csPCa and reduce unnecessary biopsies combined with systematic sampling.[25,26,98] When interpreted using structured reporting pathways such as PI-RADS, mpMRI exhibits high pooled sensitivity and moderate specificity for clinically significant disease.[39,42,43,106]

In spite of these advances, mpMRI interpretation remains partly qualitative. Rosenkrantz & Taneja[58] and Purysko et al[129] argues that there is imbalanced subjectivity pertaining to visual assessment of signal intensity patterns, lesion morphology,

dominant sequence weighting, which continues to rely on reader expertise and experience.

Furthermore, the multi-centre variabilities in acquisition protocols, different field strength, gradient performance, reconstruction methods, and post-processing pipelines may impinge quantitative measurements and diagnostic performance across institutions.[47,68,98] In addition, Hajianfar et al[69] and Merisaari et al[123] emphasized that diffusion-derived metrics in particular exhibit parameter dependence and inter-scanner variability, these perimeter may limit direct comparability without prior harmonisation.

These restraints underpin a principal methodological problem: reliable microstructural sensitivity must be counterbalanced against reproducibility and measurement repeatability. Sullivan et al[48] and Traverso et al[105] identifies that there is a critical need for rigorous quantitative evaluation frameworks to be developed in parallel with sophisticated MRI parameterisations. This tension between biological sensitivity and quantitative robustness provides the conceptual foundation for the present thesis.

1.6 Prostate Imaging Reporting and Data System (PI-RADS)

The Prostate Imaging Reporting and Data System (PI-RADS) was developed to standardise acquisition, interpretation, and reporting of multiparametric prostate MRI.[43] By providing structured criteria for lesion categorisation, PI-RADS aims to reduce inter-reader variability and improve diagnostic consistency cancer across institutions.[34,129,130]

When applied alongside Prostate Imaging Reporting and Data System (PI-RADS) interpretation criteria, mpMRI demonstrates pooled sensitivity of approximately 85-90% and specificity near 70-75% for detection of PCa.[42,43,131]

Version 2, published in 2015, expands the framework by applying a structured five-point scoring system reflecting the likelihood of clinically significant disease and established zonal dominance rules, whereby DWI is primary in the peripheral zone and T2-weighted imaging in the TZ.[34,43] Dynamic contrast-enhanced imaging (DCE) was assigned a secondary role, primarily used to upgrade equivocal PZ lesions (PI-RADS 3 to 4). Although PI-RADS v2 improved standardisation and diagnostic performance,

interobserver variability and ambiguities in transition zone interpretation remained recognised limitations.[42,43]

To address these concerns, PI-RADS version 2.1 (v2.1) was released in 2019 by the American College of Radiology (ACR).[39] The amendment did not fundamentally recalibrate the scoring framework but imported clarifications aimed at improving accuracy and reproducibility.[40,131]

Key modifications in PI-RADS v2.1[39,40,131] included:

- Refinement of TZ lesion morphology characterization, aiming to improve differentiation between typical BPH nodules and more suspicious lesions.
- Clarification of DWI scoring criteria, including emphasis on high b-value imaging.
- Technical acquisition recommendations to enhance standardisation.
- Minor adjustments to reduce reader ambiguity in peripheral zone assessment.

Published literature have highlighted that PI-RADS v2.1 predominantly driven to improve inter-reader agreement in imaging interpretation.[129,131,132] Diagnostic utility remained largely tantamount to v2, but measured refinements were designed to reduce variability and enhance consistency in reporting.[132,133] Regardless of these refinements, mpMRI interpretation remains partly qualitative and dependent on reader expertise.[58,134] The foundations of a PI-RADS value judgement lies on visual evaluation of signal intensity patterns and morphological features rather than quantitative thresholds.[39,42,43] Diffusion-weighted imaging is the dominant sequence in PZ, whereas T2-weighted morphological assessment predominates in TZ.[39,132]

Inter-reader variability remains a PI-RADS limitation, principally in TZ interpretation and in equivocal lesions.[42,58,59] In addition, heterogeneity in acquisition protocols, scanner hardware, and reconstruction parameters can potentially influence image appearance and downstream diagnostic performance.[47] These corresponding volatilities accentuates the need for QIBs that could provide reproducible, standardised metrics capable of distinguishing clinically significant from indolent disease across acquisition environments.[48,50]

1.6.1 Zonal Dominance Rules

Lesions are assigned particular PI-RADS categories ranging from 1 to 5, reflecting the probability of csPCa.[39,43] Assessment integrates T2w, DWI, and DCE within a structured scoring algorithm. Sequence dominance is determined by zonal anatomy: DWI serves as the dominant sequence in the PZ, whereas T2w imaging predominates in the TZ.[43,131]

Table 1.1 summarises lesion assessment and classification according to PI-RADS version 2.1. Adapted from ACR[39] :

Table 1.1 PI-RADS V2.1 Classification

PI-RADS Category	Peripheral Zone (PZ) – Dominant Sequence: DWI/ADC	Transition Zone (TZ) – Dominant Sequence: T2-Weighted Imaging
1	Normal appearance on ADC and DWI.	Normal-appearing transition zone or completely encapsulated round nodule consistent with benign prostatic hyperplasia (BPH).
2	Linear or wedge-shaped hypointensity on ADC with corresponding hyperintensity on high b-value DWI; no focal mass-like abnormality.	Mostly encapsulated nodule or homogeneous circumscribed nodule without a capsule; or homogeneous mildly hypointense area between nodules consistent with BPH.
3	Focal mild-to-moderate hypointensity on ADC and/or focal mild-to-moderate hyperintensity on high b-value DWI. No definite extraprostatic extension. DCE does not demonstrate focal early enhancement*.	Heterogeneous signal intensity with obscured margins not meeting criteria for categories 2, 4, or 5.
4	Focal marked hypointensity on ADC with corresponding marked hyperintensity on high b-value DWI, <1.5 cm	Lenticular or non-circumscribed homogeneous moderately hypointense lesion <1.5 cm**.
5	Same imaging characteristics as category 4, but lesion ≥ 1.5 cm in greatest dimension and/or with definite extraprostatic extension.	Same imaging characteristics as category 4, but lesion ≥ 1.5 cm in greatest dimension and/or with definite extraprostatic extension.

*PZ category 3 is upgraded to category 4 if DCE is positive.[39,40]

**If TZ category 3 with DWI score 5, then upgraded to category 4 or 5 depending on size.[129,132]

In PZ, DWI and corresponding ADC maps serve as the dominant determinants of lesion suspicion within the PI-RADS framework.[39,43] Marked diffusion restriction, reflected as low ADC values with corresponding high signal intensity on high b-value DWI, increases the likelihood of csPCa.

In the transition zone, T2-weighted morphology assumes greater importance. Lesions are assessed according to margin definition, signal homogeneity, and architectural distortion, as the heterogeneous background signal associated with benign prostatic hyperplasia can reduce the specificity of diffusion findings [129,131].

1.6.2 Diagnostic Performance and Limitations of Qualitative Interpretation

Prior meta-analyses and systematic reviews have consistently demonstrated high pooled sensitivity and moderate specificity of PI-RADS-guided mpMRI for detection of csPCa.[42,106,135,136] These findings support its integration into biopsy decision-making pathways and active surveillance strategies.[25,36,137]

However, PI-RADS does not alleviate inherent qualitative concerns that lead to variability. It was further argued that lesion categorisation relies on visual evaluation of signal intensity patterns, morphological characteristics, and enhancement behaviour within a structured interpretative framework that risks inter-reader variability.[43,129,138] Merisaari et al[123] and Shukla-Dave et al[47] further affirm that although ADC values may support PI-RADS interpretability, absolute quantitative thresholds are yet to be standardised across scanners or acquisition protocols.

Moreover, further concerns were raised regarding variability in acquisition parameters, image quality, and reader expertise continues to influence diagnostic performance.[58,68,130] Thus, while PI-RADS represents a substantial advance in structured reporting and harmonised interpretation, it does not eliminate interpretative variability nor fully leverage the quantitative information embedded within diffusion imaging.

For these reasons, these limitations emphasize a clear rationale for investigating structured quantitative feature extraction approaches that aim to improve reproducibility, stability, and cross-institutional robustness.

1.7 Radiomics in Prostate MRI

In medical research, radiomics is defined as a method that extracts large-data features using data-characterisation algorithms,[66] transforming medical images into high-dimensional quantitative data by systematically extracting mathematical descriptors

from defined regions of interest.[65,66] Radiomic analysis seeks to quantitatively characterise tissue phenotype through numerical features derived from voxel intensity distributions, spatial relationships, textural patterns, and geometric properties.[67]

The radiomic workflow typically comprises standardised image acquisition, region-of-interest delineation, feature extraction, and statistical or machine learning–based modelling.[33,50,139] From a single segmentation, hundreds to thousands of features may be computed, generating a multidimensional representation of imaging phenotype that extends beyond visually appreciable contrast differences.

Radiomic features differ substantially in mathematical construction, level of abstraction from the raw signal, and susceptibility to acquisition variability.[67,105] Zwanenburg et al[50] and Gillies et al[66] states that first-order statistics directly describe intensity distributions, whereas higher-order textural features quantify spatial dependencies and structural heterogeneity. These distinctions are critical when evaluating QIB candidates, as feature complexity may increase biological sensitivity while amplifying instability which leads to reliability and reproducibility concerns.[47,48,67,105]

1.7.1 Intensity-Based and Texture-Based Features

Intensity-based features quantify voxel signal distributions within a region of interest without incorporating spatial context. Let I_i denote voxel intensities within a segmented region containing N voxels. First-order statistics are computed directly from this distribution [50,140]:

$$\text{Mean } (\mu) = \frac{1}{N}\sum_{i=1}^{N} I_i$$

$$\text{Variance } (\sigma^2) = \frac{1}{N}\sum_{i=1}^{N} (I_i - \mu)^2$$

$$\text{Skewness} = \frac{1}{N\,\sigma^3}\sum_{i=1}^{N} (I_i - \mu)^3$$

$$\text{Kurtosis} = \frac{1}{N\,\sigma^4}\sum_{i=1}^{N} (I_i - \mu)^4$$

Entropy is computed from the discretised intensity probability distribution $p\ (\ j\ )$[140]:

$$\text{Entropy} = -\sum_{j} p\ (j) \log p\ (j)$$

The first-order statistical descriptors quantify the distributional properties of voxel intensities within a segmented region. The mean represents the average signal intensity, whereas variance captures dispersion around that mean. Skewness and kurtosis describe asymmetry and peakedness of the distribution, respectively, and entropy reflects the degree of randomness or disorder in discretised intensity values. Importantly, these metrics are agnostic to spatial arrangement and therefore summarise global signal behaviour without encoding structural organisation.[65,66,140,141]

Meanwhile, texture features extend first-order statistics by explicitly incorporating spatial relationships between discretised voxel intensities. Let an image region of interest (ROI) consist of N voxels with discretised gray levels $g(\ x\ ) \in \{1,2, \dots, G\}$, where $G$is the number of gray-level bins after intensity quantisation.[50]

1.7.1.1 Gray-Level Co-occurrence Matrix (GLCM)

The Gray-Level Co-occurrence Matrix (GLCM) models second-order intensity dependencies by quantifying the joint probability of pairs of discretised gray levels separated by a defined spatial offset.[50,142,143]

Let an image region of interest contain discretised gray levels $g(x) \in \{1,2, \dots, G\}$. For a spatial displacement defined by distance $d$and orientation θ, the co-occurrence matrix $P(i\ , j\ |\ d\ , \theta\ )$is defined as:

$$P(i\ , j\ |\ d\ , \theta) = \frac{1}{N\ d\ , \theta\ 1} \sum_{x\ \in\ ROI} 1\ \{g\ (x) = i\ \wedge\ g\ (x + \Delta\ d\ , \theta) = j\}$$

where:

- $\Delta_{\ d,\theta}$ denotes the spatial displacement vector,
- $N_{\ d,\theta}$ is the total number of valid voxel pairs,

- $\mathbf{1}\{\cdot\}$ is the indicator function.

The matrix is typically normalised such that:

$$\sum_{i=1}^{G}\sum_{j=1}^{G} P(i,j \mid d,\theta) = 1$$

From this matrix, texture descriptors are derived. For example, contrast is defined as:

$$Contrast = \sum_{i=1}^{G}\sum_{j=1}^{G}(i-j)^2 P(i,j)$$

Contrast increases as intensity differences between neighbouring voxels become larger, reflecting greater local heterogeneity.[143]

1.7.1.2 Gray-Level Run Length Matrix (GLRLM)

The Gray-Level Run Length Matrix (GLRLM) quantifies the distribution of contiguous runs of voxels with identical gray level along a specified direction θ.[50,144] A "run" is defined as a sequence of consecutive voxels having the same discretised gray level along direction θ. Let $R(i,r \mid \theta)$denote the number of runs with gray level $i$and run length r. The total number of runs is:

$$R(i,r \mid \theta)$$

After normalisation by the total number of runs N_r, features such as short-run emphasis or long-run emphasis are computed, for example:

$$Short-Run\ Emphasis = \frac{1}{N_r}\sum_{i}\sum_{r}\frac{R(i,r)}{r^2}$$

This matrix captures linear structural continuity within tissue.

1.7.1.3 Gray-Level Size Zone Matrix (GLSZM)

The Gray-Level Size Zone Matrix (GLSZM) extends the concept of homogeneity by quantifying connected regions (zones) of identical gray level independent of direction.[50,145]

$$Z(i,s)$$

where $Z(i,s)$denotes the number of connected regions (zones) of gray level i and size s. Unlike GLRLM, GLSZM does not constrain connectivity to a specific orientation.

The total number of zones is:

$$N_z = \sum_i \sum_s Z(i,s)$$

The normalised matrix is:

$$P(i,s) = \frac{Z(i,s)}{N_z}$$

A typical derived feature is zone-size variance:

$$Zone\ Size\ Variance = \sum_i \sum_s (s - \mu_s)^2\ P\ (i,s)$$

where $P(i,s)$ is the normalised matrix and μ_s is the mean zone size. The GLSZM captures spatial clustering and the size distribution of homogeneous regions, providing a direction-independent measure of structural organisation.

1.7.1.4 Gray-Level Dependence Matrix (GLDM)

The GLDM quantifies the number of neighbouring voxels within a defined distance δ whose gray-level difference from a centre voxel does not exceed a tolerance α.[50,146] It is defined as:

$$D(i,k)$$

where $D(i,k)$represents the number of voxels with gray level i that have k dependent neighbours satisfying:

$$|\ g(\ x\ ) - g(\ x_n\ )\ | \leq \alpha$$

Features such as dependence entropy or large-dependence emphasis are then computed from the normalised matrix.

As a whole, these matrices transform spatial intensity configurations into structured probability distributions from which higher-level statistical descriptors are derived. Unlike first-order features, which summarise global intensity distributions:

$$\mu = \frac{1}{N}\sum_{x} g\,(\,x\,)\,,\ \sigma^2 = \frac{1}{N}\sum_{x}(\,g(x) - \mu\,)^2$$

texture matrices explicitly encode spatial organisation and structural continuity.

However, all such matrices are inherently dependent on gray-level discretisation strategy (bin width and number of bins), neighbourhood definition (distance, orientation, connectivity), voxel resolution and anisotropy, and segmentation boundaries.

These features characterise the extent to which voxels form locally homogeneous clusters of varying size. Compared with GLCM, GLRLM, and GLSZM, GLDM directly models local intensity similarity within a tolerance band, making it particularly sensitive to subtle structural aggregation.

1.7.1.5 Conceptual Integration

Collectively, GLCM, GLRLM, GLSZM, and GLDM transform spatial intensity configurations into structured probability distributions from which higher-level statistical descriptors are computed.[50,65]

In contrast, first-order statistics summarise only the marginal intensity distribution:

$$\mu = \frac{1}{N}\sum_{x} g\,(x),\ \sigma^2 = \frac{1}{N}\sum_{x}(g\,(x) - \mu\,)^2$$

These measures describe global signal behaviour but remain agnostic to spatial organisation.

By design, texture matrices explicitly encode spatial dependency, structural continuity, clustering and local homogeneity. This higher-order representation is a concerted effort to amplify sensitivity to acquisition parameters.

All second-order and higher-order matrices are inherently dependent on:

- Gray-level discretisation strategy (bin width, bin number)
- Neighbourhood definition (distance δ, orientation, connectivity scheme)
- Voxel resolution and anisotropy
- Segmentation boundaries
- Image preprocessing and reconstruction choices

Because matrix construction precedes feature derivation, any perturbation in acquisition, reconstruction, or preprocessing may alter matrix composition and thus downstream feature values.[50,103,105] In mpMRI, these dependencies become particularly consequential. Voxel intensity distributions are vary depending on b-value selection, gradient timing, echo time, and reconstruction algorithms.[44,47,123] This is further expounded by Mackin et al[103] and Orlhac et al[147] that at a certain degree, small parameter adjustments shifts discretised gray-level histograms, creating differences within texture matrices and altering derived statistics.

1.7.2 Transform-Domain Features

Gillies et al[65] and Mayerhoefer et al[66] defines transform-domain features project images into alternative representational domains to emphasise specific spatial characteristics. An example are Fourier-based descriptors, implemented via two-dimensional Fast Fourier transform (FFT2D). The feature quantifies global frequency content by decomposing the image into sinusoidal components across spatial frequencies.[141,148]

Additionally, Gabor filters broaden frequency analysis by integrating sinusoidal waveforms with Gaussian envelopes, enabling orientation- and frequency-selective filtering at defined spatial scales.[149,150] Unlike global Fourier features, Gabor responses retain spatial localisation and directional sensitivity, allowing multi-scale characterisation of oriented patterns within tissue architecture.[149] These features summarise dominant frequency distributions without preserving spatial localisation.

1.7.2.1 Fourier Transform (FFT2D)

The two-dimensional discrete Fourier transform (DFT), typically computed via the Fast Fourier Transform (FFT2D), is defined as:[141,148]

$$F\ (u\ ,v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} I\ (x\ ,y)\ e^{-i\ 2\ \pi\left(\frac{u\ x}{M}+\frac{v\ y}{N}\right)}$$

where:

- is the image intensity at spatial coordinates $(x\ ,\ y)$,
- $F(u\ ,v)$represents the complex frequency coefficients,
- $(u\ ,\ v)$correspond to spatial frequency indices,
- $M$and Ndenote image dimensions.

Conceptually, the Fourier transform decomposes an image into a weighted combination of sinusoidal basis functions. Each coefficient F(u,v) describes the contribution of a particular spatial frequency to the overall image.[148]
The magnitude spectrum is defined as:

$$|F\ (u\ ,v)| = \sqrt{Re\ (F\ (u\ ,v))^2 + Im\ (F\ (u\ ,v))^2}$$

The magnitude quantifies how strongly a given frequency component is represented in the image. Low spatial frequencies correspond to slowly varying intensity patterns that encode global structural organisation, whereas high spatial frequencies correspond to rapid intensity transitions such as edges and fine textural detail.[141] However, classical Fourier analysis does not preserve spatial localisation. Jain and Farrokhnia[149] argues that frequency components only indicates that a pattern exists somewhere within the image, but not where it is located. The global nature of this image processing limits anatomical interpretability and motivates the use of spatially localised transforms such as Gabor filtering.

1.7.2.2 Gabor Filtering

Gabor[151] presents the foundational work for Gabor functions, Bovik et al[149] and Jain & Farrokhnia[150] expands this work for texture segmentation and 2D processing frameworks, where Gabor filters address the spatial localisation limitation of classical Fourier analysis by introducing Gaussian windowing around a sinusoidal carrier.

A two-dimensional real-valued Gabor kernel is defined as:

$$G(x, y) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right)\cos\left(\frac{2\pi x'}{\lambda} + \phi\right)$$

where:

- σcontrols the width of the Gaussian envelope (spatial scale),
- λdefines the wavelength of the sinusoidal carrier (spatial frequency),
- θdetermines orientation via rotated coordinates (x', y'),
- γcontrols ellipticity (aspect ratio),
- $\phi$is the phase offset.

The rotated coordinates are defined as:

$$x' = x\cos\theta + y\sin\theta$$

$$y' = -x\sin\theta + y\cos\theta$$

Mathematically, the Gabor response is obtained by convolution of the image $I(x, y)$ with the kernel $G(x, y)$:

$$R(x, y) = I(x, y) * G(x, y)$$

The Gaussian envelope spatially localises the sinusoidal oscillation, enabling joint spatial–frequency analysis.[149,152]

In practical terms:

- σ controls scale sensitivity

- λ determines frequency selectivity
- θ enables orientation selectivity
- γ shapes anisotropy

This makes Gabor filters particularly suited for detecting oriented architectural patterns such as glandular boundaries or stromal alignment.[153]

Although transform-domain approaches enhance sensitivity to spatial heterogeneity and structural periodicity, their derived features remain functions of voxel intensity distributions. Consequently, feature values are influenced by acquisition resolution, sampling frequency, reconstruction parameters, and discretisation strategy, all of which may affect stability under repeat imaging conditions.[50,103,153,154]

1.7.3 Conceptual Contrast and Stability Considerations

The conceptual contrast between FFT2D and Gabor is expanded in Table 1.2:

Table 1.2 Transform-Domain Features Conceptual Contrast

FFT2D	GABOR
Global frequency decomposition	Localised frequency analysis
No spatial localisation	Orientation selective
Efficient representation of periodic structure	Multi-scale capability
Limited anatomical interpretability	More biologically interpretable in tissue imaging

Despite operating in the frequency domain, transform-based features remain deterministic functions of voxel intensity distributions.[65,141] Consequently, their numerical behaviour is inherently dependent on acquisition and preprocessing characteristics:

- Acquisition resolution affects sampling frequency and Nyquist limits.
- Voxel size constrains the spatial frequencies that can be represented.
- Reconstruction filtering alters high-frequency content and edge definition.
- Gray-level discretisation modifies spectral energy distribution prior to transformation.[50,154]

Bracewell[148] emphasized that Fourier-based representations are inherently sensitive to sampling density and interpolation effects, as spectral coefficients directly reflect the underlying intensity sampling structure. Although Gabor responses provide spatial localisation, they remain dependent on scale selection and kernel parameterisation, introducing additional variability across acquisition settings.[149] Thus, while transform-domain features enhance sensitivity to spatial heterogeneity and directional organisation, their quantitative stability under repeat acquisition depends on imaging parameter consistency and preprocessing standardisation.[103,105] For quantitative biomarker development, the critical methodological issue is therefore not solely the detection of heterogeneity, but whether frequency-derived descriptors remain reproducible across platforms, parameterisations, and reconstruction pipelines.[47,48,50]

1.7.4 Gradient-Derived and Structural Descriptors

Beyond intensity- and texture-based radiomic features, a distinct class of descriptors is derived from spatial intensity derivatives. In this thesis, these features are implemented within the two dimensional or three dimensional Edges- and Corners-Based (EC2D3D) framework, referring to edge- and curvature-based two- and three-dimensional structural descriptors extracted from gradient- and Hessian-enhanced image representations.[155,156] EC2D3D operators quantify local changes in signal intensity and characterise structural properties such as edge strength, curvature, anisotropy, and geometric organisation. Rather than modelling diffusion signal behaviour or estimating biological compartments, they operate directly on the spatial intensity field.

