BIOMEDICAL ENGINEERING MSc PROGRAM
UNIVERSITY OF CRETE
TECHNICAL UNIVERSITY OF CRETE
FOUNDATION FOR RESEARCH AND TECHNOLOGY-HELLAS

Application Grade Thesis

Title: A radiomics analysis pipeline in prostate cancer imaging

Student's Name: Evangelia Ilia

Supervisor's Name: Konstantinos Marias

Date of completion: 12th June 2024

This dissertation is submitted as a partial fulfilment of the requirements for the Master's degree of Biomedical Engineering M.Sc. program

BIOMEDICAL ENGINEERING MSc PROGRAM
UNIVERSITY OF CRETE
TECHNICAL UNIVERSITY OF CRETE
FOUNDATION FOR RESEARCH AND TECHNOLOGY-HELLAS

Διπλωματική Εργασία

Τίτλος: Μια διαδικασία ανάλυσης ακτινονομικών δεδομένων στην απεικόνιση του καρκίνου του προστάτη

‘Ονομα μαθητή/ριας : Ευαγγελία Ηλία

Όνομα επιβλέποντα : Κωνσταντίνος Μαριάς

Ημερομηνία Ολοκλήρωσης : 12 Ιουνίου 2024

Acknowledgements

I would like to express my gratitude to my supervisor, Kostas Marias. Also to Katerina Nikiforaki, for her patience and attentiveness. Special thanks to George Manikis for his invaluable input and guidance. I am also profoundly grateful to my parents, my sister, my boyfriend, and my best friends for their support and encouragement throughout this journey. Thank you for believing in me and standing by my side in everything I do.

Abstract

This research aims to answer the question: what is the best preprocessing pipeline for radiomics analysis of prostate cancer MRI images and how can that benefit the performance of a radiomics model. To achieve this ~80 papers were compared on their respective preprocessing pipeline, study design , results and limitations. Through this process a pipeline proposal was formed which included bias field correction, normalization and resampling. This pipeline was then tested on the ProstateX dataset which included MRI scans from 66 patients, prostate segmentations and True/False labels for clinical significance. Following preprocessing the radiomics features were extracted and utilized as the input for model building. After comparing different classifiers, Logistic Regression was selected as a stand out. Hyperparameter tuning was used in order to find the best parameters for the model by utilizing 5-fold repeated stratified cross-validation. The dataset was used to create two models. The first was divided into a train (70%) set and a test(30%) set. The training set was used for the tuning and the training whereas the test set was used only at the end for evaluation of the final model. The model achieved an accuracy of 80.4% on the training, 75% on the test set and an AUC of 0.867 on the training and 0.791 on the test set. To reduce overdiagnosing it is important to focus on the precision metrics too. On the training set an 76.9% precision was achieved compared to a 60% on the test set. This difference means that although on the training set the model was fairly good at avoiding false positives, its performance on the test set was lacking a bit in comparison. The second model divided the dataset into 80% training and 20% hold-out set. The training set was utilized the exact same way as before while the hold-out set was kept aside and used to evaluate the performance of the trained model. This model did perform worse due to the “unseen data” aspect. Precision in the training set was 80% but dropped to 50% on the hold-out and AUC went 92,2% to 68,9%. This finding demonstrates clearly the necessity of creating larger publicly accessible datasets in order to create more reliable models that may eventually be implemented in clinical settings. Conclusively, this research was successful in proposing an effective preprocessing pipeline that achieved notable performance results on the final radiomics model test set but did not do as good on unseen data. Still, this work represents a significant step forward and may pave the way for more studies and future clinical applications.