1.7.4.1 Scharr-Based Object Properties

The Scharr operator is a first-order derivative operator designed to approximate the image gradient with improved rotational symmetry and reduced discretisation error compared with earlier operators such as Sobel.[155,157] It estimates spatial intensity change along horizontal and vertical axes through convolution with derivative kernels.

If $I\ (\ x, y\ )$ denotes image intensity, the gradient magnitude is:

$$|\nabla I| = \sqrt{\left(\frac{\partial I}{\partial x}\right)^2 + \left(\frac{\partial I}{\partial y}\right)^2}$$

The gradient vector describes the direction and magnitude of maximal first-order intensity change.[158] Scharr kernels provide an optimised discrete approximation to these derivatives, reducing angular bias introduced by finite-difference sampling.[155]

Scharr-based features typically derive from the gradient magnitude map or from thresholded edge representations. Object-level descriptors may include edge density, edge orientation distribution, minor and major axis lengths of connected edge components and relative eccentricity of detected structures.

These properties reflect boundary sharpness, structural elongation, and local architectural organisation, capturing first-order geometric variation rather than second-order curvature.[155,157]

1.7.4.2 Hessian-Based Object Properties

The Hessian matrix captures second-order spatial derivatives of the image intensity function and provides a local description of curvature:[157,158]

$$H(x,y) = \begin{bmatrix} \frac{\partial^2 I}{\partial x^2} & \frac{\partial^2 I}{\partial x\,\partial y} \\ \frac{\partial^2 I}{\partial y\,\partial x} & \frac{\partial^2 I}{\partial y^2} \end{bmatrix}$$

The eigenvalues λ_1 , λ_2 of the Hessian describe the principal curvature directions and quantify the magnitude of second-order intensity variation. Their relative magnitude and sign enable structural classification:[156,158]

- Similar magnitude eigenvalues → Blob-like structures
- One dominant eigenvalue → Ridge- or line-like structures
- Opposing sign patterns → Sheet- or saddle-like structures

In multiscale filtering frameworks, these eigenvalue relationships are used to distinguish blob-like, ridge-like, or sheet-like objects in biomedical images.[156,159] Hessian-based features therefore quantify local curvature and structural bending rather than global intensity distribution.

Object-level properties may subsequently be derived from connected regions defined by curvature or vesselness. These include morphometric descriptors such as area, minor axis length, major axis length and area variability (e.g., interquartile range).

Hessian descriptors translate differential curvature information into geometrically interpretable object metrics, reflecting architectural distortions and boundary morphology.[157,158]

1.7.4.3 Frangi-Based Object Properties

The Frangi filter is derived from the Hessian framework. It analyses relationships between Hessian eigenvalues to enhance elongated or tubular structures.[156,158] Vesselness-type functions are computed based on eigenvalue ratios, at each spatial scale, the Hessian matrix is computed and its eigenvalues λ_1, λ_2 (in 2D) are evaluated to characterise local structure.

In two dimensions, the vesselness measure is defined as:

$$V(x,y) = \exp\left(-\frac{R_B^{\,2}}{2\,\beta^{\,2}}\right)\left(1-\exp\left(-\frac{S^{\,2}}{2\,c^{\,2}}\right)\right)$$

where:

$$R_B = \frac{|\,\lambda_1\,|}{|\,\lambda_2\,|} \text{ (blobness vs line-likeness ratio)}$$

$$S = \sqrt{\lambda_1^2 + \lambda_2^2} \text{ (second-order structureness)}$$

and β and c are scale parameters controlling sensitivity to anisotropy and overall second-order response magnitude.[156]

Conceptually:

- Small R_B→ line-like structure
- Large R_B→ blob-like structure
- Large S→ strong second-order structure

The filter is typically evaluated across multiple spatial scales, with the maximum vesselness response retained at each voxel to enable scale-dependent detection of anisotropic patterns.[156]

Although originally developed for vascular enhancement, Frangi-based object properties can quantify elongated or anisotropic structures within tissue. Derived descriptors often measure: tubularity, structural continuity, scale-dependent object extent and directional anisotropy.

Thus, Frangi-based features represent a curvature-driven, eigenvalue-ratio approach to structural characterisation, extending Hessian analysis towards further object-level interpretation.

1.7.4.4 Harris-Stephens Corner Detector

Unlike the previous detectors, the Harris-Stephens operators are derived from the second-moment matrix, also known as the structure tensor.[157,160] It is constructed from locally weighted products of first-order image derivatives:

$$M(x\,,y) = \begin{pmatrix} \sum w(x,y)\, I_x^2 & \sum w\,(x,y)\, I_x I_y \\ \sum w\,(x,y)\, I_x\, I_y & \sum w(x,y)\, I_y^2 \end{pmatrix}$$

where:

- $I_x = \frac{\partial I}{\partial x}$ and $I_y = \frac{\partial I}{\partial y}$ are first-order spatial derivatives,
- $w(x,y)$is a local weighting function (typically Gaussian),
- the summation is taken over a local neighbourhood.

The eigenvalues denoted by M describe intensity variation along orthogonal directions. This means if both eigenvalues are large, intensity varies strongly in multiple directions, which indicates a corner or junction structure.

The Harris-Stephens response function is defined as:

$$R = \det(M) - k\,(\text{trace}\,(M))^2$$

where:

$$\det(M) = \lambda_1\lambda_2 \text{ and } \text{trace}\,(M) = \lambda_1 + \lambda_2$$

and k is an empirical sensitivity parameter (typically 0.04–0.06) [160].

This formulation avoids explicit eigenvalue computation while preserving sensitivity to multi-directional intensity variation.

Object-level properties derived from Harris-Stephens response maps may include: corner density, spatial distribution of corner responses, response magnitude statistics and clustering or intersection metrics. These descriptors reflect structural complexity, junction density, and architectural intersections within tissue patterns.

1.7.4.5 Shi–Tomasi Detector

Shi & Tomasi (1994) established detectors that further refined pre-existing Harris-Stephens corner operator and is based directly on the eigenvalues of the structure tensor. Given the second-moment matrix M , with eigenvalues λ_1 and λ_2 , the Shi–Tomasi response is defined as:

$$R = min\,(\lambda_1, \lambda_2)$$

Instead of combining determinant and trace terms concurrent in the Harris-Stephens formula, the Shi-Tomasi criterion directly assesses the smallest eigenvalue. A point could be classified as a corner if both principal gradient directions are sufficiently large, which ensures that intensity changes are significant in orthogonal directions. As a result, Shi-Tomasi criterion produces more stable and better-localised corner detection compared with the Harris-Stephens response, particularly in tracking and repeat-detection contexts. Derived object-level features may include: number of detected

interest points, distribution of minimum eigenvalue magnitudes and spatial clustering or dispersion characteristics. These descriptors reflect the density and spatial organisation of structural intersections within the tissue architecture.

1.7.4.6 Conceptual Positioning

The following is therefore a conceptual contrast table of the structural operators (Table 1.3).

Table 1.3 Conceptual Operators of Structural Descriptors:[155–161]

Operator	Derivative Order	Core Mathematical Basis	What It Detects	Structural Interpretation
Scharr	First-order	Spatial intensity gradient ($\partial I/\partial x$, $\partial I/\partial y$)	Edge magnitude	Boundary sharpness and elongation
Hessian	Second-order	Matrix of second derivatives	Curvature eigenvalues	Blob, ridge, or sheet-like structures
Frangi	Second-order (Hessian-derived)	Eigenvalue ratio analysis	Tubular/anisotropic patterns	Directional structural continuity
Harris-Stephens	First-order (structure tensor)	Gradient covariance matrix	Multi-directional intensity change	Corner-like structural intersections
Shi–Tomasi	First-order (structure tensor)	Minimum eigenvalue of tensor	Stable corner detection	Local structural complexity

Collectively, these operators share a common mathematical foundation in spatial derivatives of the image intensity field.[157,158] They quantify:

- Edge strength (first-order derivatives; Scharr-type operators)
- Curvature (second-order derivatives; Hessian-based operators)
- Junction complexity (structure tensor methods; Harris-Stephens and Shi–Tomasi)
- Anisotropic structural patterns (Frangi-based eigenvalue filters)

Unlike model-based diffusion features, EC2D3D operators do not estimate biological signal compartments or fit parametric diffusion models. Instead, they derive structural descriptors from spatial derivatives of the image intensity field. Within the EC2D3D framework, first- and second-order differential operators are applied to generate gradient- or curvature-enhanced maps from which object-level properties are

subsequently extracted. Because these operators rely on derivative estimation, their behaviour is influenced by image resolution, smoothing scale, and noise characteristics.[50,158] However, they remain mathematically distinct from diffusion signal modelling approaches. Their outputs reflect geometric and architectural organisation rather than compartmental signal decomposition.

Since each feature family is grounded in a distinct computational principle, including intensity statistics, spatial probability matrices, frequency transforms, or differential geometry, their stability and repeatability under varying acquisition conditions require independent empirical evaluation.[48,50,105]

1.7.5 Radiomic Feature Families

The feature families summarised in Table 1.4 differ fundamentally in their mathematical construction and in the type of structural information they encode.

Table 1.4 Feature Family

Feature Family	Mathematical Basis	Structural Focus	Local or Global
First-Order Moments	Intensity distribution statistics	Global signal behaviour	Global
Texture Matrices (GLCM, GLRLM, GLSZM, GLDM)	Spatial gray-level probability matrices	Local heterogeneity and spatial dependency	Local
FFT2D	Global frequency decomposition (Discrete Fourier Transform)	Periodic spatial patterns	Global
Gabor	Localised frequency filtering (Gaussian-modulated sinusoids)	Oriented, multi-scale patterns	Local
EC2D3D – Gradient-Based (Scharr)	First-order spatial derivatives	Edge strength and boundary sharpness	Local
EC2D3D – Hessian-Based	Second-order spatial derivatives	Curvature and structural geometry	Local
EC2D3D – Frangi-Based	Hessian eigenvalue ratios (multiscale)	Anisotropic or tubular patterns	Local (multi-scale)
EC2D3D – Corner Detectors (Harris-Stephens, Shi–Tomasi)	Structure tensor of gradients (second-moment matrix)	Structural intersections and junction complexity	Local

Deriving from first-order intensity features, as implemented in PyRadiomics toolkit, intensity-based features are global signal distributions without incorporating spatial relationships.[162] While transform-domain methods, including FFT2D[141,148] and Gabor

filters[149–151,153] project images into the frequency space to characterise periodicity and orientation-specific patterns within either a global or local region of interest. In contrast, EC2D3D operators derive structural descriptors from first- and second-order spatial derivatives of the intensity field, capturing edge strength, curvature, anisotropy, and structural intersections through differential geometry rather than signal modelling.[155–161]

Because each feature family is grounded in a distinct computational principle, including intensity statistics, their stability and repeatability under varying acquisition in mpMRI application requires independent empirical evaluation.[33,47,48,50,103,105,154]

1.7.6 High-Dimensional Modelling Challenges

In PCa research, radiomic features are often incorporated into statistical and machine learning models investigating classification and risk stratification potentials. As a result, Mizuno and Beltran[31] suggests that future imaging-derived features are combined with clinical variables and genomic data to construct multimodal predictive models.

Nevertheless, the outlook of high-dimensional modelling is accompanied by persistent methodological issues. Overfitting bias is a possible risk due to radiomic datasets frequently contain hundreds or thousands of features proportionate to modest sample sizes.[67,163] Furthermore, model performance may be actively affected by feature selection strategy, thresholding decisions, and validation design.[68,105] Observed apparent discriminative accuracy may reflect false positives through statistical fluctuation rather than stable signal, especially when repeatability and robustness are not explicitly checked.[47,48] As such, crosswise assessment of feature stability and validation among independent or repeat datasets is crucial to determine reproducible QIBs from artefacts of modelling complexity.[33,50]

1.7.7 Reproducibility as a Central Constraint

Reproducibility remains one of the principal challenges in radiomic research.[50,105] In radiomics, reproducibility is defined as stability against variety in scanner vendor, magnetic field strength, acquisition protocol, reconstruction algorithms, segmentation strategy, and gray-level discretisation.[69,103,154] Small changes in preprocessing have been shown to alter derived feature values.[50]

Test–retest research have corroborated that reliability fluctuate substantially across feature families and imaging modalities.[33,105] Strong discriminative performance in a single dataset does not necessarily indicate robustness under repeat acquisition.[48,124,163–165] A feature may separate groups effectively yet still exhibit considerable measurement variability when acquisition conditions are repeated or slightly perturbed, leading to false-positive results.

Due to feature families being grounded in different distinct computational principles from intensity statistics to differential geometry and frequency-domain transforms, it leads to increased susceptibility towards acquisition perturbations.[50,65] Whether such theoretical differences translate into meaningful variation in repeatability must be determined empirically through structured experimental design.[47] Making repeatability a central constraint in QIB development that needs addressing.[48]

1.7.8 Stability-Oriented Evaluation in Quantitative Imaging Biomarkers

As the dimensionality of radiomic datasets increases, the underlying methodological framework becomes central to meaningful biomarker evaluation. Sequential or stability-informed filtering strategies that integrate repeatability, reliability in discriminative capacity, and statistical precision have been proposed to mitigate variance-driven feature selection and reduce the risk of overfitting.[47,48,105]

Rather than assuming equivalence across feature families or imaging modalities, such structured approaches explicitly evaluate performance within predefined stability constraints.[33,50] This shifts emphasis from exploratory discrimination alone, towards reproducible quantitative behaviours.[163]

The present thesis adopts this framework to systematically assess diffusion-parameterised mpMRI radiomic features under predefined repeatability and precision criteria.

1.8 Quantitative Imaging Biomarkers and Standardisation

Quantitative imaging biomarkers (QIBs) are intended to translate medical images into objective and clinically meaningful measurements.[48,166] Unlike qualitative assessments,

which rely on visual interpretation, QIBs are expected to produce numerical values that remain consistent under defined acquisition conditions and maintain a stable relationship with relevant clinical outcomes.[50,167]

The development of a QIB therefore requires attention to three closely related but conceptually distinct dimensions: repeatability, reproducibility, and discrimination.[166] Repeatability refers to measurement stability under ideal conditions, reproducibility reflects robustness across variations in scanners or protocols, and discrimination. These dimensions reflects diagnostic and prognostic performance, reporting standards emphasize that accuracy must be interpreted alongside methodological rigor and validation design.[168,169]

1.8.1 Repeatability and Reproducibility

Sullivan et al[48] and Raunig et al[166] underpins the definition of repeatability as the agreement between repeated measurements under identical conditions within a short time interval. Shukla-Dave et al[47] further elaborates this in the context of MRI research, which typically assesses repeatability through test–retest designs using the same scanner with an unchanged acquisition protocol.

On the other hand, Buckler et al[167] defines reproducibility as an extension of the repeatability concept to measurements obtained under varying conditions, including differences in scanner hardware, institutions, operators, or acquisition settings. This in turn, validates generalisability across contexts and is essential for translation into multi-centre clinical practice.[48]

Measurement stability is commonly quantified using the intraclass correlation coefficient (ICC), which expresses the proportion of total variance attributable to between-subject variability rather than measurement error.[170,171] In general form, ICC may be written as:

$$\mathrm{ICC} = \frac{\sigma^2_{between}}{\sigma^2_{between} + \sigma^2_{error}}$$

where $\sigma^2_{between}$ represents variance between subjects and σ^2_{error} represents residual variance. ICC(3,1) is defined as:

$$\text{ICC(3,1)} = \frac{MS_R - MS_E}{MS_R + (k-1)MS_E}$$

where:

- MS_R is the mean square for rows (between subjects),
- MS_E is the residual mean square (error),
- k is the number of repeated measurements.

Unlike ICC(1), which assumes measurement conditions are randomly sampled from a broader population, ICC(3) treats measurement conditions as fixed and therefore evaluates agreement under a particular, predefined acquisition setup.[170,172] ICC(3) is appropriate for test–retest studies conducted on the same scanner and protocol, where the objective is to quantify repeatability rather than reproducibility.

Importantly, ICC values are influenced by cohort variance and study design. Because ICC represents a ratio of between-subject variance to total variance, high between-subject heterogeneity may inflate ICC even when absolute measurement error remains non-negligible.[166,173] Consequently, ICC should be interpreted alongside complementary precision metrics rather than as a standalone indicator of biomarker stability.[47,48]

1.8.2 Discrimination, Reliability and Statistical Precision

Discrimination refers to the ability of a feature to differentiate clinically relevant categories, such as low-grade versus high-grade disease.[168,174] In binary classification settings, discrimination is commonly evaluated using the receiver operating characteristic (ROC) curve.[175]

The ROC curve plots:

$$\text{True Positive Rate (Sensitivity)} = \frac{TP}{TP + FN}$$

against

$$\text{False Positive Rate} = \frac{FP}{FP + TN}$$

across varying decision thresholds. By sweeping the classification threshold from low to high values, the ROC curve characterises the trade-off between sensitivity and specificity.[176]

The area under the ROC curve (AUC) summarises overall discriminative performance:

$$\text{AUC} = \int_0^1 \text{TPR (FPR)}\, d\ \text{(FPR)}$$

The AUC can be interpreted probabilistically as:

$$\text{AUC} = P(X_{positive} > X_{negative})$$

The formula underpins the probability that a randomly selected positive case has a higher feature value than a randomly selected negative case.[174,175] Under this interpretation, AUC is mathematically equivalent to the normalised Mann–Whitney U statistic, also known as the Wilcoxon rank-sum test.[175,177] The Wilcoxon statistic evaluates whether the distributions of two independent groups differ in central tendency without assuming normality.[178] The relationship between AUC and the U statistic is:

$$\text{AUC} = \frac{U}{n_{positive} \cdot n_{negative}}$$

where U is the Mann–Whitney statistic and $n_{positive}, n_{negative}$ are group sample sizes.[177]

However, discrimination alone does not ensure robustness. A feature may demonstrate a high AUC within a single dataset yet exhibit poor repeatability under repeated acquisition. Therefore, statistical precision provides additional context for interpreting discriminative performance. Precision is mainly evaluated through the width of confidence intervals (CI) around AUC estimates. Narrow CI are typically interpreted as stable estimation, whereas wide intervals may reflect uncertainty in discriminative performance.[174,175] A 95% confidence interval for AUC could be calculated using nonparametric methods or variance estimators such as the DeLong approach, which

reports for correlated ROC curves and provides an asymptotically unbiased estimate of AUC variance.[179]

Furthermore, the probability of false-positive findings increases substantially when larger feature numbers are examined.[180,181] In order to control for multiplicity, procedures such as false discovery rate adjustment (FDR) controls the expected proportion of false discoveries among rejected hypotheses by ranking p-values and applied adaptive significance thresholds.[180]

Conceptually, precision and multiplicity control are linked to reliability. Therefore, a feature with a high AUC but wide confidence intervals may reflect statistical instability rather than reproducible discrimination. Likewise, features that achieve nominal statistical significance may represent stochastic variation unless supported by repeatability evidence.[48,163] In repeatability-focused biomarker studies, reliability must therefore be integrated with measurement stability, precision of effect estimates, and multiplicity control. Without such integration, apparent performance could not be deemed repeatable under similar acquisition conditions.

It is for these reasons, repeatability (e.g., ICC), discrimination reliability (AUC), and statistical precision (confidence interval width and FDR control) represent complementary criteria in QIB evaluation. Unitedly, these metrics are instrumental in distinguishing reproducibility from variance-driven associations.

1.8.3 Standardisation Initiatives in Quantitative Imaging

Recognition of reproducibility challenges in quantitative imaging has initialized coordinated international standardisation efforts.[48,167] The Quantitative Imaging Biomarkers Alliance (QIBA), established by the Radiological Society of North America, agreed to a consensus on technical performance targets, acquisition requirements, and precision claims for QIBs.[47,48] These technical profiles aim to ensure that obtained measurements across scanners and institutions reach predefined thresholds for repeatability, reproducibility and bias, supporting clinical and research reliability.[166]

Complementing this effort, the Image Biomarker Standardisation Initiative (IBSI) focuses specifically on radiomics. IBSI provides mathematically explicit definitions for

radiomic features, image preprocessing steps, and discretisation strategies.[50] A key trait of IBSI benchmarking is to ensure that features sharing identical names may yield numerically atypical values across software platforms unless definitions are rigorously harmonised.[50,103] Generally, these initiatives emphasise that methodological transparency, harmonisation, and empirical stability assessment are prerequisites for biomarker validity.[48,105] In mpMRI, adherence to reproducibility-oriented frameworks is particularly critical. Without structured stability, heightened microstructural sensitivity risks being counterpoised by measurement variability.

1.9 Methodological Foundations: The PRODIF Test-Reposition-Retest Cohort

The present study is compassed with the test–retest framework established in the PRODIF cohort, in which subjects underwent two prostate MRI examinations on the same day using identical scanner hardware and acquisition protocols, with patient repositioning between scans. This study design enables direct assessment of short-term repeatability under controlled technical conditions and isolates measurement variability from biological change.[33]

In the original repeatability analysis of the PRODIF dataset, radiomic features derived from DWI demonstrated substantial variability in stability, with intraclass correlation coefficients (ICC(3,1)) ranging widely across feature classes.[33] It was argued that discriminative performance alone did not guarantee repeatability, and feature selection strategies influenced stability of downstream machine learning models. These findings determined that short-term measurement stability cannot be established even when acquisition parameters are held constant. The PRODIF design therefore provides a structured experimental environment for evaluating quantitative imaging features under predefined repeatability conditions.[33]

1.10 Radiomics in PCa: Reported Performance and Methodological Limitations

Radiomic analysis of mpMRI has been widely investigated in PCa, with numerous studies report associations between radiomic features and clinically relevant endpoints, including Gleason grade group, biochemical recurrence, extracapsular extension, and treatment response.[49,139,164,182,183] Reported discriminative performance

metrics frequently demonstrate moderate to high AUC values for differentiating clinically significant from indolent disease. Moreover, Shiradkar et al[184] and Castillo et al[185] reported improved classification performance when radiomic features are integrated with PSA levels or biopsy findings.

However, due to a lack of standardized methodological approach that characterizes this literature. Prathyush Chirra et al[186] and Traverso et al[105] argues that inconsistency complications still persist in image acquisition protocols, field strength and scanner vendor, segmentation strategy, feature extraction software, gray-level discretisation methods, feature selection pipelines and validation design, which in turn, limits application generalisability. Feature dimensionality often greatly exceeds cohort size, increasing susceptibility to overfitting and instability of model estimates.[187]

Emerging data from repeatability-focused investigations highlights that radiomic stability varies substantially across feature families and imaging sequences, particularly in DWI.[33,165] This poses a principal methodological problem: which distinct radiomic features retain discriminative performance while maintaining repeatability under controlled acquisition conditions?

The unresolved methodological question is therefore not whether radiomic features can discriminate prostate cancer grade, but which features retain discriminative capacity when subjected to structured stability constraints. Against this backdrop, this thesis operationalises the hypothesis through stability-driven evaluation framework in mpMRI. Through enforcing repeatability, reliability and precision criteria, this study examines whether longer acquisition strategies could translate into clinically meaningful QIBs for csPCa.

2 Aims and Hypothesis

2.1 Hypothesis

Longer or novel multiparametric magnetic resonance imaging (mpMRI) modalities are expected to enhance reliability and repeatability, resulting in more clinically relevant data for the diagnosis and treatment of prostate cancer.

2.2 Major Aim

- Evaluate whether longer or novel multiparametric magnetic resonance imaging (mpMRI) modalities enhance the reliability of prostate cancer assessment.
- Investigate whether these modalities improve the repeatability of imaging results across different image weightings and settings.
- Assess whether the increased reliability and repeatability translate into more clinically relevant data for prostate cancer diagnosis, risk stratification, and treatment planning.

3 Research Design and Methodology

3.1 Study Population and Data Source

The dataset was derived from a PRODIF study, from which a cohort of 37 patients was selected. All patients underwent radical prostatectomy, with whole-mount histopathology serving as the reference standard and providing Gleason Grade Groups (GGG). Image data were acquired using DBSI, high-resolution DWI, OGSE, SPAIR and T2-weighted imaging, from which radiomic features were extracted, including First Order Moments, Gabor filters, FFT-based features, Edges- and Corners-2D/3D, and PyRadiomics features. Of the initial cohort of 37 patients with histopathologically confirmed GGG, feature availability varied across imaging modalities and radiomic parameters.

From the initial cohort of 37 patients, exclusions occurred during analysis due to non-computable radiomic features, missing feature values, unavailable paired test–retest scans, or mismatch between histopathological ground truth and the radiomic dataset. These exclusions were applied dynamically and at the feature- and modality-specific level rather than through a single patient-level exclusion step. As a result, the effective sample size varied across imaging modalities, lesions, and radiomic features. For DBSI low-resolution parameters, between 30 and 31 patients were retained per feature, whereas OGSE-derived parameters were available for 23 to 28 patients per feature, reflecting differences in sequence availability and successful test–retest acquisition.

3.2 MPMRI Acquisition Protocol and Imaging Modalities

All imaging was performed before radical prostatectomy to provide data for radiomic analysis. Imaging was performed using a 3T Philips Ingenuity PET/MR scanner with 32-channel cardiac coil, as previously described in the PRODIF test–retest protocol.[33]

The MRI sequences and their extracted radiomic parameters are summarised in Table 3.1. Each sequence was acquired for each patient:

Table 3.1 MRI sequences and their extracted radiomic parameters used for radiomic analysis

[1]Study-specific sub-parameters; see List of Abbreviations for definitions.

MRI Sequence	Purpose / Notes	Extracted Parameters
DBSI (Diffusion Basis Spectrum Imaging)	High-resolution diffusion imaging capturing isotropic and anisotropic components	Low Resolution AD[1], Low Resolution RD[1], Low Resolution F1AD[1], Low Resolution F1RD[1], Low Resolution RR[1], Low Resolution WR[1]
High-resolution DWI (Diffusion-Weighted Imaging)	Multi-b-value diffusion to assess tissue microstructure	ADC[1]
OGSE (Oscillating Gradient Spin Echo)	Sensitive to short diffusion times; probes microstructural restrictions	N1 ADC[1], N1 C[1], N2 ADC, N2 C[1], N4 ADC, N4 C[1], N6 AD[1], N6 C[1]
SPAIR (Spectral Attenuated Inversion Recovery)	Fat-suppressed T2-weighted imaging to enhance lesion contrast	4b ADC[1], 12b ADC[1]
T2-weighted imaging (T2w)	Standard anatomical imaging provides high soft-tissue contrast	—

Acquisition parameters, including repetition time (TR), echo time (TE), voxel size, slice thickness, b-values and number of averages, were recorded for each sequence to enable reproducible radiomic feature extraction. All images underwent standard pre-processing, including motion correction, co-registration across sequences, and intensity normalisation, to ensure consistency of radiomic features across all patients and sequences. These acquisition and pre-processing steps ensured reliable data for lesion-level radiomic analysis.

3.3 Radiomic Feature Extraction

Radiomic feature extraction was performed to quantify lesion heterogeneity and structural characteristics from the acquired MRI sequences. The extracted distinct radiomic descriptors are listed in table 3.2.

Radiomic features were extracted using both custom-designed pipelines and the PyRadiomics framework.[162] The custom radiomic feature extraction followed the MRCradiomics pipeline, previously validated in a cohort of 112 PCa patients from the PRODIF study.[33]

Table 3.2 List of radiomic feature family and extracted parameters

*Full parameterisation, scale settings, and response thresholds are detailed in Appendix 1.

Feature family	Purpose/Notes	Extracted parameters
First Order Statistics	Basic intensity-based measures	Mean, Median, 25th percentile (p25), 75th percentile (p75), skewness, kurtosis, standard deviation, range, ml, Coefficient of Variation (CV)
Gabor-filtered features	Capture textural heterogeneity across multiple frequencies and orientations, enabling quantification of microstructural heterogeneity beyond simple intensity measures.	Mean, median, p25, p75, SD, IQR, skewness, kurtosis, background-adjusted metrics (meanBG, medianBG), contrast-based metrics (Cmean, Cmedian, CNR)
FFT-based features in 2D (FFT2D)	Quantify frequency-domain characteristics of lesion textures, capturing fine and coarse patterns in tissue microstructure.	Low- and high-pass filtered features (FWHM 1–5); mean, median, SD, IQR, skewness, kurtosis, p25, p75, relative within-ROI metrics
Edges- and Corners- 2D/3D (EC2D3D)*	Multi-scale structural radiomic features derived from corner, edge, and object-based filters that characterize lesion morphology, boundary complexity and microstructural organization.	Feature families include: Corner response and density metrics (Harris-Stephens, Shi–Tomasi; ROI, background, and ratio-based). Object-based morphology descriptors (Frangi, Hessian), including area, eccentricity, axis lengths, orientation, perimeter, density, and object count (absolute and relative measures). Gradient-based structural features (Scharr) capturing edge-driven morphology and intensity descriptors. All features were extracted in both 2D and 3D across multiple scales and response thresholds[1], as detailed in the Appendix.
PyRadiomics*	Standardised benchmark radiomics framework	First-order, GLCM, GLRLM, GLSZM, GLDM, NGTDM, and shape features extracted using Original, LoG, and Wavelet filters (total 1062 features)[1]

The inclusion of both standardised (PyRadiomics) and custom-designed radiomic feature families aimed to balance methodological reproducibility with sensitivity to prostate-specific microstructural patterns. This extraction design enabled subsequent

evaluation of feature robustness, repeatability and discriminative relevance while reducing reliance on any single radiomic paradigm through inter-class analysis.

In this thesis, the term “distinct radiomic descriptor” refers to a non-duplicated radiomic feature. Moreover, “distinct radiomic feature-modality combination” refers to a non-duplicated radiomic feature-modality pairing after collapsing primary and secondary lesion entries and “distinct radiomic feature-modality-lesion” refers to a non-duplicated radiomic feature-modality pairing that was not collapsed by primary and secondary lesion entries. These terms do not imply statistical independence or decorrelation between features.

3.4 Repeatability and Reliability Analysis

To determine whether a radiomic paradigm could be classified as a valid QIB, it must be based on quantifiable features that can be reliably and objectively measured on a continuous quantitative scale. The fundamental purpose of a QIB is to provide information beyond visual assessment, thereby improving diagnostic accuracy, prediction and longitudinal monitoring imperative for treatment. Consequently, both repeatability and reliability analysis are necessary to establish that an extracted radiomic feature represents stable and meaningful measurements rather than acquisition- or processing-dependent artefacts.

Repeatability is a prerequisite for all subsequent analyses to ensure that radiomic features validate stable performance under repeated measurements obtained under proportionate conditions. A feature that appears reliable but lacks repeatability may yield apparently discriminative results while failing to reproduce them consistently. Therefore, from a quantitative imaging perspective, such features cannot be considered robust biomarkers, as instability would undermine reproducibility and limit clinical translatability.

Accordingly, repeatability checks were used to pinpoint features with higher measurement quality, ensuring that only stable and reproducible features were carried forward for reliability appraisal and downstream assessment.

3.4.1 Repeatability Study Design

Repeatability was assessed using a test-retest design under identical imaging conditions. All repeat measurements were acquired using the same MRI scanner, acquisition protocols, reconstruction parameters and segmentation strategy with no change in reader or processing workflow. Two prostate MRI examinations were performed on the same day. After the first scan, the patient was removed from the scanner, rested for 10-15 minutes, and repositioned before the second repeat scan.

This design attenuates biological change as a confounding variable, intentionally introducing small positioning variability while isolating measurement repeatability. Facilitating assessment of feature stability while assuming negligible biological change, as the time interval between each test and retest scans was short relative to expected PCa progression and captures short-term acquisition-related variability.

This design further isolates intrinsic measurement variability of radiomic features while minimising confounding variables emanating from hardware, protocol, or observer differences. The analysis aims to provide a baseline evaluation for repeatability, which is a prerequisite for subsequent reliability assessment and downstream thresholding.

Statistical analysis of repeatability was performed using intraclass correlation coefficient (ICC), specifically ICC(3,1), corresponding to a two-way mixed-effects model with absolute agreement and single measurements. The model is appropriate for single-centre repeated acquisitions under identical imaging conditions and a fixed measurement system, as in the PRODIF cohort.

3.4.2 Repeatability pipeline in R

The workflow can be illustrated as follows:
1. Data preparation: For each patient, feature sets from test and retest scans were deduplicated and merged using patient identifiers (*case*).

2. Feature Selection: Relevant radiomic features for the statistical descriptors and the group were selected.

3. Data Cleaning: Rows with missing or infinite values were excluded, degeneracy checks ensured features with zero variance were not included.

4. ICC Computation: ICC(3,1) was computed for each feature pair using a two-way mixed-effects model with absolute agreement.

5. Interpretation: ICC values were classified as Poor (<0.50), Moderate (0.50–0.74), Good (0.75–0.89), or Excellent (≥0.90).

6. Aggregation: Results from multiple lesions per patient were combined to summarize repeatability across all lesions and modalities. Results from all subsets were aggregated into a single ICC summary table for feature-level comparison.

Pseudocode 1 – Illustrative R Implementation for Repeatability.

```
# Compute ICC for one feature between repeat scans
paired <- merge(df_rep1, df_rep2, by = "case")[, feature_col]
icc_res <- ICC(paired)
icc_value <- icc_res$results$ICC[icc_res$results$type == "ICC3"]
# Interpret ICC values
Interpretation <- case_when(
 ICC < 0.50 ~ "Poor",
 ICC < 0.75 ~ "Moderate",
 ICC < 0.90 ~ "Good",
 TRUE    ~ "Excellent"
)
# Combine results into summary table
icc_results_all <- bind_rows(icc_results_all, icc_block)
```

The R package “psych” was used for ICC calculations, ICC3 was selected to evaluate absolute agreement for repeated measures. Multiple lesions per patient were analysed independently and results are combined per feature. Interpretation is separated between “Poor”, “Moderate” “Good”. “Excellent”, where values <0.50 indicate poor reliability, 0.50–0.74 moderate, 0.75–0.89 good, and ≥0.90 excellent reliability.[170,188] All results are bound into a single table.

3.4.3 Reliability Study Design

While repeatability assesses measurement stability under ideal and controlled conditions, reliability evaluates the robustness of radiomic features when clinically relevant variation is introduced. Therefore, reliability is operationalised as consistent discriminative performance across features and analysis subsets. For a radiomic feature to function as a QIB, it must demonstrate performance consistency beyond

test-retest settings, reflecting conditions encountered in clinical and research workflows.

Histopathological grading followed the International Society of Urological Pathology (ISUP) 2014 Grade Group system.[11,71] For classification analyses, lesions were dichotomized at Grade Group 3, with Grade Group 1–2 coded as 0 (low grade) and Grade Group 3–5 coded as 1 (high grade). csPCa was therefore defined as ISUP Grade Group ≥ 3.

For each radiomic feature, an ROC curve was derived using the binarized Gleason outcome and corresponding feature values. Discriminative performance was quantified using the area under the ROC curve (AUC), with confidence intervals for the AUC calculated using the *ci.auc()* function. In addition, the Wilcoxon rank-sum test was performed to assess the statistical separation between the distributions in the two Gleason categories.

The area under the receiver operating characteristic curve (AUC) summarizes diagnostic discrimination. An AUC value of 0.5 indicates no discriminative ability, whereas values below 0.5 reflect performance worse than chance, typically due to inverse association between the predictor and the predefined positive class.[189] The AUC can be interpreted probabilistically as the likelihood that a randomly selected positive case is ranked higher than a randomly selected negative case; correspondingly, the complement (1 − AUC) represents the probability that negative cases are ranked above positive cases, consistent with reversed discrimination.[190] If a feature discriminated in the opposite direction (AUC < 0.5), its values were reversed to ensure consistent interpretation relative to the clinical endpoint.

3.4.4 Reliability pipeline in R

Reliability analysis was implemented in R using a structured AUC-based pipeline. The workflow can be summarised as follows:

1. Ground Truth Preparation: Histopathological GGG values were prepared and binarized (GGG < 3 = 0; GGG ≥ 3 = 1).

2. Data validation: Non-finite or missing feature values were excluded to ensure valid statistical analysis.

3. ROC Analysis: ROC curves were generated between the binarized Gleason Grade and individual radiomic features.

4. Confidence interval appraisal: Confidence interval of the AUC was calculated from the ROC curve.

5. Discriminative assessment: AUC values were extracted to quantify feature-level discrimination. If an AUC value is <0.5, they are reversed by using (1-AUC)

6. Statistical test: Wilcoxon rank-sum test was performed to analyse differences in feature distributions between a binarized ground truth.

7. Result Aggregation: AUC values, confidence intervals, Wilcoxon statistics, and p-values were summarized and combined into a unified results table. Subsequent feature analysis would also be aggregated into a single AUC summary document.

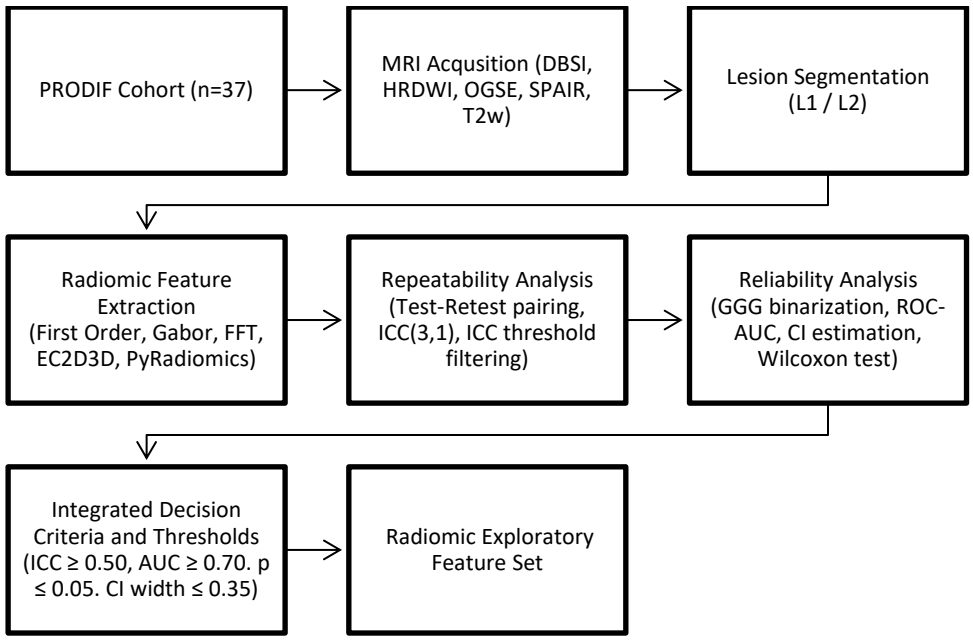
Pseudocode 2 – Illustrative R Implementation for Reliability

```
BINARISE ground truth for both lesions:
    IF GGG < 3 → class = 0
    ELSE → class = 1

  FOR each repeatable radiomic feature:
    COMPUTE ROC curve using feature as predictor
    CALCULATE AUC and confidence interval

    PERFORM Wilcoxon rank-sum test between class 0 and class 1
    STORE:
      - AUC
                    If AUC <0.30, 1-AUC
      - confidence interval
      - Wilcoxon statistic
      - p-value
      - feature identifier
      - modality / subset label
```

ROC and AUC analyses were performed using the “pROC” package in R, while Wilcoxon tests were implemented using base R functions. Multiple lesions per patient were analysed independently and results are combined at the feature level.

An overview of the complete radiomic processing and validation workflow is illustrated in Figure 3.1:

Fig 3.1 Exploratory Analysis Pipeline Diagram



3.5 Decision Criteria and Thresholds

Following repeatability and reliability analyses, both ICC-based repeatability and AUC-based reliability metrics were integrated into a unified decision framework. Feature selection was performed using predefined thresholds to ensure acceptable exploratory estimates. Radiomic features were retained if they satisfied all the following criteria: ICC ≥ 0.50, AUC ≥ 0.70, and corresponding ICC and Wilcoxon rank-sum p-values ≤ 0.05.

In addition to point estimates of discriminative performance, uncertainty was explicitly considered by constraining the width of the AUC confidence interval. Wide confidence intervals signal imprecise estimation and reduced robustness of feature performance, which is particularly relevant in moderate-size cohorts and low signal-to-noise diagnostic settings. The importance of precision in performance estimates has been highlighted in prediction modelling literature, where unreliable or highly uncertain estimates would undermine model interpretability and reliability in prediction modelling.[191]

Moreover, the Image Biomarker Standardisation initiative (IBSI) emphasizes that point estimates alone are insufficient, and that radiomic feature performance must be interpreted in conjunction with variability and uncertainty. Features exhibiting high

variability are considered unsuitable for downstream modelling, and robust technical assessment is essential before evaluating clinical utility.[50]

Consistent with this framework, Parmar et al demonstrate that most radiomic features are inherently unstable unless explicitly tested for repeatability and robustness, and that feature filtering must precede any modelling and performance-based claims.[192] Traverso et al[105] further illustrate that wide confidence intervals reflect poor precision, and that analyses based on small to moderately sized cohorts require stricter filtering due to the increased risk of overfitting.

As no universally accepted threshold exists for acceptable AUC confidence interval width in radiomics, it is for these reasons that a pragmatic upper limit width of ≤ 0.35 was selected to balance retention of informative features against the exclusion of unstable or weakly supported performance estimates.

The ICC threshold was selected to exclude features with poor repeatability according to the established ICC interpretation frameworks, while the AUC threshold ensured meaningful discriminative ability beyond chance-level classification. Statistical significance thresholds were applied to reduce the likelihood of spurious associations, and confidence interval constraints were used to prioritise features with stable and well-defined performance estimates.

Furthermore, because radiomic features were extracted separately from primary and secondary lesions when present, lesion-specific decision rules were applied:

- If Lesion 1 passed quality checks and Lesion 2 passed, both lesions were included.
- If Lesion 1 passed and Lesion 2 failed, only Lesion 1 was included.
- If Lesion 1 failed and Lesion 2 passed, the radiomic feature was excluded from analysis.
- If both lesions failed, the radiomic feature was excluded from analysis.

This approach ensured that the primary lesion was consistently represented, while restricting analysis to radiomic features derived from reliable imaging and stable feature estimates, thereby maintaining the integrity of the analysis.

3.6 QIB-Oriented Stringency Analysis

In the primary analytical framework (Section 3.5), an ICC ≥ 0.50 was used to identify radiomic features demonstrating at least moderate repeatability under test–retest conditions. This threshold was intentionally selected in the purpose of exploratory characterization of feature stability across modalities without excluding potentially informative candidates for analysis.

Nevertheless, QIBs require substantially stronger evidence of robust measurement repeatability and discriminative performance, thus QIBs are expected to demonstrate minimal measurement variability relative to biological variability. Accordingly, a secondary, QIB-oriented stringency analysis was conducted to establish radiomic features approaching biomarker-level robustness.

3.6.1 QIB-Strict Threshold Definition

Accordingly, the study was structured as a two-stage analytical framework: an exploratory characterization phase (Section 3.5) followed by a QIB-oriented validation phase applying stricter biomarker-level criteria.

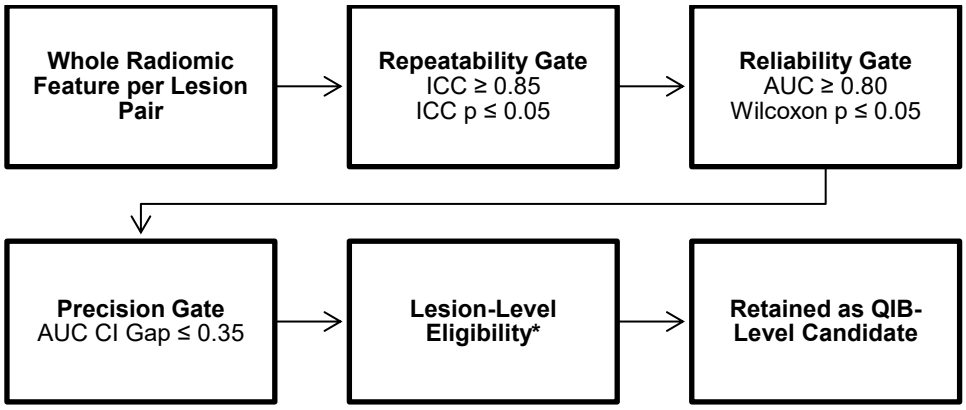
For QIB-oriented evaluation, repeatability was restricted to features demonstrating ICC ≥ 0.85, corresponding to the good-to-excellent reliability range according to Koo and Li.[170] This threshold reflects the level of measurement stability expected for quantitative biomarker applications in clinical imaging, as repeatability is a foundational trait given the risk that instability may compromise discriminative performance.

Reliability performance was restricted to AUC ≥ 0.80, representing strong classification performance that align with thresholds indicative of clinically meaningful diagnostic accuracy. Statistical robustness required is defined as: ICC p-value ≤ 0.05 and Wilcoxon rank-sum p-value ≤ 0.05. To ensure adequate precision, the AUC confidence interval width was required to be ≤ 0.35. Given the moderate cohort size (approximately

15–17 cases per class depending on modality), this threshold was selected to minimize the influence of sampling variability on discrimination estimates while remaining statistically feasible. Any feature that did not meet any of the criteria is excluded.

These stricter criteria were intentionally not applied in the initial exploratory analysis, as doing so would have obscured comparative modality-level repeatability patterns and limited broader radiomic characterization. The filtering was applied sequentially:

Fig 3.2 QIB Stringency Analysis Pipeline Diagram



*Eligibility Applied Identically to Section 3.5

Only features satisfying all criteria were considered QIB-level candidates under single-centre, small-sample conditions of this study.

3.7 Use of Generative AI Tools

AI-assisted coding support was also used to aid in drafting and troubleshooting selected programming scripts related to data preprocessing and workflow implementation. All study design, methodological decisions, statistical analyses, feature selection criteria, and interpretation of results were independently developed, executed, and verified by the author. AI tools were not used to generate results or to determine analytical conclusions. All code was reviewed, tested, and validated prior to use in the final analyses.