Περίληψη

Αυτή η έρευνα στοχεύει να απαντήσει στο ερώτημα: ποια είναι η καλύτερη προ-επεξεργαστική διαδικασία για την ανάλυση ακτινονομικών δεδομένων MRI του καρκίνου του προστάτη και πώς μπορεί αυτό να ωφελήσει την απόδοση ενός ακτινομικού μοντέλου. Για την επίτευξη αυτού του στόχου, συγκρίθηκαν περίπου 80 μελέτες ως προς την προ-επεξεργαστική τους διαδικασία, τον σχεδιασμό της μελέτης, τα αποτελέσματα και τους περιορισμούς τους. Μέσα από αυτή τη διαδικασία σχηματίστηκε μια πρόταση για προ-επεξεργασία. Αυτή η διαδικασία δοκιμάστηκε στο σύνολο δεδομένων ProstateX, το οποίο περιλάμβανε MRI σαρώσεις από 66 ασθενείς, τμηματοποιήσεις προστάτη και True/False ετικέτες για κλινική σημασία. Μετά την προ-επεξεργασία, τα ραδιομικά χαρακτηριστικά εξήχθησαν και χρησιμοποιήθηκαν ως είσοδος για την κατασκευή του μοντέλου. Μετά από σύγκριση διαφορετικών classifiers, επιλέχθηκε η Logistic Regression ως η καλύτερη. Έγινε ρύθμιση υπερ-παραμέτρων για την εύρεση των καλύτερων παραμέτρων για το μοντέλο χρησιμοποιώντας five-fold repeated stratified cross-validation. Το σύνολο δεδομένων χρησιμοποιήθηκε για τη δημιουργία δύο μοντέλων. Το πρώτο διαιρέθηκε σε ένα σετ εκπαίδευσης (70%) και ένα σετ δοκιμής (30%). Το σετ εκπαίδευσης χρησιμοποιήθηκε για τη ρύθμιση και την εκπαίδευση, ενώ το σετ δοκιμής χρησιμοποιήθηκε μόνο στο τέλος για την αξιολόγηση του τελικού μοντέλου. Το μοντέλο πέτυχε ακρίβεια 80,4% στο σύνολο εκπαίδευσης, 75% στο σύνολο δοκιμής και AUC 0,867 στο σύνολο εκπαίδευσης και 0,791 στο σύνολο δοκιμής. Για τη μείωση των υπερδιαγνώσεων είναι σημαντικό να επικεντρωθούμε και στις μετρικές ακρίβειας. Στο σύνολο εκπαίδευσης επιτεύχθηκε ακρίβεια 76,9% σε σύγκριση με 60% στο σύνολο δοκιμής. Αυτή η διαφορά σημαίνει ότι, αν και στο σύνολο εκπαίδευσης το μοντέλο ήταν αρκετά καλό στην αποφυγή ψευδώς θετικών αποτελεσμάτων, η απόδοσή του στο σύνολο δοκιμής ήταν λίγο κατώτερη σε σύγκριση. Το δεύτερο μοντέλο διαιρέθηκε σε ένα σετ εκπαίδευσης 80% και ένα σετ επαλήθευσης 20%. Το σετ εκπαίδευσης χρησιμοποιήθηκε με τον ίδιο ακριβώς τρόπο όπως πριν, ενώ το σετ επαλήθευσης κρατήθηκε στην άκρη και χρησιμοποιήθηκε για την αξιολόγηση της απόδοσης του εκπαιδευμένου μοντέλου. Αυτό το μοντέλο απέδωσε χειρότερα λόγω του ότι δοκιμάστηκε σε δεδομένα καινούρια που δεν είχε “ξαναδεί”. Η ακρίβεια στο σετ εκπαίδευσης ήταν 80% αλλά έπεσε στο 50% στο σύνολο επαλήθευσης και το AUC από 92,2% έπεσε στο 68,9%. Αυτό το εύρημα δείχνει ξεκάθαρα την ανάγκη δημιουργίας μεγαλύτερων δημόσια προσβάσιμων συνόλων δεδομένων για τη δημιουργία πιο αξιόπιστων μοντέλων που μπορεί να εφαρμοστούν τελικά σε κλινικές ρυθμίσεις. Συμπερασματικά, αυτή η έρευνα ήταν επιτυχής στην πρόταση μιας αποτελεσματικής προ-επεξεργαστικής διαδικασίας που πέτυχε αξιοσημείωτα αποτελέσματα απόδοσης στο τελικό ακτινομικό μοντέλο, αλλά δεν ήταν εξίσου καλή σε νέα δεδομένα. Παρ' όλα αυτά, αυτή η εργασία αντιπροσωπεύει ένα σημαντικό βήμα για μελλοντικές έρευνες και μπορεί να ανοίξει το δρόμο για περισσότερες μελέτες και μελλοντικές κλινικές εφαρμογές.

Table of contents

Table of Contents

List of figures

List of tables

Chapter 1: Introduction

Prostate cancer (PCa) is the sixth among cancer-related male deaths and is the second most common type of cancer to be diagnosed with the risk of developing it increasing with age. Beginning in 1989, transrectal ultrasound-guided biopsy has been the standard diagnostic tool for prostate cancer however it can cause serious side effects such as hematuria and rectal bleeding. Another important tool for the diagnosis process is Magnetic Resonance Imaging (MRI) as it reduces the need for unnecessary biopsies on low-risk lesions and minimizes over-diagnosing which is a significant problem in prostate cancer. However, it has its own drawbacks as different experts could look at the same MRI scan and come to a different conclusion. Even the same expert looking at the same scan on different days could come to a different conclusion. To address the above issues, radiomics analysis can be a very valuable tool for achieving a more accurate and reliable diagnosis.