The author retains full responsibility for the scientific content and integrity of this work.

4 Results

4.1 Study Cohort and Data Availability

A total of 37 patients from the PRODIF study comprised the initial study cohort. All patients underwent preoperative MRI using a 3T Philips Ingenuity PET/MR scanner with a 32-channel cardiac coil. All patients subsequently received radical prostatectomy, with whole-mount histopathology serving as the reference standard and providing GGG classification.

The cohort consisted of 18 low-grade (GGG < 3) and 19 high-grade (GGG ≥ 3) cases, providing a balanced class distribution for ROC–AUC analysis.

After case matching between GGG and feature- and modality-based documentation, DBSI retained 30-31 cases per feature while OGSE retained 23-28 cases per feature. Feature availability varied due to non-computable radiomic features, missing feature values, unavailable paired test–retest scans, or mismatch between histopathological ground truth and the radiomic dataset. Paired test–retest acquisitions were available for DBSI- and OGSE-derived parameters.

Class distribution between low- and high-grade groups remained stable across all evaluated DBSI and OGSE subsets, with 15–17 cases per class for DBSI and 13–16 cases per class for OGSE. All evaluated subsets retained representation from both Gleason categories, permitting stable ROC–AUC estimation.

A total of 1,564 radiomic features were analysed. Of these, 10 (0.64%) were standalone first-order features, 13 (0.83%) were Gabor-based features, 90 (5.75%) were FFT2D-derived features, 392 (25.06%) were Edges- and Corners-based 2D/3D (EC2D3D) features, and 1,059 (67.71%) were PyRadiomics-derived features. The detailed distribution is presented in Table 4.1.

Table 4.1. Initial Radiomic Feature Space Composition by Feature Family

Feature family	Total distinct radiomic descriptor (n)	Percentage of total (%)
First-order (standalone)	10	0.64%

Feature family	Total distinct radiomic descriptor (n)	Percentage of total (%)
Gabor-filters	13	0.83%
FFT2D	90	5.75%
EC2D3D (Edges & Corners 2D/3D)	392	25.06%
PyRadiomics	1,059	67.71%
Total	1,564	100%

HRDWI and SPAIR imaging results were excluded due to insufficient paired test–retest observations (n = 1 per feature) respectively. Although T2-weighted imaging retained all 37 cases, no paired test–retest acquisitions were available. Consequently, T2-weighted parameters were excluded from repeatability-based evaluation. A minimum of three paired observations was required for ICC estimation. For ROC–AUC analysis, features were evaluated only when both Gleason categories were represented within the available subset.

Exclusions were applied at the feature- and modality-specific level rather than as a global patient-level exclusion, such that individual patients could contribute to analyses in modalities where valid data were available.

Study availability and inclusion are detailed in Table 4.2:

Table 4.2 Study Cohort and Modality-Specific Data Availability

Imaging modality	Initial cohort (n=37)	Repeatability analysis (paired test-retest ICC)	ROC–AUC n range (Low / High grade)	Included in final evaluation
DBSI (low-res)	37	30–31 per feature	30–31 (15–17 / 15–17)	Yes
OGSE	37	23–28 per feature	23–28 (13–16 / 13–16)	Yes
HRDWI	37	1 paired test-retest per feature	1 paired case per feature	No (insufficient sample)
SPAIR	37	1 paired test-retest per feature	1 paired case per feature	No (insufficient sample)
T2-weighted	37	No paired retest	Not evaluated (repeatability prerequisite not met)	No (repeatability unavailable)

Accordingly, effective sample sizes varied across modalities, lesions, and features and are reported alongside the corresponding statistical results.

4.2 Repeatability Results (ICC Analysis)

All radiomic features were evaluated for repeatability using ICC(3,1) and classified as Poor (<0.50), Moderate (0.50–0.74), Good (0.75–0.89), or Excellent (≥0.90).

A total of 53,176 feature–lesion observations were initially eligible for repeatability assessment. Prior to ICC estimation, observations were excluded if they contained non-finite values, fewer than three paired test–retest measurements or demonstrated degenerate distributions (defined as zero or near-zero variance across repeated measurements) which precludes stable ICC estimation.

Following this quality-control step, 52,636 observations remained for ICC(3,1) evaluation, corresponding to the removal of 540 observations (1.02%) due to degeneracy or insufficient variability. Degeneracy removal did not disproportionately affect any single modality.

4.2.1 Modality-Specific Repeatability

Repeatability results are presented in Table 4.3 for DBSI and Table 4.4 for OGSE.

Table 4.3. Distribution of ICC(3,1) Repeatability Classifications for DBSI Low-Resolution Parameters (%)

Derived Parameter	Poor (%)	Moderate (%)	Good (%)	Excellent (%)
Axial Diffusivity (AD)	35.59	16.59	15.98	31.83
Apparent Diffusion Coefficient (ADC)	35.60	16.71	18.28	29.40
Fiber Fraction Axial Diffusivity (F1AD)	55.83	13.62	9.26	21.28
Fiber Fraction Radial Diffusivity (F1RD)	52.40	13.97	10.06	23.56
Fractional Anisotropy (FA)	47.47	16.53	14.32	21.69
Fiber Ratio (FR)	47.14	14.24	12.37	26.25
Radial Diffusivity (RD)	34.95	17.49	17.07	30.49
Restricted Ratio (RR)	47.98	18.38	12.99	20.65
Water Ratio (WR)	33.39	15.94	18.93	31.75

Derived Parameter	Poor (%)	Moderate (%)	Good (%)	Excellent (%)
Modality Average[1]	43.37	15.94	14.36	26.32

[1]Modality averages represent the mean proportion of ICC classifications across all derived parameters.

Table 4.4. Distribution of ICC(3,1) Repeatability Classifications for OGSE Parameters (%)

Derived Parameter	Poor (%)	Moderate (%)	Good (%)	Excellent (%)
N1 ADC	79.09	6.95	2.28	11.68
N1 C	82.56	6.15	2.32	8.98
N2 ADC	73.16	10.80	4.02	12.02
N2 C	68.18	10.89	4.24	16.68
N4 ADC	74.24	7.78	5.60	12.38
N4 C	79.05	8.14	3.41	9.40
N6 ADC	74.85	8.75	4.54	11.87
N6 C	43.11	16.57	15.51	24.81
Modality Average[1]	71.78	9.50	5.24	13.48

[1]Modality averages represent the mean proportion of ICC classifications across all derived parameters.

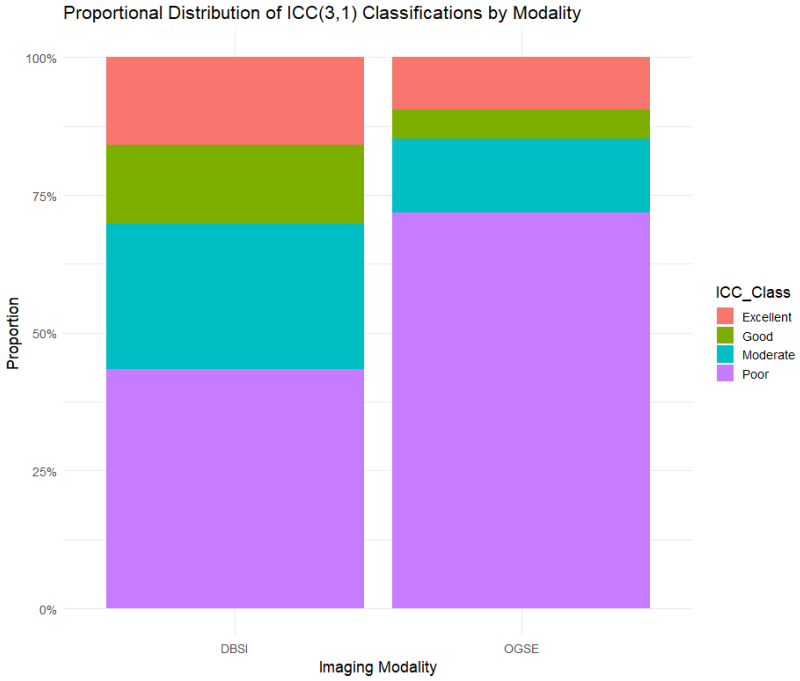
For DBSI low-resolution parameters, 56.62% of features achieved ICC ≥ 0.50, including 14.36% classified as “Good” (ICC 0.75–0.89), 26.32% classified as “Excellent” (ICC ≥ 0.90).

In contrast, OGSE-derived parameters demonstrated substantially lower repeatability. Only 28.22% of features achieved ICC ≥ 0.50, including 5.24% classified as “Good” and 13.48% classified as “Excellent”.

Among DBSI-derived parameters, Axial Diffusivity (AD) and Water Ratio (WR) exhibited the highest proportions of features classified as “Excellent” (31.83% and 31.75%, respectively). Within OGSE-derived parameters, N6 C demonstrated the highest proportion of “Excellent” classifications (24.81%). Radiomic feature families were extracted identically across DBSI and OGSE parameterisations, ensuring comparability of repeatability distributions between modality groups.

The distributional contrast between DBSI and OGSE repeatability profiles is further illustrated in Figure 4.1.

Fig 4.1. Distribution of ICC(3,1) Repeatability by Imaging Modality



The distribution illustrates the predominance of Poor repeatability in OGSE relative to DBSI, while DBSI demonstrates a rightward shift toward higher ICC classifications. Overall, DBSI-derived parameters demonstrated a more favourable repeatability distribution compared with OGSE-derived parameters.

4.3 Reliability Results (AUC Analysis)

A total of 53,176 distinct radiomic feature-modality-lesion pairings were evaluated separately for the primary and secondary lesions, resulting in a total of 106,352 ROC-AUC observations. Reliability assessment includes AUC estimates, corresponding confidence intervals and Wilcoxon rank-sum tests.

As repeatability was predefined as a prerequisite, only observations with ICC ≥ 0.50 were considered eligible for subsequent reliability evaluation. Moreover, both ICC and AUC filtering also follow the lesion filtering described in Section 3.5. If both lesions failed repeatability criteria, the corresponding radiomic feature was excluded from further reliability analysis.

The ICC-evaluable observations (n = 52,636) correspond to the total eligible feature–lesion pairings after degeneracy removal. Out of the total 52,636 ICC-evaluable feature–lesion observations, only 20,683 observations remained after ICC filtering of ≥ 0.50 and lesion filtering, which is a 39.50% retention.

ICC magnitude (≥0.50) was required prior to discrimination evaluation. Statistical significance of ICC ($p \leq 0.05$) was enforced in the final integrated filtering stage.

According to the predefined decision framework (Section 3.5), reliability is defined as AUC ≥ 0.70, and corresponding Wilcoxon rank-sum p-values ≤ 0.05 with an established confidence interval gap of 0.35. No feature with AUC ≤ 0.30 yielded an inverted AUC ≥ 0.70 that satisfied additional statistical and precision constraints.

Out of a total of 106,352 AUC observations, over 7,377 observations have an AUC ≥ 0.7 along with lesion filtering, or a retention rate of 6.93%. Intersection of the ICC-eligible and AUC-eligible subsets yielded 4,020 observations.

4.4 Integrated Repeatability and Reliability Filtering

The filtering framework was applied sequentially, beginning with repeatability eligibility, followed by discrimination thresholds, and finally statistical significance and precision constraints. This hierarchical structure ensured that only features demonstrating measurement stability, discriminative ability, and statistical robustness were retained.

A final integrated filtering step required an ICC p-value ≤ 0.05, Wilcoxon p-value ≤ 0.05, and AUC confidence interval width ≤ 0.35, a final retained set of 216 observations remained, corresponding to a 0.20% retention rate relative to the total 106,352 AUC observations. The low final retention rate reflects the high dimensionality of the radiomic feature space and the conservative nature of the predefined hierarchical filtering framework, rather than post-hoc exclusion.

Repeatability eligibility was enforced prior to discrimination filtering through modality- and statistic-level matching. As this approach may introduce duplicate lesion-level entries, feature-level distinction was subsequently resolved to prevent artificial inflation of retained counts.

Of the 216 radiomic observations meeting all predefined repeatability, discrimination, statistical significance, and precision criteria, 120 (55.6%) were derived from DBSI parameters and 96 (44.4%) from OGSE parameters. Within OGSE-derived features, the N6 C parameter accounted for the largest proportion of retained observations (n = 48; 22.2% of the total retained set), while within DBSI-derived features, low resolution DBSI with restricted ratio accounted the largest proportion in DBSI (n = 26; 12.03% of the total retained set).

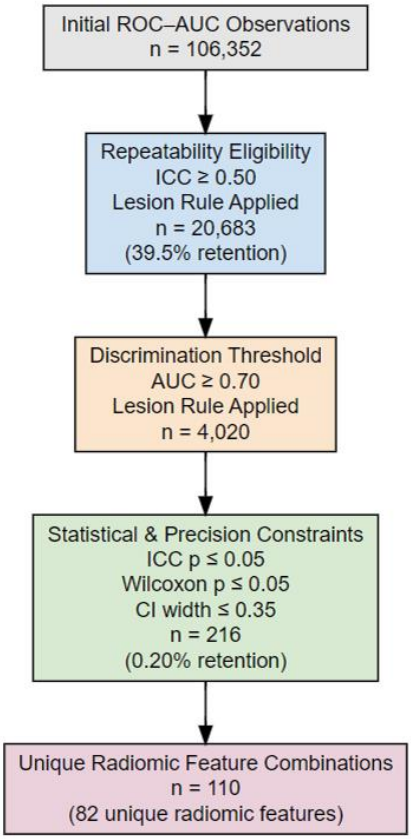
The 216 retained observations reflect lesion-level results. After collapsing duplicate primary and secondary lesion results, this corresponded to 110 distinct radiomic descriptor combinations that consist of 82 radiomic descriptors.

Benjamini–Hochberg false discovery rate (FDR) correction was applied to the Wilcoxon p-values within this filtered subset of 216 observations. All retained features remained statistically significant following correction (adjusted p-values 0.016–0.020). FDR correction was applied only within the final candidate subset to control for multiplicity among retained hypotheses, rather than across the entire feature space. This approach avoids diluting statistical power by correcting for hypotheses that were never eligible for discovery, and the sequential filtering criteria were defined a priori and were not optimized based on discrimination p-values.

Sequential filtering was applied beginning with repeatability eligibility (ICC ≥ 0.50 with lesion-specific decision rule), followed by discrimination thresholding (AUC ≥ 0.70), and concluding with statistical and precision constraints (ICC $p \leq 0.05$, Wilcoxon $p \leq 0.05$, AUC confidence interval width ≤ 0.35). From 106,352 initial ROC–AUC observations, 216 lesion-level observations (0.20%) satisfied all criteria, corresponding to 110 distinct radiomic feature-modality combinations and 82 distinct radiomic descriptor retained for final evaluation. The framework was intentionally conservative to reduce false-positive discovery in a high-dimensional radiomic setting. All thresholds were predefined prior to integrated filtering to avoid post-hoc optimization.

The statistical constraints applied in this framework (Fig 4.2) relate to significance and precision of individual feature performance estimates and do not represent redundancy or inter-feature correlation filtering.

Fig 4.2. Filtering flowchart of Reliability and Repeatability.



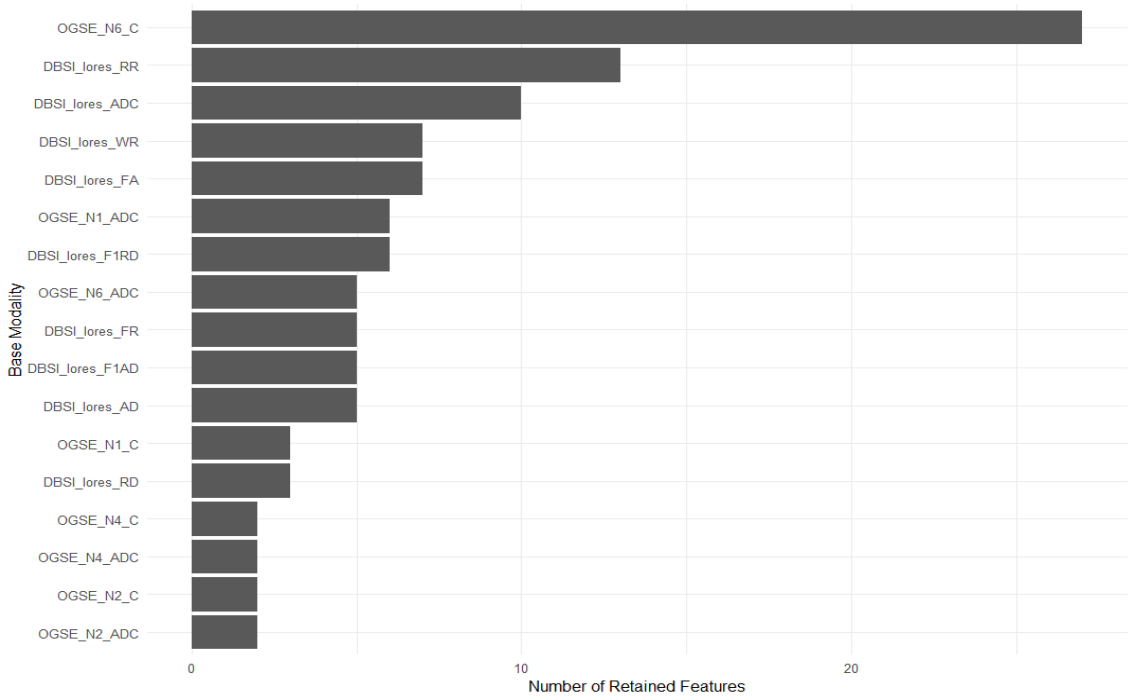
After collapsing duplicate lesion entries, 216 lesion-level observations corresponded to 110 distinct radiomic feature–modality combinations, representing 82 distinct radiomic descriptors.

Out of the 110 distinct radiomic descriptor combinations retained after integrated filtering, 61 (55.5%) were derived from DBSI parameters and 49 (44.5%) from OGSE parameters. The distribution of retained distinct radiomic feature-modality combinations by modality group is presented in Table 4.5. A further breakdown by individual base modality is illustrated in Figure 4.3.

Table 4.5 Retained distinct radiomic descriptor feature-modality combinations divided by modality group

Modality group	Distinct radiomic feature-modality combinations (n)	Percentage (%)
DBSI	61	55.5%
OGSE	49	44.5%
Total	110	100%

Fig 4.3. Number of retained distinct radiomic feature-modality combinations by base modality



Within OGSE-derived features, the N6 C parameter accounted for the largest proportion of retained observations (n = 27; 24.5% of the total retained set), while within DBSI-derived features, low resolution DBSI with restricted ratio accounted the largest proportion in DBSI (n = 13; 11.82% of the total retained set) along with low resolution DBSI with ADC (n = 10, 9.09% of the total retained set).

4.5 Lesion-Based Interpretation

Furthermore, of the 110 distinct radiomic descriptor combinations retained after integrated filtering, 98 (89.09%) were retained under the filtering framework in both primary and secondary lesions, indicating lesion-level robustness. Only 12 features (10.9%) were retained in a single lesion. This high concordance across lesion locations suggests substantial lesion-level robustness, supporting spatial robustness of retained features. A breakdown of lesion-level concordance could be emphasized in Table 4.6.

Table 4.6 Lesion-Level Concordance of retained distinct feature-modality combinations

Modality group	Retained in both lesions (n, %)	Single-lesion retention (n, %)
DBSI	57 (93.4%)	4 (6.6%)

Modality group	Retained in both lesions (n, %)	Single-lesion retention (n, %)
OGSE	41 (83.7%)	8 (16.3%)
Overall	98 (89.1%)	12 (10.9%)

Lesion-level concordance was higher for DBSI-derived features (93.4%) compared with OGSE-derived features (83.7%), indicating greater cross-lesion stability within DBSI parameters under the present acquisition and processing framework.

4.6 Feature Family Distribution

Of the 82 distinct radiomic descriptors retained after integrated filtering, 56 (68.3%) were derived from PyRadiomics and 26 (31.7%) from EC2D3D. Notably, no standalone first-order, Gabor, or FFT-derived features satisfied all predefined repeatability, discrimination, statistical significance, and precision criteria.

The enrichment ratio was calculated as the proportion of retained features within a feature family divided by that family's proportion in the initial feature pool. An enrichment ratio >1 indicates over-representation after filtering, whereas a ratio <1 indicates under-representation relative to baseline composition.

Table 4.7 Distribution of distinct radiomic descriptors within feature family by percentage.

Feature family	Distinct radiomic descriptor	Distinct radiomic feature-modality combinations	% Filtered	% Initial	Enrichment ratio
PyRadiomics	56	62	68.3%	67.71%	1.01
EC2D3D	26	48	31.7%	25.06%	1.27
First-order	0	0	0%	0.64%	0
Gabor	0	0	0%	0.83%	0
FFT2D	0	0	0%	5.75%	0
Total	82	110	100%	100%	--

When normalized to the initial feature space distribution, EC2D3D-derived features demonstrated a moderate enrichment ratio of 1.27, whereas PyRadiomics-derived features showed a proportional retention (enrichment ratio 1.01).

4.6.1 Edges- and Corners-Based Features

Within Edges- and Corners-based feature family (Table 4.8), Scharr-based object property descriptors constituted the largest proportion (n = 18, 37.5%), followed by Harris-Stephens corner detectors (n = 12, 25%) and Hessian-based object properties (n = 10, 20.8%). Frangi-based descriptors (n = 5, 10.4%) and Shi–Tomasi detectors (n = 3, 6.3%) comprised smaller fractions.

Table 4.8 Distribution of Edges- and Corners-based feature family subclass

Operator	Distinct radiomic feature-modality combinations (n)	Percentage of EC2D3D (%)
Scharr-based object properties	18	37.5%
Hessian-based object properties	10	20.8%
Frangi-based object properties	5	10.4%
Harris-Stephens corner detectors	12	25.0%
Shi–Tomasi detector	3	6.3%
Total	48	100%

Among these, the Scharr-derived object property *Ax2len_mean_mm* demonstrated the highest cross-modality retention. This descriptor appeared in 46 lesion-level observations across multiple DBSI and OGSE parameters, corresponding to 18 distinct radiomic feature-modality combinations. After collapsing duplicate primary and secondary lesion entries, this corresponded to a single structural descriptor repeatedly retained under varying diffusion parameterizations. The descriptor was retained across multiple DBSI and OGSE parameterizations, demonstrating consistent cross-modality robustness.

4.6.2 PyRadiomics Features

Within the PyRadiomics-derived subset (n = 62 distinct radiomic feature-modality combinations), second-order texture matrices predominated (Table 4.7). Gray Level Co-occurrence Matrix (GLCM) features accounted for the largest proportion (n = 23, 37.1%), followed by Gray Level Run Length Matrix (GLRLM) features (n = 12, 19.4%) and Gray

Level Size Zone Matrix (GLSZM) features (n = 9, 14.5%). Gray Level Dependence Matrix (GLDM) features comprised 11.3%, while wavelet-derived first-order features represented 8.1%. The remaining 9.7% consisted of other wavelet–texture combinations. This distribution indicates that higher-order spatial texture descriptors dominated the retained PyRadiomics feature space following repeatability and discrimination filtering.

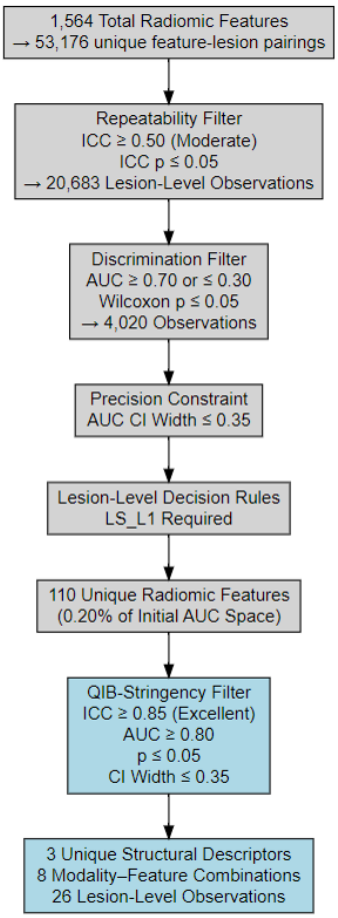
Table 4.9 Distribution of PyRadiomics feature family subclass

Texture matrix / category	Distinct radiomic feature-modality combinations (n)	Percentage of PyRadiomics (%)
GLCM (Gray Level Co-occurrence Matrix)	23	37.1%
GLRLM (Gray Level Run Length Matrix)	12	19.4%
GLSZM (Gray Level Size Zone Matrix)	9	14.5%
GLDM (Gray Level Dependence Matrix)	7	11.3%
First-order (wavelet-derived)	5	8.1%
Other wavelet-texture combinations	6	9.6%
Total	62	100%

4.7 QIB-Stringency Analysis

To evaluate biomarker-level robustness, a secondary QIB-oriented filtering framework was applied to the lesion-level observations retained after integrated repeatability and discrimination filtering (Section 4.4). This secondary analysis was conducted using predefined, stricter thresholds consistent with quantitative biomarker-oriented evaluation. Under this stricter framework, features were required to satisfy ICC ≥ 0.85, AUC ≥ 0.80 ICC p ≤ 0.05, Wilcoxon p ≤ 0.05, AUC confidence interval width ≤ 0.35 and Lesion-specific inclusion rules identical to Section 3.5. Only features meeting all criteria were considered QIB-level candidates under the single-centre test–retest conditions of this study. The sequential application of these thresholds is illustrated in Fig 4.4.

Fig 4.4 Sequential QIB-Oriented Thresholding Pipeline Applied After Integrated Filtering



Application of the QIB-stringency criteria resulted in:

- 26 lesion-level observations remained
- Corresponding to 8 distinct radiomic feature-modality combinations, utilizing 3 distinct radiomic descriptors.
- Representing 8 distinct feature–modality combinations (7.3% of 110) and 3.65% of the 82 distinct radiomic descriptors. This corresponds to a 92.7% reduction compared with the total possible combinations in the exploratory framework.

All retained QIB-level candidates were derived exclusively from Edges- and Corners-based 2D/3D operators. No PyRadiomics-derived radiomic features satisfied the QIB-stringency criteria.

4.7.1 Distribution of Retained QIB-Level Features

Two of the three retained features were Scharr-derived structural descriptors:

- Scharr Object Properties Axis 2 Length Mean (mm)
- Scharr Object Properties Relative Eccentricity

The Scharr Axis 2 Length Mean descriptor was retained across five OGSE parameterisations:

- OGSE N=1 (ADC and C)
- OGSE N=2 (ADC)
- OGSE N=4 (ADC and C)

The Scharr Relative Eccentricity descriptor was retained within OGSE N=6 parameterisations:

- OGSE N=6 ADC
- OGSE N=6 C

The remaining feature was a Hessian-derived structural descriptor:

- Hessian ($\sigma = 0.025$, threshold = 15.0) Object Properties Area Interquartile Range (mm^2)

This feature was retained within the low-resolution DBSI Restricted Ratio maps and was the only DBSI-derived feature meeting all QIB-stringency criteria.

4.7.2 Summary of QIB-Level Candidate Structure

Across all retained observations:

- Mean ICC values ranged from 0.90 to 1.00
- Mean AUC values ranged from 0.82 to 0.84
- AUC confidence interval widths ranged from 0.32 to 0.34

The QIB-stringency analysis therefore isolated a small subset of structural object-property descriptors demonstrating concurrent high repeatability, strong discriminative performance, statistical significance, and acceptable precision under the applied test–retest framework.

4.7.3 Distribution of QIB-Level Candidates

The QIB-stringency analysis yielded 3 distinct radiomic descriptors, observed across 8 distinct modality–feature combinations and 26 lesion-level observations. This pattern indicates cross-parameter robustness of specific structural radiomic properties rather than isolated modality-specific effects.

Table 4.10. QIB-Level Radiomic Candidates After Stringent Sequential Filtering, Ranked by Mean AUC.

Imaging parameter	Distinct radiomic descriptor	Lesion-level observations	Mean ICC	Mean AUC	CI gap
OGSE (N=6) – ADC	Scharr Edge-Based Relative Eccentricity	2	1.00	0.84	0.33
OGSE (N=6) – C	Scharr Edge-Based Relative Eccentricity	2	1.00	0.84	0.33
DBSI (Low-Resolution) – Restricted Ratio	Hessian-Derived Structural Area Variability (IQR)	2	0.90	0.83	0.33
OGSE (N=2) – ADC	Scharr Edge-Based Minor Axis Length (Mean)	2	1.00	0.83	0.32
OGSE (N=1) – ADC	Scharr Edge-Based Minor Axis Length (Mean)	2	1.00	0.83	0.32
OGSE (N=4) – ADC	Scharr Edge-Based Minor Axis Length (Mean)	2	1.00	0.82	0.34
OGSE (N=4) – C	Scharr Edge-Based Minor Axis Length (Mean)	2	1.00	0.82	0.34
OGSE (N=1) – C	Scharr Edge-Based Minor Axis Length (Mean)	2	1.00	0.82	0.33

Scharr Edge-Based Relative Eccentricity demonstrated the highest mean AUC (0.84) within OGSE N=6 ADC and OGSE N=6 C parameterisations, while maintaining ICC values of 1.00 under test–retest conditions. Furthermore, Scharr Edge-Based Minor Axis Length (Mean) exhibited consistent retention across five distinct OGSE parameterisations, indicating reproducible structural sensitivity across varying diffusion encoding conditions.

The only DBSI-derived QIB-level candidate was the Hessian-Derived Structural Area Variability (Interquartile Range), retained within low-resolution restricted ratio maps, with ICC values between 0.88 and 0.93 and mean AUC of 0.83.

4.8 Summary of Key Findings

106,352 lesion-level ROC-AUC observations were derived within the hierarchical filtering framework. After degeneracy removal and repeatability preselection (ICC ≥ 0.50), 20,683 observations remained eligible for discrimination analysis. Application of predefined AUC threshold (≥ 0.70), lesion-specific inclusion rules, and precision criteria further reduced the candidate set to 4,020 observations.

Final integrated filtering required statistical significance for both repeatability (ICC $p \leq 0.05$) and discrimination (Wilcoxon $p \leq 0.05$), together with an AUC confidence interval width ≤ 0.35. This resulted in 216 lesion-level observations (0.20% of the initial AUC space), corresponding to 110 distinct radiomic feature-modality combinations.

After collapsing duplicate lesion entries, 216 lesion-level observations corresponded to 110 distinct radiomic feature–modality combinations, representing 82 distinct radiomic descriptors.

At the feature-modality combination level (n=110)

- 62 (56.4%) were derived from PyRadiomics
- 48 (43.6%) were derived from Edges- and Corners-based 2D/3D operators (EC2D3D)

At the descriptor level (n=82)

- 56 (68.3%) were derived from PyRadiomics
- 26 (31.7%) were derived from Edges- and Corners-based 2D/3D operators (EC2D3D)

No standalone first-order, Gabor, or FFT-derived features satisfied all predefined criteria.

Lesion-level concordance analysis demonstrated 98 of 110 distinct radiomic feature-modality combinations (89.1%) were independently retained in both primary and secondary lesions. Concordance was higher for DBSI-derived features (93.4%) compared with OGSE-derived features (83.7%), indicating stronger cross-lesion stability within DBSI parameterisations.

Within OGSE-derived features, the N6 C parameter accounted for the largest share of retained distinct radiomic descriptor (n = 27; 24.5% of the total retained set), reflecting concentration of robust features within specific OGSE parameterisations rather than uniform distribution across OGSE metrics.

Within Edges- and Corners-based 2D/3D features, Scharr-based object property descriptors predominated, with the structural descriptor Ax2len_mean_mm demonstrating the broadest cross-modality retention (46 lesion-level observations across 14 base modalities). Within the PyRadiomics subset, higher-order spatial texture descriptors, particularly GLCM-based metrics, constituted the largest proportion of retained features.

Following QIB-stringency analysis, 26 lesion-level observations remained, corresponding to 3 distinct radiomic descriptors and 8 distinct radiomic feature-modality combinations that comprises modality-feature combinations, representing a 92.7% reduction relative to the 110 retained feature–modality combinations in the exploratory thresholding framework.

Ranked by mean AUC, Scharr Edge-Based Relative Eccentricity within OGSE N=6 C and ADC parameterisations demonstrated the highest discriminative performance. The only DBSI-derived QIB-level candidate was Hessian-Derived Structural Area Variability (IQR).

Generally, the hierarchical filtering framework substantially reduced the high-dimensional radiomic feature space while preserving a subset of features demonstrating consistent repeatability, discriminative ability, statistical robustness, and lesion-level stability.

5 Discussion

5.1 Modality-Specific Repeatability Dynamics

Repeatability varied meaningfully across imaging modalities. DBSI demonstrated higher lesion-level concordance and a greater proportion of parameters within the good-to-excellent ICC range compared with OGSE. While ICC-based preselection has been widely advocated in radiomics to mitigate instability in high dimensional feature spaces, reported thresholds and sequencing strategies remain heterogenous across the literature.[105,165,166] In the present framework, repeatability was enforced prior to discrimination, thereby positioning measurement stability as a prerequisite for performance evaluation.

The findings of this thesis suggest modality-dependent differences in how diffusion parameterisation influences downstream radiomic stability. As within DBSI-derived parameters, Axial Diffusivity (AD) and Water Ratio (WR) exhibited the highest proportions of features classified as “Excellent” (31.83% and 31.75%, respectively). Among OGSE-derived parameters, the N6 configuration demonstrated the highest proportion of “Excellent” classifications (24.81%).

csPCa is characterised by increased cellular density and restricted diffusion, and diffusion-weighted imaging remains central to PI-RADS v2.1 assessment, underscoring its clinical importance.[110,112,193] The broader repeatability observed in DBSI-derived parameters may relate to its multi-compartment modelling framework, which decomposes intra-voxel diffusion signals into anisotropic and isotropic components and may thereby stabilise diffusion metrics under test–retest conditions.[99,123,124,194] However, this interpretation remains inferential and was not directly tested within the present study.

Although DBSI exhibited broader repeatability retention, OGSE yielded fewer retained features overall but proportionally more discriminatively robust structural descriptors under QIB-oriented thresholds. By probing diffusion dynamics across defined frequency regimes, OGSE modulates sensitivity to diffusion restriction and may enhance contrast related to fine-scale tissue architecture.[119,195] Diffusion techniques differ not only in

acquisition design but also in the microstructural length scales they interrogate, which may contribute to differences in radiomic repeatability and stability profiles.

The shorter effective diffusion times achievable with OGSE may increase sensitivity to subtle microstructural variation, potentially contributing to the retention of specific structural descriptors with higher discriminative performance under QIB-oriented filtering. In the present study, OGSE parameterisations were disproportionately represented within the QIB-level candidate subset. Nevertheless, given the single-centre design and moderate sample size, these observations require independent validation before broader generalization. The proposed mechanistic interpretation remains inferential and was not directly tested within this study.

The balance between repeatability and discrimination is central to QIB development. In high-dimensional radiomics, repeatability functions as a statistical gatekeeper, as features that fail to reproduce under controlled conditions risk inflating results through noise-driven variance. High repeatability does not ensure robust discrimination, and conversely, high AUC without stability may reflect variance-driven separation rather than true biological signal.[163,181,187] Importantly, intraclass correlation coefficient (ICC) estimates are influenced by inter-subject variability, such that elevated ICC values may reflect lesion heterogeneity rather than measurement precision.[170–173] Altogether, these modality-specific repeatability dynamics contribute mechanistic context for the redistribution of radiomic feature families under stability-oriented filtering.[50]

5.2 Structural Versus Intensity-Derived Radiomic Paradigms

The initial feature pool comprised 1,564 descriptors: 10 first order (0.64%), 13 Gabor (0.83%), 90 FFT2D (5.75%), 392 EC2D3D (25.06%), and 1,059 PyRadiomics (67.71%). After exploratory thresholding, 82 distinct descriptors were retained (56 PyRadiomics; 26 EC2D3D), corresponding to 110 feature–modality combinations (62 PyRadiomics; 48 EC2D3D).

Implementation of QIB-oriented thresholds further reduced the candidate set to 8 distinct feature–modality combinations representing 3 distinct radiomic features. Of which, seven were Scharr-based features paired with OGSE parameterizations (87.5%),

and one was a Hessian-derived feature paired with DBSI Restricted Ratio maps. Notably, no PyRadiomics-derived features satisfied biomarker-level criteria under the imposed repeatability and precision constraints.

5.2.1 Stability and Limitations of Standardized Texture Matrices (PyRadiomics)

The proportional retention of PyRadiomics features likely resonates its standardized and extensive design, encompassing first-order, shape, and higher-order texture matrices (GLCM, GLRLM, GLSZM, GLDM).[162] Within the retained subset, GLCM features predominated (37.1%), consistent with their dense representation within the PyRadiomics feature pool. One possible justification is that GLCM quantifies gray-level co-occurrence at specific spatial offsets, quantifying second-order spatial dependencies between discretized voxel intensities.[142,143,196] Comparable matrix-based features such as: GLRLM and GLSZM extends this convention to run-length and connected zone configurations, respectively.[144,145]

In spite of numerical prominence during exploratory filtering, PyRadiomics-derived texture features did not carry over under QIB-level constraints. Considering these features depend on discretized voxel intensity relationships and predefined neighbourhood configurations, PyRadiomics features may be sensitive to small spatial perturbations introduced by repositioning or minor acquisition variability.[50,105,165] Such sensitivity could alter matrix composition prior to feature derivation, resulting in insufficient stability under stricter ICC (≥0.85) and precision requirements.

5.2.2 Robustness of Gradient-Derived Structural Descriptors (EC2D3D)

Although fewer EC2D3D descriptors passed exploratory thresholds compared with PyRadiomics, EC2D3D was the only family retained under QIB-level criteria. This shift suggests that structural descriptors tolerated escalating stability constraints more effectively than intensity-derived matrices.

EC2D3D operators are gradient-derived structural filters that derive geometry-oriented properties, such as axis length, eccentricity, object count, from edge-enhanced representations. Within the initial exploratory framework, Scharr-based object properties demonstrated the greatest proportional survival within EC2D3D, seen in 46

lesion-level observations across 14 base modalities and representing 18 modality-feature combinations (37.5% of EC2D3D).

The Scharr operator computes first-order spatial derivatives optimized for rotational symmetry and accurate gradient magnitude estimation.[141,155] When combined with object-property extraction, local intensity transitions are translated into higher-order geometric descriptors. This process reduces reliance on absolute gray-level distributions and instead encodes boundary-defined morphology.

Under repositioning-induced perturbation, small voxel-level intensity shifts may meaningfully alter co-occurrence statistics and other texture matrices. In contrast, macroscopic lesion boundaries are less likely to change substantially across short-interval repeat acquisition. This abstraction from raw intensity toward structural geometry may partly explain the preferential survival of edge-derived morphometric descriptors. Previous work has shown that many radiomic features are sensitive to acquisition and preprocessing variability, particularly those dependent on discretized intensity relationships.[105,154,192] Gradient-based operators emphasize spatial transitions rather than absolute intensities and may therefore provide greater resilience to minor acquisition-related perturbations.

Following the predefined QIB-stringency thresholding pathway, two Scharr-derived structural descriptors satisfied biomarker-level candidacy criteria. Scharr Edge-Based Relative Eccentricity displayed the highest mean AUC across retained candidates, whereas Scharr Edge-Based Minor Axis Length (Mean) showcased consistent retention across five distinct OGSE parameterisations.

Both features are sourced from ellipse fitting applied to edge-defined components within the region of interest. Minor axis length quantifies structural compactness that reflect the average short-axis dimension of detected components, while relative eccentricity summarizes architectural elongation as a function of the major-to-minor axis ratio, quantifying higher values that indicate major elongated structures and lower values reflecting more circular geometry.

In addition to the Scharr-based descriptors, a single Hessian-derived structural operator satisfied QIB-stringency criteria. Hessian operators compute second-order spatial derivatives and enhance curvature- and blob-like structures within an image. Hessian-derived Area Interquartile Range quantifies dispersion in object area following curvature-based filtering. Notably, low-resolution DBSI RR was the only DBSI parameterization associated with this retained Hessian descriptor. This suggests that curvature-based structural abstraction, when applied to compartmentalized diffusion maps, may provide a stable representation of spatial heterogeneity under short-interval acquisition variability.

Restricted Ratio (RR) in DBSI represents the proportion of diffusion signal attributed to restricted isotropic components, which are commonly associated with increased cellular density and reduced extracellular space. Structural descriptors derived from RR maps may therefore summarize the geometric organization of high-restriction domains rather than their absolute intensity magnitude. However, direct biological correlation was not evaluated in this study.

Likewise, these descriptors translate boundary-defined lesion morphology into compact quantitative measures. Their survival under stringent repeatability and precision constraints suggests that edge-defined structural geometry may provide a stable representation of lesion organization under short-interval acquisition variability.

More broadly, abstraction from voxel-level intensity toward object-level morphology appears to buffer the effects of minor perturbations introduced by repositioning. While this interpretation remains inferential, the preferential survival of curvature- and edge-derived descriptors under QIB-level constraints supports the notion that geometric representation may offer greater repeatability than intensity-dependent metrics.

5.2.3 Stability-First Filtering Reshapes the Radiomic Feature Landscape

When repeatability (ICC ≥ 0.85) and precision constraints were enforced prior to discrimination, gradient-derived structural and object-based descriptors consistently outperformed intensity-based texture matrices in terms of robustness. This aligns with

prior radiomics literature demonstrating that many features are sensitive to acquisition variability unless explicitly filtered for stability.[105,165,192]

The repositioning design functioned as a geometric stress test. Under such perturbations, descriptors derived from lesion boundaries showed greater resilience relative to intensity-dependent matrices. Robust radiomic performance was not uniformly distributed across modality parameterizations, implying biomarker-oriented radiomic analysis may require modality-specific optimization rather than uniform feature transfer across sequences.

Although OGSE exhibited lower overall repeatability, certain radiomic features appeared to enhance boundary contrasts that favour gradient-based structural descriptors. These retained radiomic descriptors had quantified lesion structural geometry extracted from edge-enhanced maps, these radiomics may provide greater measurement stability under repositioning while preserving discriminative information.

By and large, these findings footholds the central hypothesis, that applying repeatability metrics and statistical precision prior to reliability analysis alters the composition of retained radiomic candidates. This shift highlights the importance of stability-oriented preselection in QIB development.[50,166] emphasizing the fallacy of discrimination-only biomarker pipelines that may overlook structurally stable candidates that emerge only under stability-oriented filtering frameworks.

5.3 Cross-Parameter Robustness and Structural Recurrence Under QIB-Level Constraints

Amongst the 14 evaluated base modalities, Scharr Edge-Based Minor Axis Length satisfied QIB-level criteria across five distinct OGSE settings, such patterns may indicate cross-modality robustness instead of parameter-specific performance.

OGSE modulates effective diffusion time and thereby alters sensitivity to microstructural length scales.[119–121] Within the present acquisition framework, specific OGSE parameterizations may have enhanced lesion boundary contrast in a manner favourable to gradient-derived structural descriptors.

Other structural operators, including Hessian- and Frangi-based filters, did not demonstrate comparable cross-parameter recurrence within OGSE settings, despite sharing similar parametric outputs such as Minor Axis Length. Scharr emphasizes first-order intensity transitions and boundary contours,[155] whereas Hessian operators enhance curvature-defined, blob-like regions, and Frangi filters preferentially detect tubular or line-like structures through eigenvalue analysis of the Hessian matrix.[156–158] The preferential retention of Scharr-derived descriptors within OGSE may therefore reflect compatibility between diffusion-enhanced boundary contrast and gradient-based boundary abstraction, rather than curvature- or vessel-oriented feature detection.

In contrast, no intensity-based texture matrices from PyRadiomics, FFT2D, Gabor filters, or standalone first-order features demonstrated comparable cross-parameter persistence under QIB constraints. Texture matrices rely on discretized gray-level relationships and predefined neighbourhood configurations,[50,143] which may be more sensitive to minor spatial perturbations introduced in a test–reposition–retest design.

Cross-parameter recurrence therefore represents an additional dimension of robustness beyond repeatability and discrimination alone. Descriptors that remain stable across independent parameterisations may better approximate parameter-robust structural properties, an important consideration in QIB development.[105,166]

Although OGSE parameterizations are not statistically independent, recurrence across diffusion-time variations suggests relative stability under parameter modulation rather than strict independence.

In contrast, the only DBSI-derived QIB-level candidate (Hessian Area Interquartile Range on Restricted Ratio maps) did not demonstrate comparable multi-parameter recurrence. Hessian Area IQR reflects dispersion in the size of curvature defined components following segmentation. DBSI Restricted Ratio maps represent the proportion of isotropic restricted diffusion, often associated with densely packed microdomains. It is therefore plausible that spatially coherent restricted-diffusion clusters of restricted diffusion align more closely with curvature-sensitive detection than with boundary-gradient enhancement.

These mechanistic interpretations remain hypothesis-generating, as direct diffusion-time sensitivity analyses and histopathological correlation were not performed in the present study.

5.4 Methodological Implications for QIB-Oriented Radiomics

In high-dimensional radiomics, the order and strictness of filtering criteria fundamentally determine which features are ultimately retained and how the candidate space is structurally composed. In this thesis, retained radiomic descriptors and feature–modality combinations were derived using a repeatability-first strategy, followed by discrimination and precision constraints. This approach aligns with recommendations in QIB-candidacy research emphasizing technical performance before clinical inference.[48,166]

The underlying principle of this framework is that reproducibility must precede discrimination, because discriminative reliability without measurement stability risks inflating false-positive findings and limiting generalisability.[105,181] At the same time, contrariwise, a highly repeatable descriptor that lacks discriminative power may still have utility in quantitative longitudinal characterization, depending on clinical context.

Two analytical pipelines were implemented with distinct aims. The first provided exploratory characterization of modality- and feature-level behaviour. The second applied stricter thresholds to identify QIB-level candidates within the PRODIF dataset.