The goal of this study is to develop a pre-processing pipeline for prostate cancer MRI images to address the lack of standardization of this step in the radiomics process, aiming to improve the reproducibility and performance of radiomic classification models. The inclusion of a classification radiomics model in a clinical setting could help physicians in the diagnosis process, in decision making and minimize inter-reader variability-related issues.

Therefore for this study the following questions need to be addressed :

- What is the most effective preprocessing pipeline according to the literature ?
- How strong is the performance of the final radiomics classification model?

This research is divided into two phases. The first phase is a literature review which involves a comparative analysis of ~80 studies on their preprocessing pipeline, study design and final result for the purpose of proposing the most effective pipeline. The second phase is the model development which includes the implementation of the preprocessing steps on a prostate MRI dataset, the ProstateX , the extraction of radiomic features and the development of a model for classification on clinical significance.

The thesis begins with a state-of-the-art that discusses relevant literature on prostate cancer, diagnosis methods, MRI and radiomics. Following that, the Methodology chapter describes the research design, the dataset, the preprocessing pipeline proposal, and its implementation for developing a model. Then, the Analysis chapter contains all the details of the research process and the results. The Discussion chapter discusses the research results and compares it to relevant literature. Finally the Conclusion summarizes the findings, their contributions and limitations, and includes suggestions for future research.

Chapter 2: State-of-the-art

2.1 Introduction

Prostate cancer (PCa) is estimated to be the sixth leading cause of cancer related deaths among men while also being the second most commonly diagnosed cancer among them. Statistically, the risk of a man developing prostate cancer increases as he ages with the average age of diagnosis being 67. The risk is higher for those who have family history with the disease. Although a lot of research has gone into alternative accurate and reliable diagnostic methods, transrectal ultrasound-guided biopsy has been the golden standard since 1989. There are, however, drawbacks that accompany an invasive procedure such as a biopsy, most common ones in the case of prostate cancer being hematuria and rectal bleeding.

Due to the increased probability of diagnosis with age, the potential benefits of routine screening have been looked into in order to secure an early-stage diagnosis. Although results show that it could help reduce death rates in certain cases, it also increases the risk of overdiagnosis of low-risk lesions. Loeb and Bjurlin et al. in 2014 compared a number of studies based on epidemiology, clinical factors and biopsy data conducted in order to address overdiagnosis However, the estimates fall between a very wide range from 1.7% up to 67%, highlighting the need of careful screening and treatment planning. Avoiding such drawbacks is vital as low-risk tumors can be harmless. Treating them with aggressive and intensive therapies could cut the patient's life short or simply not make any difference on their lifespan, all while reducing their quality of life.

Studies have been conducted attempting to look into the potential of MRI as a diagnostic tool compared to biopsy. In those, MRI was proven to be noninferior to biopsy in diagnosing clinically significant cancer. With imaging, overdiagnosis was lessened and biopsy was able to be avoided in some cases. Ahmed et al. compared the diagnostic accuracy of MRI and TRUS biopsy and found that utilizing MRI lowers the amount of men that need a biopsy by 27%. Kasivisvanathan et al. conducted a study including 500 men and divided them into two random groups : MRI and TRUS biopsy. In the cases of a positive diagnosis through MRI, a biopsy was performed as well. In the following 30 days the complications reported by the patients were a lot more rare in the MRI group.

Due to this, the suggestion to update guidelines and move towards an image-based diagnostic procedure for prostate cancer has been brought about. However caution is necessary when considering such a change. A notable study in 2018 by Rouviere O. et al. showed that when comparing the diagnostic accuracy of MRI and biopsy, 5.2% of clinically significant prostate cancer would have been missed had there not been a biopsy performed.

Combining MRI and biopsy was found to be the most reliable diagnostic method. This highlights the risk of some patients getting a negative diagnosis despite having cancer. Also, in general, studies often specifically include men with high prostate-specific antigen PSA levels meaning they are more likely to be positively diagnosed and thus the study could be biased. Issues also arise when considering human error which is to be expected especially in fast paced, high stress environments. Low interobserver agreement has been documented in clinical settings which can interfere with decision-making and could possibly lead to unnecessary treatments. It is therefore essential to find additional tools to lower the risk misclassification before moving away from biopsy.

The above highlights the need for a diagnostic approach that could potentially limit the need for invasive biopsy procedures, lessen overdiagnosing and overtreating while also being equally (if not more) reliable in identifying clinical significant prostate cancer. MRI-radiomics and image analysis has been shown to be a strong contender in the search for reliable biomarkers as it promises to limit the need for biopsies and overcome issues that stem from inter-reader disagreement.

2.2 Medical Imaging in Prostate Cancer

Nuclear Resonance Imaging (NMR), or Magnetic Resonance Imaging (MRI) as it is known today, was first used to image the prostate in 1982 using a magnet of 0.08 Tesla (T). With the advancements of technology the hardware and the software were updated over the following decades which led to the introduction of higher field strength of 1.5 and 3.0 T. These advancements have not only increased the image resolution but also reduced the amount of time needed to obtain the image. A magnet of ≥1.5 T is considered necessary for the prostate while most research centers support that 3 T is the ideal field strength.

An endorectal coil is a medical device that can be inserted into a patient's rectum aiding in proper placement of the prostate during an MRI examination. An inflatable balloon attached to a probe lessens local movement in an effort to acquire higher quality images. It was first introduced back in 1989 but as MRI technology has improved, its use has gone down. A lot of research has attempted to conclude whether its use is necessary. It has been shown that especially for 3T it could be omitted and notable results have also been achieved even with 1.5T. The patients' examination experience and comfort has been greatly enhanced as a result of this.

During an MRI a number of sequences are acquired. Biparametric (bp) MRI combines T1- and T2-weighted (T1W, T2W) sequences with Diffusion Weighted Images (DWI). Multiparametric (mp) MRI adds Dynamic Contrast Enhanced imaging(DCE).

- The T2W imaging sequence gives high-resolution pictures of prostate anatomy, exhibiting normal peripheral zone signal intensity as well as cancer signal intensity. However, it has decreased accuracy in the transitional and anterior zones due to lower baseline T2 signal and benign prostatic hyperplasia (BPH) nodules. Functional sequences are necessary because T2W alone lacks the sensitivity and specificity necessary to locate prostate cancer.
- The T1W imaging sequence assists in distinguishing between a tumor and post-biopsy hemorrhage.
- The DWI sequence assesses water diffusion in prostate cancer, which has lower diffusion due to closely packed cells when compared to healthy tissue. When paired with T2WI, apparent diffusion coefficient (ADC) maps yield sensitivity and specificity that are 85–90% higher than those obtained after radical prostatectomy.
- Finally, a T1 sequence, intravenous gadolinium bolus, and fast scans are employed to obtain the DCE image. Cancer symptoms include increased blood flow, neovascularity, and leaky capillaries. As a result, a perfusion vs. time curve is produced and used as a diagnostic tool. There are three types of curves: normal prostate tissue, BPH or prostatitis, and high grade prostate cancer.

As previously stated, bpMRI protocols, in contrast to mpMRI protocols, lack DCE. As a result, it offers three primary advantages: faster examination times, reduced expenses, and an absence of side effects involving contrast agents.

A grading system was created to evaluate MRI images to be utilized as a diagnostic and treatment planning tool. The system is called Prostate Imaging-Reporting and Data System (PI-RADS). The latest version, v2, was published in 2019. According to radiopedia, after obtaining the MRI images, each lesion will be graded on a 1-5 scale that indicates the probability of clinically significant cancer. The grades translate as follows :

PI-RADS 1: very low
PI-RADS 2: low
PI-RADS 3: intermediate
PI-RADS 4: high
PI-RADS 5: very high

PI-RADS v2 prostate MRI offers a high sensitivity and relatively low specificity in identifying clinically significant prostate cancer. A systematic evaluation of 21 trials discovered a sensitivity of 89 percent and a specificity of 73% to confirm a prostate cancer diagnosis. The PRECISION trial found that rates of detection for clinically relevant prostate cancer in men with no prior biopsy ranged from 12 to 60%. However, the reported sensitivities are lower (75-85 percent), implying that MRI-guided biopsy for only PI-RADS 3 lesions or higher could fail to identify some prostate malignancies.

PI-RADS 3 lesions are generally thought to be indicative of clinically significant malignancy, although the importance of these malignancies is contradictory. Depending on the lesion, the diagnostic yield of biopsy varies. Some doctors advocate using PSA density to guide decision-making. An MRI graded as PI-RADS 3 or lower could serve as a justification to avoid biopsy in older patients with substantial morbidities, however this is not accepted practice.

It is evident that although MRI examinations are an important tool for prostate cancer patients, it, alone, cannot be a reliable guide for clinical decision making as its interpretation heavily depends on (and varies based on) radiologists.

The question then becomes whether there is a better imaging method for prostate cancer.

Ultrasound(US) is used to identify and diagnose prostate cancer in the early stages. It can be performed in an office setting, it is widely available, low-cost, and allows for real-time imaging. However, the tissue contrast between malignant and benign tissue is limited. A solution that gets around some of the drawbacks of the separate modalities is mpMRI-US fusion imaging. However, it is relatively expensive and necessitates either fusion-device specialized training or extensive experience. Registration mistakes may also occur during MRI-ultrasound fusion.