Under exploratory thresholds (ICC ≥ 0.50; AUC ≥ 0.70), PyRadiomics-derived texture matrices accounted for the majority of retained feature–modality combinations. Yet, when thresholds were escalated to QIB-oriented levels (ICC ≥ 0.85; AUC ≥ 0.80), geometry-derived EC2D3D descriptors became predominant.

The observed shift in feature-family composition after threshold escalation indicates that biomarker candidacy is not an intrinsic property of a radiomic descriptor alone but depends on the evaluation framework applied. These findings suggest that performance-first approaches may disproportionately over-represent intensity-dependent texture matrices, whereas stability-first frameworks preferentially preserve geometry-derived descriptors.

The additional requirement of an AUC confidence interval width ≤ 0.35 further influenced feature survival by enforcing a minimum precision standard. Confidence interval width reflects estimation uncertainty, which is particularly sensitive to moderate cohort sizes. By constraining precision, the framework reduces the likelihood of retaining features whose apparent performance is driven by sampling variability rather than stable signal, which echo the recommendations underpinned in quantitative imaging and prediction modelling literature.[50,105,191,192]

Everything considered, the results indicate that biomarker candidacy in radiomics emerges from methodological design choices as much as from intrinsic feature properties. Repeatability-oriented and reliability-aware filtering therefore plays a central role in QIB development.

5.5 Conceptual and Translational Implications for Quantitative Imaging Biomarkers

5.5.1 Repeatability vs Reliability: A Necessary Tension

The present framework explicitly distinguishes between measurement stability and discriminative performance through the concurrent evaluation of intraclass correlation (ICC) and ROC–AUC. These metrics quantify fundamentally different properties of a radiomic feature–modality combination and may not co-vary under high-dimensional screening.[170,175]

Without repeatability constraints, elevated AUC values may reflect variance-driven separation, sampling fluctuation, or acquisition-dependent signal instability.[163,181] Inversely, a feature that is repeatable yet lacks discriminatory power offers limited value for diagnostic stratification.[171,173]

The harmony between repeatability and reliability becomes evident across successive stages of the filtering hierarchy. When discrimination was considered in isolation, a substantial subset of feature-modality combinations appeared promising. However, thresholding synthesis of repeatability, statistical significance, and precision criteria, led to a 0.20% reduction from the initial observations, corresponding to 110 distinct feature–modality combinations and 82 descriptors. Further escalation to QIB-level

thresholds contracted this set to 8 feature–modality combinations representing 3 structural descriptors.

This progressive contraction illustrates that apparent discriminative performance is not synonymous with biomarker robustness.

QIB development therefore requires simultaneous optimization of repeatability and reliability thresholding.[48,166] The hierarchical filtering strategy enforced repeatability prior to discriminative reliability and subsequently integrating statistical and precision constraints. Comparison of exploratory and QIB-level thresholds concluded that feature candidacy is sensitive to the balance imposed between ICC and AUC criteria.

Robust QIB identification rests upon deliberate calibration of these axes within a predefined methodological structure, which reduces the probability that variance-driven artifacts are advanced as candidate biomarkers.[105,181]

5.5.2 Precision as a Structural Filter in Radiomics

Although no universal consensus exists regarding acceptable confidence interval width thresholds in radiomics, reliance on point estimates alone is statistically insufficient for evaluating feature robustness. Both AUC and ICC are sample-dependent statistics, and their uncertainty increases as effective sample size decreases.[170,175] In moderately sized cohorts, numerically high AUC values may be accompanied by wide confidence intervals, reflecting substantial estimation variability. Without explicit precision constraints, such features risk being retained as candidate QIBs based on unstable performance estimates.

The initial cohort of this study comprised of 37 patients, with modality-specific effective sample sizes ranging from 23 to 31 per feature and 13 to 17 cases per Gleason category. Under these conditions, ROC–AUC variance is non-trivial, and statistical significance alone does not guarantee precision.[179]

By defining a confidence interval width threshold (≤ 0.35) as a precision prerequisite into the hierarchical filtering framework, it critically affects magnitude size and penalizes imprecision rather than inadequate discrimination capabilities.

Precision control is particularly relevant in high-dimensional radiomic settings, where thousands of candidate features are evaluated simultaneously. In such settings, sampling variability alone may generate spuriously elevated AUC values, especially in moderate sample sizes.[163,181] Incorporating confidence interval filtering reduces the probability that stochastic variance is misinterpreted as robust discriminative signal.

Within a QIB-oriented framework emphasizing measurement reliability and reproducibility, explicit precision constraints are methodologically aligned with biomarker qualification principles.[48,166] Although standardized CI thresholds have yet to be established in radiomics, the present findings demonstrate that CI width meaningfully contributed to dimensional reduction and strengthens the robustness profile of retained candidates.

The absence of formal precision control in many radiomic studies may contribute to retention of features with unstable performance estimates to persist in candidate sets. These findings underscore the potential value of incorporating CI-based filtering as a complementary safeguard in biomarker-oriented analyses.[50,105]

5.5.3 Abstraction Hierarchy in Radiomic Stability

Following QIB-stringency analysis, only edges- and corners-based 2D/3D structure-derived radiomic descriptors satisfied integrated repeatability, discrimination, statistical significance, and precision criteria under the present test–reposition–retest framework. Intensity-proximal descriptors, including standalone first-order statistics, Gabor-derived features, FFT2D features, and higher-order texture matrices from PyRadiomics, did not persist under biomarker-level constraints.

This pattern suggests a form of abstraction hierarchy in radiomic stability. Radiomic features differ in how closely they operate to raw voxel intensities and in the extent of transformation applied before feature computation.

First-order statistics directly summarize voxel intensity distributions within the region of interest. Texture matrices such as GLCM quantify discretized gray-level relationships between neighbouring voxels by estimating how frequently intensity level *i* occurs adjacent to level *j* at predefined spatial.[50,143] Derived measures such as contrast,

homogeneity, correlation, and entropy are therefore functions of voxel-level adjacency distributions. Because these matrices depend on discretized intensities and spatial configuration, small repositioning-related voxel shifts may alter gray-level bin assignments and adjacency counts, redistributing probability mass within the matrix. In a test–retest setting, such proximity to voxel-level signal may increase sensitivity to acquisition variability.[105,165]

In contrast, edges- and Corners-based 2D/3D operators apply sequential transformations prior to descriptor extraction. Gradient enhancement, object segmentation, and morphometric summarization, including axis length, eccentricity, and area variability, introduce higher-order geometric abstraction. The resulting descriptors quantify object-level structural organization rather than direct voxel-intensity relationships. This abstraction reduces dependence on individual voxel intensity assignments and may buffer minor spatial perturbations introduced by repositioning.

The preferential survival of geometry-derived feature–modality combinations under QIB-level thresholds supports the hypothesis that increasing structural abstraction is associated with enhanced stability in high-dimensional radiomic screening. Within the present diffusion-based mpMRI framework, descriptors derived from boundary-enhanced representations demonstrated greater resilience to repeatability and precision constraints than intensity-dependent features.

These observations do not imply a universal hierarchy of feature robustness. Rather, they suggest that level of structural abstraction represents a meaningful dimension when radiomic features are evaluated under stability-oriented constraints. The apparent resilience of geometry-derived descriptors in this single-centre test–reposition–retest setting should therefore be interpreted as context-specific rather than definitive.

Independent multi-centre validation will be necessary to determine whether abstraction-related stability persists across different scanners, acquisition protocols, and institutional workflows. Only through such external validation can the generalisability of these findings be established within broader QIB development.

5.5.4 Cross-Parameter Recurrence as an Additional Robustness Dimension

Beyond repeatability and discrimination, the present analysis revealed an additional pattern of robustness: cross-parameter recurrence. Following hierarchical filtering, several radiomic descriptors were retained across multiple diffusion parameterizations rather than being confined to a single acquisition setting.

Within the exploratory framework, a number of descriptors appeared in multiple different feature–modality combinations, indicating that their performance was not restricted to a single diffusion encoding. This pattern was most apparent in the Scharr-based EC2D3D subclass, where the same structural descriptor was repeatedly retained across distinct parameterizations.

Under QIB-stringency criteria, this recurrence became more selective but remained observable. Scharr Edge-Based Relative Eccentricity was retained within both OGSE N=6 ADC and OGSE N=6 C maps, while Scharr Edge-Based Minor Axis Length persisted across five OGSE parameterizations. In contrast, the only DBSI-derived QIB candidate (Hessian Area Interquartile Range on Restricted Ratio maps) was confined to a single feature–modality pairing.

The asymmetry in recurrence patterns suggests that robustness may operate at multiple hierarchical levels:

- Lesion level (consistency across primary and secondary lesions)
- Parameter level (persistence across diffusion-time settings)
- Modality level (retention within DBSI or OGSE groups)
- Descriptor level (repetition of the same geometric construct under varying encodings)

Cross-parameter recurrence may therefore represent an additional dimension of stability beyond repeatability alone. A descriptor that remains viable under varying diffusion-time sensitivities may approximate a parameter-robust structural property, rather than reflecting optimization to a specific acquisition setting.

In high-dimensional radiomics, where thousands of candidate features are evaluated, such recurrence patterns may help distinguish structurally meaningful descriptors from acquisition-dependent performance artifacts. Features confined to a single parameterization may reflect acquisition-specific contrast behaviour specific to that setting, whereas recurrent descriptors may capture geometric properties that persist despite modulation of diffusion encoding.

Although direct biological validation was not undertaken, the repeated retention of gradient-derived structural descriptors across OGSE parameterizations supports the interpretation that certain geometric properties may exhibit relative invariance to diffusion-time modulation. From a QIB development perspective, such invariance is desirable, as robustness across acquisition conditions strengthens translational credibility.

5.5.5 Implications for Future Radiomic Framework Design

Enforcing repeatability as a prerequisite to discrimination appears to mitigate variance-driven performance inflation in high-dimensional feature space. Noting the differences in exploratory and QIB-oriented constraints, the shift demonstrates that the architecture of the analytical framework directly influences which feature families emerge as candidates for QIB development.

Incorporation of explicit precision constraints introduced an additional safeguard against sampling variability. Confidence interval filtering materially contributed to dimensional contraction and strengthened the robustness profile of retained descriptors, suggesting that future radiomic frameworks may benefit from integrating precision control as a standard component of feature evaluation.

Moreover, observed abstraction pattern implies that object-level geometric descriptors may be more resilient to acquisition variability than voxel-level intensity statistics. Frameworks that explicitly account for abstraction level during feature evaluation may therefore enhance stability-oriented selection processes and improve alignment with QIB development principles.

Cross-parameter recurrence emerged as an additional dimension of robustness. Descriptors that persisted across multiple diffusion parameterizations demonstrated behaviour suggestive of reduced dependency on a single specific acquisition setting. These summarized metrics may therefore provide complementary evidence of parameter-stable performance in QIB pipelines. Enforcing multi-dimensional robustness assessment to mitigates overfitting and enhances translational reliability.

5.6 Integration of Aims, Findings and Translational Scope

5.6.1 Alignment With Study Hypothesis and Major Aim

The major aim of this study was to evaluate whether longer or novel mpMRI modalities enhance the reliability of PCa assessment, improve repeatability across image weightings and parameter settings, and ultimately translate into clinically relevant information for diagnosis, risk stratification, and treatment planning.

The findings demonstrate that repeatability and reliability differ across diffusion-based modalities, and that QIB-level survival was selectively observed among structural, geometry-derived descriptors under integrated stability and precision constraints. Enhancement was not uniform across modalities or feature families. Rather than identifying a single superior modality, the results highlight modality-specific and descriptor-specific differences in robustness profiles.

Accordingly, the hypothesis that longer or novel mpMRI modalities would uniformly improve discrimination, and stability was only partially supported. Instead, the data indicates that novel diffusion parameterizations influence reliability in structured and heterogeneous ways. Any observed reliability enhancement appears conditional on both acquisition characteristics and the level of structural abstraction embodied in the radiomic descriptor within the constraints of the present dataset.

5.6.2 Repeatability Across Weightings and Parameterizations

A secondary aim was to assess whether novel or extended diffusion modalities improve repeatability across settings and parameterizations. Lesion-level concordance analysis demonstrated high cross-lesion stability for the majority of retained feature–modality

combinations. However, repeatability was not uniformly distributed across all lesions, parameters, or modalities. Broad repeatability was observed within DBSI-derived features during initial ICC thresholding, whereas cross-parameter structural robustness under QIB-stringency was more evident within specific OGSE parameterizations.

Notably, cross-parameter recurrence of structural descriptors, particularly Scharr-derived operators, was evident across multiple OGSE diffusion-time settings. In contrast, certain modalities retained numerous features at exploratory thresholds but failed to maintain persistence under stricter QIB-level criteria. These findings suggest that repeatability appears to operate along multiple dimensions: lesion-level stability, parameter-level persistence, and descriptor-level recurrence.

Instead of indicating uniform modality superiority of any single modality, repeatability emerged as structured and context-dependent across the evaluated diffusion frameworks.

5.6.3 From Technical Robustness to Clinical Relevance

The integrated filtering framework identified eight distinct feature–modality combinations corresponding to three structural radiomic descriptors that satisfied QIB-level candidacy criteria for repeatability, discrimination, statistical significance, and precision. These candidates demonstrated methodological robustness under stability-oriented screening.

However, the study did not directly evaluate clinical decision impact, incremental diagnostic performance over established clinical frameworks, or treatment modification. Accordingly, the results of this thesis establish technical biomarker robustness rather than confirmed clinical benefit. The retained structural descriptors represent QIB-candidacy with potential relevance for PCa risk stratification and assessment, but their translational value requires validation in larger, multi-centre cohorts and comparison against established clinical standards.

5.6.4 Methodological Contribution to QIB Development

This thesis contributes methodologically to QIB-oriented radiomics by implementing a structured, repeatability-first filtering architecture that integrates discrimination and precision constraints within a predefined hierarchy. Incorporating repositioning into the repeatability design introduced controlled variability and strengthens the evaluation of measurement stability under realistic acquisition perturbations, the addition of confidence interval width threshold established an explicit precision criterion.

Cross-parameter recurrence was also examined as a complementary dimension of robustness beyond repeatability and reliability analysis. The concept of abstraction hierarchy further contextualizes stability findings, which suggests that structurally abstracted features may exhibit greater resilience under perturbation in contrast with intensity-proximal features.

This interpretation is grounded in observed survival patterns rather than from a formal hierarchical modelling. All thresholds were predefined prior to integrated filtering, minimizing post hoc optimization and reinforcing the architectural nature of the contribution.

The principal advancement of this work lies not only in identifying specific candidate descriptors, but in demonstrating how stability-oriented framework design fundamentally shapes radiomic feature selection.

5.7 Scope, Limitations and Generalizability

The present analysis was conducted within the PRODIF cohort which utilizes a single-centre test–retest design. This framework enabled controlled assessment of repeatability and reliability under standardized acquisition conditions. However, multi-centre validation across multiple institutions and scanner platforms is required to determine the generalizability of the findings. The study cohort constituted 37 patients, resulting in moderate effective sample sizes across modality-specific analyses.

Although sufficient for structured repeatability and reliability assessment, moderate sample sizes inherently may limit statistical precision and influence AUC and ICC

estimates. Study designs that include larger cohorts would permit narrower confidence intervals, more stable variance estimates and more robust subgroup analysis. No independent external validation dataset was available, and radiomic–histopathological correlation beyond Gleason grade grouping was not performed.

Although the relatively balanced grade distribution reduces extreme class imbalance effects, sampling variability cannot be fully excluded. Consequently, while structural descriptors demonstrated technical robustness under integrated filtering criteria, their biological specificity and incremental clinical utility remain to be independently verified.

An additional consideration concerns the operational definition of csPCa. In this thesis, binary classification was performed using a Grade Group ≥ 3 threshold. Although many imaging studies use ≥ Grade Group 2, the stricter threshold was selected to increase biological separation between clearly indolent and high-grade tumours. Because Grade Group 2 represents an intermediate category with variable Gleason pattern 4 composition, it may introduce phenotypic overlap in quantitative modelling. Nevertheless, this decision limits direct comparability with studies using a ≥ Grade Group 2 definition, and future validation under alternative thresholds is warranted. The stricter threshold may also have contributed to the strong lesion-level discrimination observed under QIB-level filtering.

Parameter interdependence within OGSE diffusion-time settings represents an additional consideration. Although cross-parameter recurrence was observed, diffusion parameterizations are not statistically independent, and further investigation is required to determine whether similar robustness patterns persist across broader acquisition frameworks and alternative diffusion models. Overall, the findings demonstrate strong internal methodological validity under controlled acquisition conditions within the constraints of the study design, but with limited external generalizability. The results should therefore be interpreted as hypothesis-generating at a translational level and as a foundation for future multi-centre validation studies aimed at confirming reproducible QIBs in PCa.

6 Conclusion

The evidence assembled in this thesis aims to provide empirical evidence on whether longer or novel mpMRI modalities reinforce reliability and repeatability of PCa assessment, and whether such improved robustness could potentially translate into clinically meaningful radiomic biomarkers. Using a structured repeatability-first framework that is further compounded by discrimination and precision constraints, the study evinces that biomarker robustness in prostate MRI is settled by the interaction between acquisition design, descriptor abstraction and statistical stability.

Across 106,352 lesion-level ROC-AUC observations, sequential integration of repeatability, statistical significance and confidence interval precision criteria reduced the candidate space to 216 retained observations, corresponding to 110 distinct feature-modality combinations that consisted of 82 distinct radiomic descriptors. Escalation to QIB-oriented thresholds further tightened this set to 8 distinct feature-modality combinations that represent 3 distinct radiomic structural descriptors. This progressive contraction highlights that discriminative performance alone is insufficient for under measurement variability.

The findings partially support the study hypothesis. Novel diffusion parameterizations influenced reliability and robustness profiles; however, enhancement was not uniform across modalities or feature families. DBSI validated broader repeatability retention across radiomic descriptors, whereas specific OGSE parameterizations exhibited selective structural robustness under QIB-level constraints.

Geometric structural descriptors derived from Scharr and Hessian operators within the EC2D3D feature family demonstrated greater resilience than intensity-based texture matrices under test–reposition–retest conditions. Scharr edge-based relative eccentricity and minor axis length (mean) showed consistent retention across five OGSE parameterisations, while Hessian area interquartile range on DBSI restricted-ratio maps achieved QIB-level candidacy. These outcomes attests that structurally abstracted features may capture diagnostically relevant morphology with improved stability compared to voxel-level intensity statistics.

With respect to the study design, this thesis contributes a structured QIB-oriented evaluation pipeline that quantifies repeatability, reliability and precision within hierarchical thresholds. Incorporating repositioning introduced controlled perturbation to evaluate measurement stability under realistic acquisition variability.

While QIB-level candidates were identified, the contribution of this work is based in establishing technical robustness, not in claiming superiority over established diagnostic frameworks such as PI-RADS V2.1. Clinical translation will require external validation, multi-centre reproducibility assessment across scanners, and evaluation of incremental value beyond conventional diffusion metrics and radiologist interpretation.

In light of these findings, through integrating a strict thresholding hierarchy, this thesis demonstrates that stability-first prioritization provides a clinically anchored pathway for identifying reproducible radiomic candidates in PCa imaging. Future work should extend this framework to larger cohorts and multi-centre datasets to determine whether structurally stable radiomic descriptors can meaningfully enhance diagnostic confidence, risk stratification accuracy, and treatment planning in PCa care.

6.1 Generative AI Usage Disclosure

The author used generative artificial intelligence (AI) tools, including large language model–based assistants, to support selected aspects of manuscript preparation. These tools were employed for language editing, grammar correction, structural refinement, and clarification of academic expression. Limited AI-assisted prompting was used to reorganize text and improve coherence in the Introduction and Discussion. AI tools were also used in a limited capacity to assist with drafting and troubleshooting programming scripts related to data preprocessing and workflow implementation. All generated code was reviewed, tested, and validated by the author prior to inclusion in the final analyses.

No AI tools were used to generate research hypotheses, design the study, define statistical thresholds, conduct independent data analyses, interpret results, or formulate scientific conclusions. All scientific content, methodological decisions, analyses, and interpretations were independently developed and verified by the author, who assumes full responsibility for the integrity and accuracy of the work.

Acknowledgements

This thesis would not have been possible without the support and contribution of multiple individuals throughout this writing and examination process. This achievement is as much all of yours as it is mine.

First and foremost, I hereby express my heartfelt gratitude to my supervisor, Harri Merisaari, for his level guidance, patience and intuition. His capacity to shape the way I approach scientific problems truly inspire me to push further. I am thankful for his trust, his critical feedback, and the space he gave me to grow.

I would like to accord my sincere appreciation to the collaborators of the PRODIF study. Their research became the touchstone for this thesis, and it is truly an honour to expound upon it. I am also truly grateful to my examiners for their time, expertise, and careful evaluation of this work.

I also extend my acknowledgements and gratitude to the MDP Biomedical Imaging programme team at the University of Turku and Åbo Akademi. It was within the environment they fostered where my interest in quantitative imaging deepened, and I remain beholden for the inspiration and opportunities they provided.

My sincerest exaltations to my cohort members, classmates, research peers and clinical peers, thank you for the late-night discussions, random chats, honest feedback, and shared frustrations. Those minutes of collaboration and support were just as important as the formal milestones. I extend my thanks to the librarians, faculty members, and especially the study participants that help formulate the intellectual basis of this research.

I am indebted to my family. To my parents and my sister, your faiths in me never wavered, even when mine did. Your patience and quiet encouragement carried me further than you had realized. To my friends as well, many thanks for being by my side no matter how bright the sun or how windy the blizzard.

And to my partner, you have been beside me since the very beginning and you have inspired me to work hard. Your encouragement, understanding, and steady presence have meant more than I can fully express.

Above all, I am grateful to Allah for the courage, constraint, and resilience that have been granted to me throughout this journey. To Him belongs all praise.

And finally, I extend my sincere appreciation to the reader for the time and attention given to this work. This thesis was truly not easy to finish, and by the time you read this, just know that you have my upmost appreciation for making it this far.

As I pen these final words, I state that this thesis marks not an endpoint, but the beginning of our continued commitment to thoughtful and rigorous scientific inquiry. I move forward from this chapter with humility, resilience, and a renewed sense of purpose.

Throughout this work, stability was required before discrimination, and evidence before assertion. In that spirit, I remain guided by the principle:

Nullius in Verba

"Take nobody's word for it".

Kiitos.

References

1. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. *CA Cancer J Clin*. 2023;73(1):17-48. doi:10.3322/caac.21763

2. Bergengren O, Pekala KR, Matsoukas K, et al. 2022 Update on Prostate Cancer Epidemiology and Risk Factors—A Systematic Review. *Eur Urol*. 2023;84(2):191-206. doi:10.1016/j.eururo.2023.04.021

3. Wang G, Zhao D, Spring DJ, DePinho RA. Genetics and biology of prostate cancer. *Genes Dev*. 2018;32(17-18):1105-1140. doi:10.1101/gad.315739.118

4. Young TK, Kelly JJ, Friborg J, Soininen L, Wong KO. Cancer among circumpolar populations: an emerging public health concern. *Int J Circumpolar Health*. 2016;75. https://api.semanticscholar.org/CorpusID:3807396

5. Benke IN, Leitzmann MF, Behrens G, Schmid D. Physical activity in relation to risk of prostate cancer: a systematic review and meta-analysis. *Ann Oncol Off J Eur Soc Med Oncol*. 2018;29(5):1154-1179. doi:10.1093/annonc/mdy073

6. Ferlay J, Ervik M, Lam F, et al. Global cancer observatory: cancer today. *Lyon Int Agency Res Cancer*. 2020;20182020:557-567.

7. Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2024;74(3):229-263. doi:https://doi.org/10.3322/caac.21834

8. Schafer EJ, Laversanne M, Sung H, et al. Recent Patterns and Trends in Global Prostate Cancer Incidence and Mortality: An Update. *Eur Urol*. 2025;87(3):302-313. doi:https://doi.org/10.1016/j.eururo.2024.11.013

9. Shore ND, Morgans AK, El-Haddad G, Srinivas S, Abramowitz M. Addressing Challenges and Controversies in the Management of Prostate Cancer with Multidisciplinary Teams. *Target Oncol*. 2022;17(6):709-725. doi:10.1007/s11523-022-00925-7

10. Epstein JI, Walsh PC, Carmichael M, Brendler CB. Pathologic and clinical findings to predict tumor extent of nonpalpable (stage T1c) prostate cancer. *JAMA*. 1994;271(5):368-374.

11. Cornford P, Tilki D, van den Bergh RCN, et al. *EAU-EANM-ESTRO-ESUR-ISUP-SIOG Guidelines on Prostate Cancer*. (Briers E, ed.). European Association of Urology Guidelines Office; 2024.

12. Schröder FH, Hugosson J, Roobol MJ, et al. Screening and prostate-cancer mortality in a randomized European study. *N Engl J Med*. 2009;360(13):1320-1328. doi:10.1056/NEJMoa0810084

13. Catalona WJ, Southwick PC, Slawin KM, et al. Comparison of percent free PSA, PSA density, and age-specific PSA cutoffs for prostate cancer detection and staging. *Urology*. 2000;56(2):255-260. doi:10.1016/s0090-4295(00)00637-3

14. Wei JT, Barocas D, Carlsson S, et al. Early Detection of Prostate Cancer: AUA/SUO Guideline Part I: Prostate Cancer Screening. *J Urol*. 2023;210(1):45-53. doi:10.1097/JU.0000000000003491

15. Maffei D, Giganti F, Moore CM. Seminar: Revisiting the value of PSA-based prostate cancer screening Essay No 5: Should men undergo MRI before prostate biopsy? (Pro). *Urol Oncol Semin Orig Investig*. 2023;41(2):88-91. doi:10.1016/j.urolonc.2022.04.016

16. Andriole GL, Crawford ED, Grubb RL 3rd, et al. Mortality results from a randomized prostate-cancer screening trial. *N Engl J Med*. 2009;360(13):1310-1319. doi:10.1056/NEJMoa0810696

17. Gulati R, Gore JL, Etzioni R. Comparative effectiveness of alternative prostate-specific antigen--based prostate cancer screening strategies: model estimates of potential benefits and harms. *Ann Intern Med*. 2013;158(3):145-153. doi:10.7326/0003-4819-158-3-201302050-00003

18. Ilic D, Djulbegovic M, Jung JH, et al. Prostate cancer screening with prostate-specific antigen (PSA) test: a systematic review and meta-analysis. *BMJ*. 2018;362:k3519. doi:10.1136/bmj.k3519

19. Grubb RL 3rd, Pinsky PF, Greenlee RT, et al. Prostate cancer screening in the Prostate, Lung, Colorectal and Ovarian cancer screening trial: update on findings from the initial four rounds of screening in a randomized trial. *BJU Int*. 2008;102(11):1524-1530. doi:10.1111/j.1464-410X.2008.08214.x

20. Hugosson J, Roobol MJ, Månsson M, et al. A 16-yr Follow-up of the European Randomized study of Screening for Prostate Cancer. *Eur Urol*. 2019;76(1):43-51. doi:10.1016/j.eururo.2019.02.009

21. Tosoian JJ, Zhang Y, Xiao L, et al. Development and Validation of an 18-Gene Urine Test for High-Grade Prostate Cancer. *JAMA Oncol*. 2024;10(6):726-736. doi:10.1001/jamaoncol.2024.0455

22. Klein EA, Chait A, Hafron JM, et al. The Single-parameter, Structure-based IsoPSA Assay Demonstrates Improved Diagnostic Accuracy for Detection of Any Prostate Cancer and High-grade Prostate Cancer Compared to a Concentration-based Assay of Total Prostate-specific Antigen: A Preliminary Report. *Eur Urol*. 2017;72(6):942-949. doi:10.1016/j.eururo.2017.03.025

23. Wei JT, Barocas D, Carlsson S, et al. Early Detection of Prostate Cancer: AUA/SUO Guideline Part II: Considerations for a Prostate Biopsy. *J Urol*. 2023;210(1):54-63. doi:10.1097/JU.0000000000003492

24. Bjurlin MA, Carter HB, Schellhammer P, et al. Optimization of initial prostate biopsy in clinical practice: sampling, labeling and specimen processing. *J Urol*. 2013;189(6):2039-2046. doi:10.1016/j.juro.2013.02.072

25. Kasivisvanathan V, Rannikko AS, Borghi M, et al. MRI-Targeted or Standard Biopsy for Prostate-Cancer Diagnosis. *N Engl J Med*. 2018;378(19):1767-1777. doi:10.1056/NEJMoa1801993

26. Ahmed HU, El-Shater Bosaily A, Brown LC, et al. Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. *Lancet Lond Engl*. 2017;389(10071):815-822. doi:10.1016/S0140-6736(16)32401-1

27. Tombal B. Over- and Underdiagnosis of Prostate Cancer: The Dangers. *Tech Asp Urooncological Interv*. 2006;5(6):511-513. doi:10.1016/j.eursup.2006.02.019

28. Gulati R. Reducing Prostate Cancer Overdiagnosis. *N Engl J Med*. 2022;387(23):2187-2188. doi:10.1056/NEJMe2214658

29. Dushimova Z, Iztleuov Y, Chingayeva G, et al. Overdiagnosis and Overtreatment in Prostate Cancer. *Diseases*. 2025;13(6). doi:10.3390/diseases13060167

30. Sandhu S, Moore CM, Chiong E, Beltran H, Bristow RG, Williams SG. Prostate cancer. *Lancet Lond Engl*. 2021;398(10305):1075-1090. doi:10.1016/S0140-6736(21)00950-8

31. Mizuno K, Beltran H. Future directions for precision oncology in prostate cancer. *The Prostate*. 2022;82 Suppl 1(Suppl 1):S86-S96. doi:10.1002/pros.24354

32. Chiam K, Carle C, Hughes S, et al. Use of multiparametric magnetic resonance imaging (mpMRI) in active surveillance for low-risk prostate cancer: a scoping review on the benefits and harm of mpMRI in different biopsy scenarios. *Prostate Cancer Prostatic Dis*. 2021;24(3):662-673.

33. Merisaari H, Taimen P, Shiradkar R, et al. Repeatability of radiomics and machine learning for DWI: Short-term repeatability study of 112 patients with prostate cancer. *Magn Reson Med*. 2020;83(6):2293-2309.

34. Barentsz JO, Weinreb JC, Verma S, et al. Synopsis of the PI-RADS v2 Guidelines for Multiparametric Prostate Magnetic Resonance Imaging and Recommendations for Use. *Eur Urol*. 2016;69(1):41-49. doi:10.1016/j.eururo.2015.08.038

35. Ghai S, Haider M. Multiparametric-MRI in diagnosis of prostate cancer. *Indian J Urol*. 2015;31(3):194.

36. van der Leest M, Cornel E, Israël B, et al. Head-to-head Comparison of Transrectal Ultrasound-guided Prostate Biopsy Versus Multiparametric Prostate Resonance Imaging with Subsequent Magnetic Resonance-guided Biopsy in Biopsy-naïve Men with Elevated Prostate-specific Antigen: A Large Prospective Multicenter Clinical Study. *Eur Urol*. 2019;75(4):570-578. doi:10.1016/j.eururo.2018.11.023

37. Rouvière O, Puech P, Renard-Penna R, et al. Use of prostate systematic and targeted biopsy on the basis of multiparametric MRI in biopsy-naive patients (MRI-FIRST): a prospective, multicentre, paired diagnostic study. *Lancet Oncol*. 2019;20(1):100-109. doi:10.1016/S1470-2045(18)30569-2

38. Park JJ, Kim CK. Paradigm Shift in Prostate Cancer Diagnosis: Pre-Biopsy Prostate Magnetic Resonance Imaging and Targeted Biopsy. *Korean J Radiol*. 2022;23(6):625-637.

39. Radiology AC of. Prostate Imaging - Reporting and Data System 2.1. Published online 2019.

40. Scott R, Misser SK, Cioni D, Neri E. PI-RADS v2.1: What has changed and how to report. June 1, 2021. http://www.sajr.org.za/index.php/SAJR/article/view/2062

41. Turkbey B, Brown AM, Sankineni S, Wood BJ, Pinto PA, Choyke PL. Multiparametric prostate magnetic resonance imaging in the evaluation of prostate cancer. *CA Cancer J Clin*. 2016;66(4):326-336. doi:10.3322/caac.21333

42. Woo S, Suh CH, Kim SY, Cho JY, Kim SH. Diagnostic Performance of Prostate Imaging Reporting and Data System Version 2 for Detection of Prostate Cancer: A Systematic Review and Diagnostic Meta-analysis. *Eur Urol*. 2017;72(2):177-188. doi:10.1016/j.eururo.2017.01.042

43. Weinreb JC, Barentsz JO, Choyke PL, et al. PI-RADS Prostate Imaging – Reporting and Data System: 2015, Version 2. *Eur Urol*. 2016;69(1):16-40. doi:10.1016/j.eururo.2015.08.052

44. Jones DK. *Diffusion MRI. London*. Oxford University Press; 784 p; 2010.

45. Godley KC. Focused Diffusion-Weighted Imaging in Prostate Cancer FODIP. In: 2017. https://api.semanticscholar.org/CorpusID:59347259

46. Tamada T, Ueda Y, Ueno Y, Kojima Y, Kido A, Yamamoto A. Diffusion-weighted imaging in prostate cancer. *Magn Reson Mater Phys Biol Med*. 2022;35(4):533-547.

47. Shukla-Dave A, Obuchowski NA, Chenevert TL, et al. Quantitative imaging biomarkers alliance (QIBA) recommendations for improved precision of DWI and DCE-MRI derived biomarkers in multicenter oncology trials. *J Magn Reson Imaging JMRI*. 2019;49(7):e101-e121. doi:10.1002/jmri.26518

48. Sullivan DC, Obuchowski NA, Kessler LG, et al. Metrology Standards for Quantitative Imaging Biomarkers. *Radiology*. 2015;277(3):813-825. doi:10.1148/radiol.2015142202

49. Stoyanova R, Takhar M, Tschudi Y, et al. Prostate cancer radiomics and the promise of radiogenomics. *Transl Cancer Res*. 2016;5(4):432-447. doi:10.21037/tcr.2016.06.20

50. Zwanenburg A, Vallières M, Abdalah MA, et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. *Radiology*. 2020;295(2):328-338. doi:10.1148/radiol.2020191145

51. Paulsen F, Böckers TM, Waschke J. *Sobotta Anatomy Textbook*. Elsevier Health Sciences; 2018.

52. Standring S. *GRAY'S ANATOMY : The Anatomical Basis of Clinical Practice.* 42nd ed. Elsevier Health Sciences; 2020.

53. Wein AJ. *Campbell-Walsh-Wein Urology Twelfth Edition Review.* Elsevier; 2020.

54. McNeal JE. The zonal anatomy of the prostate. *The Prostate*. 1981;2(1):35-49. doi:10.1002/pros.2990020105

55. Yacoub JH, Oto A. MR Imaging of Prostate Zonal Anatomy. *Radiol Clin*. 2018;56(2):197-209. doi:10.1016/j.rcl.2017.10.003

56. Hricak H, Dooms GC, McNeal JE, et al. MR imaging of the prostate gland: normal anatomy. *AJR Am J Roentgenol*. 1987;148(1):51-58. doi:10.2214/ajr.148.1.51

57. Cao N, Lu Q, Si J, et al. The Characteristics of the Transitional Zone in Prostate Growth With Age. *Urology*. 2017;105:136-140. doi:10.1016/j.urology.2017.03.010

58. Rosenkrantz AB, Taneja SS. Radiologist, be aware: ten pitfalls that confound the interpretation of multiparametric prostate MRI. *AJR Am J Roentgenol*. 2014;202(1):109-120. doi:10.2214/AJR.13.10699

59. Rosenkrantz AB, Duszak RJ, Babb JS, Glover M, Kang SK. Discrepancy Rates and Clinical Impact of Imaging Secondary Interpretations: A Systematic Review and Meta-Analysis. *J Am Coll Radiol JACR*. 2018;15(9):1222-1231. doi:10.1016/j.jacr.2018.05.037

60. Ali A, Du Feu A, Oliveira P, Choudhury A, Bristow RG, Baena E. Prostate zones and cancer: lost in transition? *Nat Rev Urol*. 2022;19(2):101-115. doi:10.1038/s41585-021-00524-7

61. Turkbey B, Huang R, Vourganti S, et al. Age-related changes in prostate zonal volumes as measured by high-resolution magnetic resonance imaging (MRI): a cross-sectional study in over 500 patients. *BJU Int*. 2012;110(11):1642-1647. doi:10.1111/j.1464-410X.2012.11469.x

62. Laschkar S, Montagne S, De Kerviler E, et al. Zonal anatomy of the prostate using magnetic resonance imaging, morphometrics, and radiomic features: impact of age-related changes. *Br J Radiol*. 2022;95(1131):20210156. doi:10.1259/bjr.20210156

63. Lee CH, Akin-Olugbade O, Kirschenbaum A. Overview of Prostate Anatomy, Histology, and Pathology. *Endocrinol Metab Clin North Am*. 2011;40(3):565-575. doi:10.1016/j.ecl.2011.05.012

64. Cheng L, Montironi R, Bostwick DG, Lopez-Beltran A, Berney DM. Staging of prostate cancer. *Histopathology*. 2012;60(1):87-117. doi:10.1111/j.1365-2559.2011.04025.x

65. Mayerhoefer ME, Materka A, Langs G, et al. Introduction to Radiomics. *J Nucl Med Off Publ Soc Nucl Med*. 2020;61(4):488-495. doi:10.2967/jnumed.118.222893

66. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology*. 2016;278(2):563-577. doi:10.1148/radiol.2015151169

67. Yip SSF, Aerts HJWL. Applications and limitations of radiomics. *Phys Med Biol*. 2016;61(13):R150-166. doi:10.1088/0031-9155/61/13/R150

68. Bleker J, Kwee TC, Yakar D. Quality of Multicenter Studies Using MRI Radiomics for Diagnosing Clinically Significant Prostate Cancer: A Systematic Review. *Life Basel Switz*. 2022;12(7). doi:10.3390/life12070946

69. Hajianfar G, Hosseini SA, Bagherieh S, Oveisi M, Shiri I, Zaidi H. Impact of harmonization on the reproducibility of MRI radiomic features when using different scanners, acquisition parameters, and image pre-processing techniques: a phantom study. *Med Biol Eng Comput*. 2024;62. doi:10.1007/s11517-024-03071-6

70. Rebello RJ, Oing C, Knudsen KE, et al. Prostate cancer. *Nat Rev Dis Primer*. 2021;7(1):9. doi:10.1038/s41572-020-00243-0

71. Epstein JI, Egevad L, Amin MB, et al. The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma: Definition of Grading Patterns and Proposal for a New Grading System. *Am J Surg Pathol*. 2016;40(2). https://journals.lww.com/ajsp/fulltext/2016/02000/the_2014_international_society_of_urological.10.aspx

72. Epstein JI, Zelefsky MJ, Sjoberg DD, et al. A Contemporary Prostate Cancer Grading System: A Validated Alternative to the Gleason Score. *Eur Urol*. 2016;69(3):428-435. doi:10.1016/j.eururo.2015.06.046

73. Pierorazio PM, Walsh PC, Partin AW, Epstein JI. Prognostic Gleason grade grouping: data based on the modified Gleason scoring system. *BJU Int*. 2013;111(5):753-760. doi:10.1111/j.1464-410X.2012.11611.x

74. Sarkar RR, Parsons JK, Bryant AK, et al. Association of Treatment With 5α-Reductase Inhibitors With Time to Diagnosis and Mortality in Prostate Cancer. *JAMA Intern Med*. 2019;179(6):812-819. doi:10.1001/jamainternmed.2019.0280

75. Hoffman RM, Clanon DL, Littenberg B, Frank JJ, Peirce JC. Using the free-to-total prostate-specific antigen ratio to detect prostate cancer in men with nonspecific elevations of prostate-specific antigen levels. *J Gen Intern Med*. 2000;15(10):739-748. doi:10.1046/j.1525-1497.2000.90907.x

76. Yusim I, Krenawi M, Mazor E, Novack V, Mabjeesh NJ. The use of prostate specific antigen density to predict clinically significant prostate cancer. *Sci Rep*. 2020;10(1):20015. doi:10.1038/s41598-020-76786-9

77. Vickers AJ, Brewster SF. PSA Velocity and Doubling Time in Diagnosis and Prognosis of Prostate Cancer. *Br J Med Surg Urol*. 2012;5(4):162-168. doi:10.1016/j.bjmsu.2011.08.006

78. Constantin T, Savu DA, Bucur Ștefana, Predoiu G, Constantin MM, Jinga V. The Role and Significance of Bioumoral Markers in Prostate Cancer. *Cancers*. 2021;13(23). doi:10.3390/cancers13235932

79. Loeb S, Lilja H, Vickers A. Beyond prostate-specific antigen: utilizing novel strategies to screen men for prostate cancer. *Curr Opin Urol*. 2016;26(5):459-465. doi:10.1097/MOU.0000000000000316

80. Russo GI, Regis F, Castelli T, et al. A Systematic Review and Meta-analysis of the Diagnostic Accuracy of Prostate Health Index and 4-Kallikrein Panel Score in Predicting Overall and High-grade Prostate Cancer. *Clin Genitourin Cancer*. 2017;15(4):429-439.e1. doi:10.1016/j.clgc.2016.12.022

81. Farha MW, Salami SS. Biomarkers for prostate cancer detection and risk stratification. *Ther Adv Urol*. 2022;14:17562872221103988. doi:10.1177/17562872221103988

82. Liu Y, Hatano K, Nonomura N. Liquid Biomarkers in Prostate Cancer Diagnosis: Current Status and Emerging Prospects. *World J Mens Health*. 2025;43(1):8-27.

83. Gandaglia G, Albers P, Abrahamsson PA, et al. Structured Population-based Prostate-specific Antigen Screening for Prostate Cancer: The European Association of Urology Position in 2019. *Eur Urol*. 2019;76(2):142-150. doi:10.1016/j.eururo.2019.04.033

84. Bell KJL, Del Mar C, Wright G, Dickinson J, Glasziou P. Prevalence of incidental prostate cancer: A systematic review of autopsy studies. *Int J Cancer*. 2015;137(7):1749-1757. doi:10.1002/ijc.29538

85. Mitsiades IR, Onozato M, Iafrate AJ, et al. ERBB2/HOXB13 co-amplification with interstitial loss of BRCA1 defines a unique subset of breast cancers. *Breast Cancer Res BCR*. 2024;26(1):185. doi:10.1186/s13058-024-01943-1

86. Umarane PB, Nerli RB, Malik SC. Integrating molecular diagnostics for early prostate cancer detection. *Oncoscience*. 2025;12:58-64. doi:10.18632/oncoscience.620

87. Pilarski R. The Role of BRCA Testing in Hereditary Pancreatic and Prostate Cancer Families. *Am Soc Clin Oncol Educ Book*. 2019;(39):79-86. doi:10.1200/EDBK_238977

88. Guerrios-Rivera L, Howard LE, Wiggins EK, et al. Metabolic syndrome is associated with aggressive prostate cancer regardless of race. *Cancer Causes Control CCC*. 2023;34(3):213-221. doi:10.1007/s10552-022-01649-9

89. Di Francesco S, Robuffo I, Caruso M, et al. Metabolic Alterations, Aggressive Hormone-Naïve Prostate Cancer and Cardiovascular Disease: A Complex Relationship. *Medicina (Mex)*. 2019;55(3). doi:10.3390/medicina55030062

90. Hsing AW, Sakoda LC, Chua SJ. Obesity, metabolic syndrome, and prostate cancer. *Am J Clin Nutr*. 2007;86(3):s843-857. doi:10.1093/ajcn/86.3.843S

91. Serefoglu EC, Altinova S, Ugras NS, Akincioglu E, Asil E, Balbay MD. How reliable is 12-core prostate biopsy procedure in the detection of prostate cancer? *Can Urol Assoc J J Assoc Urol Can*. 2013;7(5-6):E293-298. doi:10.5489/cuaj.11224

92. Descotes JL. Diagnosis of prostate cancer. *Asian J Urol*. 2019;6(2):129-136. doi:10.1016/j.ajur.2018.11.007

93. Malewski W, Milecki T, Tayara O, et al. Role of Systematic Biopsy in the Era of Targeted Biopsy: A Review. *Curr Oncol Tor Ont*. 2024;31(9):5171-5194. doi:10.3390/curroncol31090383

94. Kawa SM, Stroomberg HV, Larsen SB, et al. A Nationwide Analysis of Risk of Prostate Cancer Diagnosis and Mortality following an Initial Negative Transrectal Ultrasound Biopsy with Long-Term Followup. *J Urol*. 2022;208(1):100-108. doi:10.1097/JU.0000000000002491

95. Osama S, Serboiu C, Taciuc IA, et al. Current Approach to Complications and Difficulties during Transrectal Ultrasound-Guided Prostate Biopsies. *J Clin Med*. 2024;13(2). doi:10.3390/jcm13020487

96. Hagiwara A, Fujita S, Kurokawa R, Andica C, Kamagata K, Aoki S. Multiparametric MRI: From Simultaneous Rapid Acquisition Methods and Analysis Techniques Using Scoring, Machine Learning, Radiomics, and Deep Learning to the Generation of Novel Metrics. *Invest Radiol*. 2023;58(8):548-560.