Another imaging modality is Computed Tomography(CT). Because of its poor soft-tissue contrast and lack of molecular information, CT is not the primary imaging modality for prostate cancer. It is used to evaluate nodal and distant metastases, but its efficacy is limited when compared to sophisticated hybrid imaging approaches like PET/CT. Furthermore, unlike MRI, it exposes the patient to radiation, which has the potential to cause cancer. Despite its limitations, CT is still part of the recommendations by the American Urological Association for individuals with intermediate- to high-risk PCa.

PET or Positron Emission Tomography provides supplementary data for tumor stage, characterisation, and metastatic involvement. However, it is costly and presents technological (e.g., attenuation correction) and/or clinical problems (e.g., radiation exposure). Hybrid approaches exist like PET/CT, as mentioned before, and PET/MRI. The latter is considered superior due its higher soft tissue contrast and lower radiation exposure. However both modalities are not very widely available and require specially trained professionals.

1.3 Radiomics Analysis

In nuclear medicine in particular and in medical imaging generally, the topic of radiomics has drawn considerable interest. Despite the lack of a precise definition, the objective of radiomics is to extract quantitative information from medical images by recognizing intricate patterns that are difficult for the human eye to interpret or measure.

Radiomics is the study of tumor phenotypes using a wide range of image-derived, quantitative data such as intensity, shape, texture, and so on. Radiomics tools derive image features from immense patient datasets, assessing tumor shape, size, and gray level intensity distribution. They aid in the understanding of the correlations between tumor imaging features, genetic traits, phenotype, and treatment responses. Radiomic examination of tumor subvolumes or habitats offers an imaging measurement for tumor heterogeneity, revealing different tumor cell clones.

Oncology radiomic studies include classification tasks as well as clinical outcome prediction using time-to-event analysis. Classification is the process of categorizing a population into groups such as benign versus malignant, clinically significant versus insignificant, tumor stage, and metastases. Based on clinical outcomes, predictive models classify individuals into risk groups. Therefore, radiomics has the ability to act as a “virtual biopsy” because, unlike traditional biopsies, it uses noninvasive imaging that allows for assessment of the entire tumor and can be applied at various points in time.

As previously stated, the problem with PCa diagnosis is the invasive procedures and lack of ability to entirely rely on MRI imaging. Radiomics may be able to provide a solution to both by doing a "virtual biopsy" using tools that can uncover patterns with diagnostic value that the human eye cannot. Future therapy planning and diagnosis for patients with prostate cancer may take a different path attributable to the extraction of features from MRI images.

The extraction of high-dimension feature data to describe attributes of ROIs is the core of radiomics. In actuality, radiomics extracts two kinds of features: "semantic" and "agnostic" features. Semantic features are frequently utilized in the radiology vocabulary to describe ROIs as they are linked to anatomical or physiological properties of the imaged tissue.
Agnostic features, on the other hand, are data-driven and aim to identify lesion heterogeneity using quantitative descriptors. They are derived without prior knowledge of the underlying anatomical or physiological properties.

Some commonly calculated features in radiomics analysis are :

- Shape Features:
  Volume: The space occupied by a region of interest in the image.
  Surface Area: The total area of the boundary surface of a structure in the image.

- Intensity Features:
  Mean Intensity: The average pixel intensity value within a defined region.
  Standard Deviation: A measure of the spread or dispersion of intensity values within a region.

- Texture Features:
  Entropy: A measure of randomness or disorder in pixel intensities.
  Contrast: The difference in intensity between neighboring pixels.
  Homogeneity: Describes the similarity of intensity values in an image region.

- Statistical Features:
  Skewness: Indicates the asymmetry of the intensity distribution.
  Kurtosis: Measures the "tailedness" of the intensity distribution.

- Histogram-Based Features:
  Percentile: The value below which a given percentage of intensity values fall.
  Mean Absolute Deviation: The average absolute difference between each intensity value and the mean.
- Spatial Features:
  Gray-Level Co-occurrence Matrix (GLCM) Features: Descriptive statistics capturing relationships between pixel pairs in an image.
  Run Length Matrix Features: Quantifies the length and occurrence of homogeneous runs of pixels.

- Wavelet Features:
  Wavelet Energy: Represents the energy distribution in different frequency components of the image.