97. Shao L, Liang C, Yan Y, et al. An MRI–pathology foundation model for noninvasive diagnosis and grading of prostate cancer. *Nat Cancer*. 2025;6(10):1621-1637. doi:10.1038/s43018-025-01041-x

98. de Rooij M, Hamoen EHJ, Fütterer JJ, Barentsz JO, Rovers MM. Accuracy of Multiparametric MRI for Prostate Cancer Detection: A Meta-Analysis. *Am J Roentgenol*. 2014;202(2):343-351. doi:10.2214/AJR.13.11046

99. Afzali M, Pieciak T, Newman S, et al. The sensitivity of diffusion MRI to microstructural properties and experimental factors. *J Neurosci Methods*. 2021;347:108951. doi:10.1016/j.jneumeth.2020.108951

100. Shirani A, Sun P, Trinkaus K, et al. Diffusion basis spectrum imaging for identifying pathologies in MS subtypes. *Ann Clin Transl Neurol*. 2019;6(11):2323-2327. doi:10.1002/acn3.50903

101. Sun P, George A, Perantie DC, et al. Diffusion basis spectrum imaging provides insights into MS pathology. *Neurol Neuroimmunol Neuroinflammation*. 7(2):e655. doi:10.1212/NXI.0000000000000655

102. Sun F, Huang Y, Wang J, Hong W, Zhao Z. Research Progress in Diffusion Spectrum Imaging. *Brain Sci*. 2023;13:1497. doi:10.3390/brainsci13101497

103. Orlhac F, Lecler A, Savatovski J, et al. How can we combat multicenter variability in MR radiomics? Validation of a correction procedure. *Eur Radiol*. 2021;31(4):2272-2280. doi:10.1007/s00330-020-07284-9

104. Dickinson L, Ahmed HU, Allen C, et al. Magnetic Resonance Imaging for the Detection, Localisation, and Characterisation of Prostate Cancer: Recommendations from a European Consensus Meeting. *Eur Urol*. 2011;59(4):477-494. doi:10.1016/j.eururo.2010.12.009

105. Traverso A, Wee L, Dekker A, Gillies R. Repeatability and Reproducibility of Radiomic Features: A Systematic Review. *Int J Radiat Oncol Biol Phys*. 2018;102(4):1143-1158. doi:10.1016/j.ijrobp.2018.05.053

106. Vargas HA, Hötker AM, Goldman DA, et al. Updated prostate imaging reporting and data system (PIRADS v2) recommendations for the detection of clinically significant prostate cancer using multiparametric MRI: critical evaluation using whole-mount pathology as standard of reference. *Eur Radiol*. 2016;26(6):1606-1612. doi:10.1007/s00330-015-4015-6

107. Al-Bourini O, Seif Amir Hosseini A, Giganti F, et al. T1 Mapping of the Prostate Using Single-Shot T1FLASH: A Clinical Feasibility Study to Optimize Prostate Cancer Assessment. *Invest Radiol*. 2023;58(6):380-387. doi:10.1097/RLI.0000000000000945

108. Hricak H, Choyke PL, Eberhardt SC, Leibel SA, Scardino PT. Imaging prostate cancer: a multidisciplinary perspective. *Radiology*. 2007;243(1):28-53. doi:10.1148/radiol.2431030580

109. Akin O, Sala E, Moskowitz CS, et al. Transition zone prostate cancers: features, detection, localization, and staging at endorectal MR imaging. *Radiology*. 2006;239(3):784-792. doi:10.1148/radiol.2392050949

110. Ueno Y, Tamada T, Sofue K, Murakami T. Diffusion and quantification of diffusion of prostate cancer. *Br J Radiol*. 2022;95(1131):20210653. doi:10.1259/bjr.20210653

111. Lucarelli NM, Villanova I, Maggialetti N, et al. Quantitative ADC: An Additional Tool in the Evaluation of Prostate Cancer? *J Pers Med*. 2023;13(9). doi:10.3390/jpm13091378

112. Yücel Oğuzdoğan G, Adıbelli ZH, Şefik E, Arslan FZ. Apparent Diffusion Coefficient Values as a Complementary Tool in Prostate Gland Disease: Retrospective Evaluation of Apparent Diffusion Coefficient Values with Pathological Data Guided by PI-RADSv2.1. *Bull Urooncology*. 2023;22(1):28-34.

113. Rezaeian A, Ostovari M, Hoseini-Ghahfarokhi M, Khanbabaei H. Diffusion-weighted magnetic resonance imaging at 1.5 T for peripheral zone prostate cancer: the influence of the b-value combination on the diagnostic performance of apparent diffusion coefficient. *Pol J Radiol*. 2022;87:e215-e219. doi:10.5114/pjr.2022.115715

114. Brancato V, Cavaliere C, Salvatore M, Monti S. Non-Gaussian models of diffusion weighted imaging for detection and characterization of prostate cancer: a systematic review and meta-analysis. *Sci Rep*. 2019;9(1):16837. doi:10.1038/s41598-019-53350-8

115. Medved M, Soylu-Boy FN, Karademir I, et al. High-resolution diffusion-weighted imaging of the prostate. *AJR Am J Roentgenol*. 2014;203(1):85-90. doi:10.2214/AJR.13.11098

116. Saouaf R, Xie Y, Kim S, et al. High-resolution Diffusion-weighted Imaging to Detect Changes in Tumor Size and ADC, and Predict Adverse Biopsy Histology during Prostate Cancer Active Surveillance. *Cancer Res Commun*. 2024;4(3):938-945. doi:10.1158/2767-9764.CRC-24-0009

117. Nakamoto A, Onishi H, Tsuboyama T, et al. High-resolution Diffusion-weighted Imaging of the Prostate Using Multiplexed Sensitivity-encoding: Comparison with the Conventional and Reduced Field-of-view Techniques. *Magn Reson Med Sci*. 2025;24(1):58-65. doi:10.2463/mrms.mp.2023-0039

118. Portnoy S, Flint JJ, Blackband SJ, Stanisz GJ. Oscillating and pulsed gradient diffusion magnetic resonance microscopy over an extended b-value range: implications for the characterization of tissue microstructure. *Magn Reson Med*. 2013;69(4):1131-1145. doi:10.1002/mrm.24325

119. Zhu A, Michael ES, Li H, et al. Engineering clinical translation of OGSE diffusion MRI. *Magn Reson Med*. 2025;94(3):913-936. doi:10.1002/mrm.30510

120. Does MD, Parsons EC, Gore JC. Oscillating gradient measurements of water diffusion in normal and globally ischemic rat brain. *Magn Reson Med*. 2003;49(2):206-215. doi:10.1002/mrm.10385

121. Xu J, Does MD, Gore JC. Quantitative characterization of tissue microstructure with temporal diffusion spectroscopy. *J Magn Reson San Diego Calif 1997*. 2009;200(2):189-197. doi:10.1016/j.jmr.2009.06.022

122. Baron CA, Beaulieu C. Oscillating gradient spin-echo (OGSE) diffusion tensor imaging of the human brain. *Magn Reson Med*. 2014;72(3):726-736. doi:10.1002/mrm.24987

123. Merisaari H, Toivonen J, Pesola M, et al. Diffusion-weighted imaging of prostate cancer: effect of b-value distribution on repeatability and cancer characterization. *Magn Reson Imaging*. 2015;33(10):1212-1218. doi:10.1016/j.mri.2015.07.004

124. Merisaari H, Movahedi P, Perez IM, et al. Fitting methods for intravoxel incoherent motion imaging of prostate cancer on region of interest level: Repeatability and gleason score prediction. *Magn Reson Med*. 2017;77(3):1249-1264. doi:10.1002/mrm.26169

125. Iyama Y, Nakaura T, Kidoh M, et al. Fat Suppressed Contrast-Enhanced T1-Weighted Dynamic Magnetic Resonance Imaging at 3T: Comparison of Image Quality Between Spectrally Adiabatic Iversion Recovery and the Multiecho Dixon Technique in Imaging of the Prostate. *J Comput Assist Tomogr*. 2017;41(3):382-387. doi:10.1097/RCT.0000000000000540

126. Delfaut EM, Beltran J, Johnson G, Rousseau J, Marchandise X, Cotten A. Fat Suppression in MR Imaging: Techniques and Pitfalls. *RadioGraphics*. 1999;19(2):373-382. doi:10.1148/radiographics.19.2.g99mr03373

127. Tannus A, Garwood M. Adiabatic pulses. *NMR Biomed*. 1997;10(8):423-434.

128. Visser F. *Application Tip: Optimizing SPIR and SPAIR Fat Suppression.* Philips Healthcare; 2004.

129. Purysko AS, Baroni RH, Giganti F, et al. PI-RADS Version 2.1: A Critical Review, From the AJR Special Series on Radiology Reporting and Data Systems. *Am J Roentgenol*. 2021;216(1):20-32. doi:10.2214/AJR.20.24495

130. Padhani AR, Barentsz J, Villeirs G, et al. PI-RADS Steering Committee: The PI-RADS Multiparametric MRI and MRI-directed Biopsy Pathway. *Radiology*. 2019;292(2):464-474. doi:10.1148/radiol.2019182946

131. Beyer TS Heinz Peter; Weber, Marc André; Thierfelder, Kolja M. PI-RADS 2.1 – Image Interpretation: The Most Important Updates and Their Clinical Implications. *RöFo - Fortschritte Auf Dem Geb Röntgenstrahlen Bildgeb Verfahr*. 2020;193(07):787-796. doi:10.1055/a-1324-4010

132. Sundaram AD. PI-RADS 2.1: A Practical Overview. *J Gastrointest Abdom Radiol*. 2024;07(03):169-182. doi:10.1055/s-0044-1787295

133. Yang L, Li M, Zhang MN, Yao J, Song B. Association of prostate zonal volume with location and aggressiveness of clinically significant prostate cancer: A multiparametric MRI study according to PI-RADS version 2.1. *Eur J Radiol*. 2022;150:110268. doi:10.1016/j.ejrad.2022.110268

134. Cinar I, Cinar E. Assessment of interobserver variability in Gleason grading for prostate carcinoma. *North Clin Istanb*. 2025;12(3):337-343. doi:10.14744/nci.2025.11456

135. Drost FJH, Osses, DF, Nieboer, D, Steyerberg, EW, Bangma, CH, Roobol, MJ, Schoots I. Prostate MRI, with or without MRI-targeted biopsy, and systematic biopsy for detecting prostate cancer. *Cochrane Database Syst Rev*. 2019;(4). doi:10.1002/14651858.CD012663.pub2

136. Moldovan PC, Van den Broeck T, Sylvester R, et al. What Is the Negative Predictive Value of Multiparametric Magnetic Resonance Imaging in Excluding Prostate Cancer at Biopsy? A Systematic Review and Meta-analysis from the European Association of Urology Prostate Cancer Guidelines Panel. *Eur Urol*. 2017;72(2):250-266. doi:10.1016/j.eururo.2017.02.026

137. Mehralivand S, Shih JH, Rais-Bahrami S, et al. A Magnetic Resonance Imaging–Based Prediction Model for Prostate Biopsy Risk Stratification. *JAMA Oncol*. 2018;4(5):678-685. doi:10.1001/jamaoncol.2017.5667

138. Giganti F, Allen C, Emberton M, Moore CM, Kasivisvanathan V. Prostate Imaging Quality (PI-QUAL): A New Quality Control Scoring System for Multiparametric Magnetic Resonance Imaging of the Prostate from the PRECISION trial. *Eur Urol Oncol*. 2020;3(5):615-619. doi:10.1016/j.euo.2020.06.007

139. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. 2017;14(12):749-762. doi:10.1038/nrclinonc.2017.141

140. Casella G, Berger RL. *Statistical Inference*. Thomson Learning; 2002. https://books.google.fi/books?id=0x_vAAAAMAAJ

141. Gonzalez RC, Woods RE. *Digital Image Processing, Global Edition*. Pearson Education; 2018. https://books.google.fi/books?id=p74oEAAAQBAJ

142. Wibmer A, Hricak H, Gondo T, 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(10):2840-2850. doi:10.1007/s00330-015-3701-8

143. Haralick RM, Shanmugam K, Dinstein I. Textural Features for Image Classification. *IEEE Trans Syst Man Cybern*. 1973;SMC-3(6):610-621. doi:10.1109/TSMC.1973.4309314

144. Galloway MM. Texture analysis using gray level run lengths. *Comput Graph Image Process*. 1975;4(2):172-179. doi:10.1016/S0146-664X(75)80008-6

145. Thibault G, Fertil B, Navarro CL, et al. Texture indexes and gray level size zone matrix. Application to cell nuclei classification. In: 2009. https://api.semanticscholar.org/CorpusID:18989026

146. Amadasun M, King R. Textural features corresponding to textural properties. *IEEE Trans Syst Man Cybern*. 1989;19(5):1264-1274. doi:10.1109/21.44046

147. Mackin D, Fave X, Zhang L, et al. Measuring Computed Tomography Scanner Variability of Radiomics Features. *Invest Radiol*. 2015;50(11):757-765. doi:10.1097/RLI.0000000000000180

148. Bracewell RN. *The Fourier Transform and Its Applications*. McGraw Hill; 2000. https://books.google.fi/books?id=ZNQQAQAAIAAJ

149. Jain AK, Farrokhnia F. Unsupervised texture segmentation using Gabor filters. *Pattern Recognit*. 1991;24(12):1167-1186. doi:10.1016/0031-3203(91)90143-S

150. Bovik AC, Clark M, Geisler WS. Multichannel texture analysis using localized spatial filters. *IEEE Trans Pattern Anal Mach Intell*. 1990;12(1):55-73. doi:10.1109/34.41384

151. Gabor D. Theory of communication. Part 1: The analysis of information. *J Inst Electr Eng - Part III Radio Commun Eng*. 1946;93(26):429-441. doi:10.1049/ji-3-2.1946.0074

152. Daugman JG. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J Opt Soc Am A*. 1985;2(7):1160-1169. doi:10.1364/JOSAA.2.001160

153. Caelli T, Moraglia G. On the detection of Gabor signals and discrimination of Gabor textures. *Vision Res*. 1985;25(5):671-684. doi:10.1016/0042-6989(85)90173-7

154. Shafiq-Ul-Hassan M, Zhang GG, Latifi K, et al. Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. *Med Phys*. 2017;44(3):1050-1062. doi:10.1002/mp.12123

155. Scharr H. *Optimal Operators in Digital Image Processing [Elektronische Ressource] /*. 2014.

156. Frangi AF, Niessen WJ, Vincken KL, Viergever MA. Multiscale vessel enhancement filtering. In: Wells WM, Colchester A, Delp S, eds. *Medical Image Computing and Computer-Assisted Intervention — MICCAI'98*. Springer Berlin Heidelberg; 1998:130-137.

157. Szeliski R. *Computer Vision: Algorithms and Applications*. Springer International Publishing; 2022. https://books.google.fi/books?id=QptXEAAAQBAJ

158. Lindeberg T. Feature Detection with Automatic Scale Selection. *Int J Comput Vis*. 1998;30:77-116. doi:10.1023/A:1008045108935

159. Sato Y, Nakajima S, Shiraga N, et al. Three-dimensional multi-scale line filter for segmentation and visualization of curvilinear structures in medical images. *Med Image Anal*. 1998;2(2):143-168. doi:10.1016/s1361-8415(98)80009-1

160. Harris CG, Stephens MJ. A Combined Corner and Edge Detector. In: *Alvey Vision Conference*. 1988. https://api.semanticscholar.org/CorpusID:1694378

161. Shi J, Tomasi. Good features to track. In: *1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. 1994:593-600. doi:10.1109/CVPR.1994.323794

162. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res*. 2017;77(21):e104-e107. doi:10.1158/0008-5472.CAN-17-0339

163. Varoquaux G. Cross-validation failure: Small sample sizes lead to large error bars. *NeuroImage*. 2018;180(Pt A):68-77. doi:10.1016/j.neuroimage.2017.06.061

164. Hiremath A, Shiradkar R, Merisaari H, et al. Test-retest repeatability of a deep learning architecture in detecting and segmenting clinically significant prostate cancer on apparent diffusion coefficient (ADC) maps. *Eur Radiol*. 2021;31(1):379-391. doi:10.1007/s00330-020-07065-4

165. Schwier M, van Griethuysen J, Vangel MG, et al. Repeatability of Multiparametric Prostate MRI Radiomics Features. *Sci Rep*. 2019;9(1):9441. doi:10.1038/s41598-019-45766-z

166. Raunig DL, McShane LM, Pennello G, et al. Quantitative imaging biomarkers: A review of statistical methods for technical performance assessment. *Stat Methods Med Res*. 2015;24(1):27-67. doi:10.1177/0962280214537344

167. Buckler AJ, Bresolin L, Dunnick NR, Sullivan DC. A collaborative enterprise for multi-stakeholder participation in the advancement of quantitative imaging. *Radiology*. 2011;258(3):906-914. doi:10.1148/radiol.10100799

168. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies. *Radiology*. 2015;277(3):826-832. doi:10.1148/radiol.2015151516

169. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. Reporting recommendations for tumor marker prognostic studies (REMARK). *J Natl Cancer Inst*. 2005;97(16):1180-1184. doi:10.1093/jnci/dji237

170. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*. 2016;15(2):155-163. doi:10.1016/j.jcm.2016.02.012

171. Shrout P, Fleiss J. Intraclass correlations: Uses in assessing rater reliability. *Psychol Bull*. 1979;86:420-428. doi:10.1037/0033-2909.86.2.420

172. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods*. 1996;1(1):30-46. doi:10.1037/1082-989X.1.1.30

173. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res*. 2005;19(1):231-240. doi:10.1519/15184.1

174. Pepe MS. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford University Press; 2003. doi:10.1093/oso/9780198509844.001.0001

175. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982;143(1):29-36. doi:10.1148/radiology.143.1.7063747

176. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem*. 1993;39(4):561-577.

177. Bamber D. The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *J Math Psychol*. 1975;12(4):387-415. doi:10.1016/0022-2496(75)90001-2

178. H. B. Mann, D. R. Whitney. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *Ann Math Stat*. 1947;18(1):50-60. doi:10.1214/aoms/1177730491

179. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44(3):837-845.

180. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B Methodol*. 1995;57(1):289-300. doi:10.1111/j.2517-6161.1995.tb02031.x

181. Ioannidis JPA. Why most published research findings are false. *PLoS Med*. 2005;2(8):e124. doi:10.1371/journal.pmed.0020124

182. Algohary A, Viswanath S, Shiradkar R, et al. Radiomic features on MRI enable risk categorization of prostate cancer patients on active surveillance: Preliminary findings. *J Magn Reson Imaging*. 2018;48(3):818-828. doi:10.1002/jmri.25983

183. Cuocolo R, Stanzione A, Ponsiglione A, et al. Clinically significant prostate cancer detection on MRI: A radiomic shape features study. *Eur J Radiol*. 2019;116:144-149. doi:10.1016/j.ejrad.2019.05.006

184. Shiradkar R, Ghose S, Jambor I, et al. Radiomic features from pretreatment biparametric MRI predict prostate cancer biochemical recurrence: Preliminary findings. *J Magn Reson Imaging*. 2018;48(6):1626-1636. doi:10.1002/jmri.26178

185. Castillo J, de la Iglesia A, Leiva M, Jodar M, Oliva R. Proteomics of human spermatozoa. *Hum Reprod Oxf Engl*. 2023;38(12):2312-2320. doi:10.1093/humrep/dead170

186. Prathyush Chirra, Patrick Leo, Michael Yim, et al. Multisite evaluation of radiomic feature reproducibility and discriminability for identifying peripheral zone prostate tumors on MRI. *J Med Imaging*. 2019;6(2):024502. doi:10.1117/1.JMI.6.2.024502

187. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol*. 2019;110:12-22. doi:10.1016/j.jclinepi.2019.02.004

188. Mokkink LB, Eekhout I, Boers M, van der Vleuten CP, de Vet HC. Studies on Reliability and Measurement Error of Measurements in Medicine – From Design to Statistics Explained for Medical Researchers. *Patient Relat Outcome Meas Vol*. 2023;14:193-212.

189. Hajian-Tilaki K. Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. *Casp J Intern Med*. 2013;4(2):627-635.

190. Berrar D, Flach P. Caveats and pitfalls of ROC analysis in clinical microarray research (and how to avoid them). *Brief Bioinform*. 2012;13(1):83-97. doi:10.1093/bib/bbr008

191. Riley RD, Snell KIE, Ensor J, et al. Minimum sample size for developing a multivariable prediction model: Part I – Continuous outcomes. *Stat Med*. 2019;38(7):1262-1275. doi:10.1002/sim.7993

192. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJWL. Machine Learning methods for Quantitative Radiomic Biomarkers. *Sci Rep*. 2015;5(1):13087. doi:10.1038/srep13087

193. Lahoti AM, Lakhotiya AR, Ingole SM, Dhok AP, Mudaliar PN. Role and Application of Diffusion-weighted Imaging in Evaluation of Prostate Cancer. *Indian J Med Paediatr Oncol*. 2018;39:349-354. doi:10.4103/ijmpo.ijmpo_41_17

194. Cross AH, Song SK. "A new imaging modality to non-invasively assess multiple sclerosis pathology". *J Neuroimmunol*. 2017;304:81-85. doi:10.1016/j.jneuroim.2016.10.002

195. Drobnjak I, Zhang H, Ianuş A, Kaden E, Alexander DC. PGSE, OGSE, and sensitivity to axon diameter in diffusion MRI: Insight from a simulation study. *Magn Reson Med*. 2016;75(2):688-700. doi:10.1002/mrm.25631

196. Vadakkenveettil B. Grey Level Co-Occurrence Matrices: Generalisation and Some New Features. *Int J Comput Sci Eng Inf Technol*. 2012;2:151-157. doi:10.5121/ijcseit.2012.2213

Appendices

Appendix 1. Expanded Names of MRI Sequences and Derived Parameters

Diffusion Basis Spectrum Imaging (DBSI)

- Low-Resolution Axial Diffusivity
- Low-Resolution Radial Diffusivity
- Low-Resolution Fiber Fraction Axial Diffusivity
- Low-Resolution Fiber Fraction Radial Diffusivity
- Low-Resolution Restricted Fraction
- Low-Resolution Water Fraction

High-Resolution Diffusion-Weighted Imaging (High-Resolution DWI)

- Apparent Diffusion Coefficient

Oscillating Gradient Spin-Echo (OGSE)

Definition of N:
N denotes the oscillation frequency index of the diffusion-sensitizing gradient waveform in Oscillating Gradient Spin-Echo imaging. Each N-value corresponds to a specific effective diffusion time, enabling probing of tissue microstructure at different temporal scales.

- N1 Apparent Diffusion Coefficient
- N1 Diffusion Dispersion Coefficient
- N2 Apparent Diffusion Coefficient
- N2 Diffusion Dispersion Coefficient
- N4 Apparent Diffusion Coefficient
- N4 Diffusion Dispersion Coefficient

- N6 Axial Diffusivity
- N6 Diffusion Dispersion Coefficient

Spectral Attenuated Inversion Recovery (SPAIR)

- Four b-value Apparent Diffusion Coefficient
- Twelve b-value Apparent Diffusion Coefficient

T2-Weighted Imaging (T2W)

No quantitative diffusion-derived parameters extracted

Appendix 2. Analysis Pipeline in R

The full analysis pipeline and raw AUC/ICC results are available in the following GitLab repository:

Repository: prodif-thesis-radiomics
Platform: GitLab
URL: https://gitlab.utu.fi/aaichs/prodif-thesis-radiomics
Commit version used for this thesis:

dcc0305182fbba570a8290e1b8db86e4cdee0afe

Accessed: 01 March 2026

All analyses reported in this thesis were performed using this exact commit to ensure reproducibility. Some patient-related data were not available for confidentiality purposes.