1.4 Preprocessing in Radiomics

Harmonization, though less extensively researched in MRI than in PET and CT, is important because of MRI's technical limitations and dependence on a number of parameters. As far as possible, inconsistencies between images that are directly related to the acquisition process must be removed in order for radiomics to be able to extract meaningful information. The technical difficulties with MRI include non-standard signal intensities that can be affected by differences in scanner models, coils, sequence types, acceleration and acquisition parameters. Additionally, all vendors will introduce noise to the images, meaning signal variability that is not part of the desired signal, to some extent. As a result, numerous solutions to this problem have been developed and are discussed in the following paragraphs.

- *Normalization:* Signal intensity normalization is used to account for inter-scanner and inter-patient variations by adjusting the range of signal intensities within a Region Of Interest (ROI). This is accomplished by either transforming the ROI histogram to match a reference signal intensity histogram or by computing the mean and standard deviation of the signal intensity gray-levels within the ROI. No official guidelines have

been created, despite the fact that several articles in the imaging literature have underlined the fundamental value of intensity normalization. In 2014, Russell T. Shinohara and colleagues took the initiative to propose a set of seven principles, dubbed the Statistical Principles of Image Normalizing (SPIN), with the aim of defining the normalizing process and establishing specific objectives. According to their proposal, the normalization process should result in images 1. that have a consistent interpretation across multiple locations throughout the same tissue, 2. are repeatable, 3. maintain the intensity ranking, 4. have comparable distributions of the same ROIS within and between patients, 5. are unaffected by biological anomalies or variations in population, 6. are not impacted too much by noise or artifacts, and 7. do not result in loss of diagnostic information.

- *Interpolation:* Radiomics analysis commonly incorporates image interpolation, which is a method of image resizing through upsampling and downsampling. This improves texture feature extraction and assures an unbiased representation of spatially related radiomics features. With this method a series of high-resolution organ or tissue images is obtained , which is commonly used in medical applications in imaging modalities such as CT and MRI. Heterogeneous voxel sizes, structural breaks and even surfaces include problems that arise due to the fact that the distance between slices is usually larger than the pixel size. By interpolating multiple slices, this method helps generate volume data with isotropic dimensions, and it solves image resolution problems by providing a detailed and accurate picture of the target structures.

- *Bias field correction:* The bias field is a low signal that can blur the MRI image and make them inhomogeneous. This bias field varies not only among centers and vendors, but also between patients, even when using the same vendor or acquisition parameters. The bias field reduces repetitive image features such as edges and patterns and it also changes the intensity values of the image pixels, resulting in different gray level distributions in the image for the same tissue. This not only impacts the accuracy of image processing algorithms that rely on gray-level values of pixels, but also any method that depends on the assumption of spatial invariance in the processed image. Before inputting MRI images into any algorithms, it is necessary to first apply a correction pre-processing step to account for the bias field. There are several methods of correction, the most common being the N4 Bias Field Correction. With its fast execution and multiresolution process it effectively addresses the image inhomogeneities.

- *Discretization:* Discretization is another step in the pre-processing of MRI images which converts the original intensity values into a set of gray levels. This process then results in what is called an intensity histogram that is made up of intensity bins. Two

types of discretization methods exist. In the Fixed Bin Number (FBN) method the number of bins is predetermined, but the bin width may vary depending on the range of the data. The goal is to divide the data into a pre-specified number of bins. On the other hand, in the Fixed Bin Size (FBS) method, the width of each bin is the predetermined value and it remains constant throughout the histogram. No hard evidence has been found to favor one method or the other yet. However the selection of the ideal values is crucial as these methods were developed for weighted MRI data with no reference point or voxel value for intensity standardization. As a result, even seemingly insignificant changes in discretization method intensity range and gray level count have been shown to have a negative impact on the reproducibility of the resulting radiomic features.

- *Registration and alignment:* The process of aligning images spatially with one another is known as image registration. Images can be registered into the same coordinate system to produce fusion images and enable a variety of quantitative analyses. In medical imaging, registration serves the purpose of aligning multiple images to ensure anatomy correspondence. Algorithms for registration can be classified as linear or non-linear. Non-linear registration allows for local deformation with elasticity, whereas linear registration includes rigid or affine transformations.

- *Noise reduction:* Noise artifacts in MRI images include body temperature, subject time, and thermal variables. The subject's time within the MR machine is inversely correlated with the thermal factor of the machine, and extended exposure can raise body temperature. MRI noise can visually deteriorate images and lead to a false diagnosis. If a particular tissue or location has a low signal to noise ratio (SNR), it also impairs quantitative imaging and reduces the usefulness of MRI. For this reason, improving both qualitative and quantitative metrics requires an effective MRI reconstruction procedure that makes use of denoising techniques. Although some sources use the phrase more generally to mean anything that eliminates noise, noise reduction—also referred to as noise suppression or denoising—commonly refers to the numerous algorithmic ways to reduce the aforementioned noise in image files once they are formed. Various approaches, primarily filtering techniques, are applied to images including morphological filters, statistical filters, frequency filters (discrete Fourier transform), and spatial filters (convolutions).

1.5 Future Directions and Challenges: A Look into Literature

In 2020, Simon Bernatz et al. conducted a comparative analysis of many machine learning algorithms that used radiomics and clinical parameters to predict clinically-significant PCa.

Regarding pre-proprecessing, they make reference to a crucial point that applies to all cancer studies, not just those on prostate cancer. They state that the Imaging Biomarkers Standardization Initiative (IBSI) does not cover image preprocessing. Although the IBSI does explain the definitions and the general consensus of a lot of preprocessing techniques, it does not recommend which to use first, which to use next, or even which techniques to use at all. It does urge the readers to include thorough explanations of the preprocessing steps of their research to counter the difficulty in reproducing radiomics studies. To ensure integrity and comparability, the researchers in the aforementioned work chose to use unaltered images without any preprocessing.

A number of other papers have also either skipped the pre-processing step altogether or performed limited preprocessing.

In 2019, Florent Tixier et al. explored the potential of preoperative MRI-radiomics features to enhance the prediction of survival in glioblastoma patients. They admit to a possible lack of prognostic value for their research due to a lack of preprocessing. However, they clarify that this was a purposeful decision to allow their findings to be used in a clinical context.

In 2022, Cui Feng et al. created a radiomics nomogram for grade prediction of Bladder Cancer. They went with a zero preprocessing method but did mention that MRI scanners can cause inhomogeneities and background interferences. It is their opinion however that those systematic errors cannot be completely eliminated.

In 2023, Mohammad Mirza-Aghazadeh-Attari et al. studied the additive value of radiomics features in staging Hepatocellular Carcinoma. Although they did perform some minor preprocessing (registration and normalization), they referred to it as a possible contributor to reduced reproducibility.

In 2023 Maria-Fatima Chilaca-Rosas et al. studied the diagnostic performance of selected MRI-derived radiomics in distinguishing progression-free and overall survival in patients with midline glioma and the H3F3AK27M mutation. They chose to limit the data to only 2 scanners because any more scanners require harmonisations due to different acquisition parameters and variations, termed the "scanner effect". They also refer to bias field correction methods as a promising solution to this and highlight the need for standardization of such methods in order to be used for future research.

Finally, a notable mention would be a paper by Sandra Fiset et al that,in 2019, studied the repeatability of radiomics features in cervical cancer. A 4-step preprocessing pipeline was performed on the images. However, lack of bias field correction was mentioned as a

limitation and the writers underlined the need for research on the field's effect on reproducibility.

The evaluated studies stress the pivotal role of image preprocessing in radiomics research, recognizing its diverse methodologies and impact on clinical relevance. With initiatives like IBSI lacking specific guidelines, researchers must detail preprocessing methods for enhanced reproducibility. Some studies intentionally opt for minimal preprocessing for clinical utility, emphasizing the need for standardization in techniques to ensure reliable and comparable outcomes in future radiomics research.

1.6 Radiomics in Prostate Cancer Diagnosis: Bridging Gaps, Addressing Challenges, and Shaping the Future

Following an examination of existing literature in radiomics, it is clear that the approaches used in preprocessing have a major impact on the reproducibility and clinical relevance of research. The lack of precise criteria by organizations such as the Imaging Biomarkers Standardization Initiative (IBSI), emphasizes the need for researchers to be transparent and disclose their preprocessing methods in order to improve the reproducibility of their findings. While some studies purposefully use little or no preprocessing to coincide with clinical settings, others consider the implications for prognostic value and reproducibility. In order to ensure strong and reproducible results for future radiomics research, there needs to be for standardization of the preprocessing pipeline.

As previously discussed the typical transrectal ultrasound-guided biopsy, despite being a historical gold standard, has a number of limitations such as invasiveness and associated problems. The exploration of routine screening aims to acquire early-stage diagnoses, but the risk of overdiagnosis of low-risk lesions arises further concerns regarding unnecessary treatments. Recent studies comparing the diagnosis accuracy of MRI versus biopsy reveal that MRI has the potential to be a non-inferior option in finding clinically significant cancer. The use of MRI and diagnostic techniques shows promise in terms of reducing the need for invasive biopsies and minimizing associated consequences.

Despite the above, caution is advised when relying on image-based diagnostic procedures. Studies show potential limitations, including missed diagnoses which creates the need of combining both MRI and biopsies to improve reliability of the results. The limitations of MRI are highlighted further by human error, low inter-reader agreement, and potential biases in study populations. Due to this, the use of MRI-radiomics as predictive biomarkers is a promising option for decreasing invasive procedures and overdiagnosis and, also, improving the precision of diagnosing prostate cancer. As research on radiomics continues, a balance between quality of life, clinical relevance, study reproducibility, and diagnostic accuracy will be critical factors in determining the best practices.

Chapter 3: Research methodology

3.1 Introduction

In this chapter the methodology of the research will be described step-by-step. It begins with an overview of the entire process and then the following paragraphs go into more depth on the specifics of each step.

3.2 Overview

This research is divided in two phases, a comprehensive comparative literature review followed by an experimental application and validation phase. The design aims to compare existing preprocessing pipelines and develop a pipeline proposal for prostate cancer MRI images. The rationale behind this design is to take advantage of established methodologies, create a proposal and test it.

➢ *Phase 1: Comparative Literature Review*

In the first phase, a thorough comparative analysis of approximately 80 research papers was conducted. This review included:

- *Scope and selection criteria:* ~ 80 papers were selected on the premise that they specified the preprocessing steps they used on MRI images. Half of those were dedicated to prostate cancer while the rest to different cancers. This was done due the limited amount of literature on prostate cancer that explicitly discussed preprocessing in order to obtain a larger view of the recent practices on preprocessing pipelines.
- *Analysis Parameters:* The selected papers were compared on multiple parameters namely, cohort size, multi/single center design, multi/single vendor design, evaluation metrics and limitations. This comparison aimed to identify preprocessing pipelines used in the studies with the best design and results and with the minimum limitations.
- *Outcome:* Based on this comparative analysis, a preprocessing pipeline for prostate cancer MRI images was proposed, and it included the most effective and widely used techniques identified in the literature.

➢ *Phase 2: Experimental Testing and Validation*

The proposed preprocessing pipeline was implemented and validated, according to the following steps:

- *Dataset selection:* The ProstateX dataset, public prostate cancer MRI dataset was selected.
- *Pipeline Application:* The proposed preprocessing pipeline was applied to the ProstateX dataset, and preprocessing and model-building were fully implemented in Python.

- *Model Development:* After the extraction of the radiomics features they were used as input to create a radiomics classification model to predict clinically significant prostate cancer. A number of classifiers were compared and evaluated on their performance. The best one, Logistic Regression, was used to develop the final model.
- *Training and Validation:* The dataset was divided into a train set, used for tuning and training and a test set used for final evaluation (70/30) to create a train-test model. It was also divided into 80/20 for a train and hold-out model where the latter was kept as unseen data.

3.3 Establishing the Preprocessing Pipeline

In the literature review the papers were divided into sub-groups with similar objectives. 1-2 papers of each group were selected for standing out based on the results and the overall design of the study. Then from those the final proposal of the preprocessing steps was concluded upon.

Following the comparative literature review, the following pipeline was established :

- *Bias field correction* was applied to correct intensity non-uniformities within the MRI images. The N4ITK algorithm, implemented using SimpleITK in Python, was utilized for this correction. This step is important for enhancing the homogeneity of the images, in order to improve the reliability of intensity-based features.
- *Z-score normalization* was used to lessen the impact of differences in images and patients on feature extraction by calculating the mean and standard deviation of image intensities and adjusting pixel values accordingly.
- *Resampling* was used to ensure that all pictures and segmentations shared the same voxel size and dimensions. SimpleITK was utilized to perform resampling, with segmentation masks serving as the reference size.

3.4 Dataset Description

This study employed the ProstateX dataset, which is a large, publicly available collection intended exclusively for prostate cancer imaging studies. It is offered via the SPIE-AAPM-NCI Prostate MR Image Cancer Detection Challenge.

➢ *Prostate X Dataset:*

- *Source:* The Cancer Imaging Archive (TCIA) hosts the ProstateX dataset.
- *Content:* The dataset contains multiparametric MRI (mpMRI) scans from 346 people. The imaging modalities include T2-weighted (T2W) MRI, diffusion-weighted imaging

(DWI), dynamic contrast-enhanced (DCE) MRI, and proton density-weighted (PD-W) MRI. We used the T2W modality for the analysis.

- *Annotations:* The dataset includes annotations from radiologists of lesion locations coordinates, and clinical significance based on biopsy Gleason score. However, segmentation masks are only provided for 66 patients, which is critical for radiomics feature extraction. Each patient may have one or more lesions, located in their prostate.

➢ *Dataset Utilization :*

- *Preprocessing Pipeline Application:* The proposed preprocessing pipeline, developed from the comparative literature review, was applied to the ProstateX dataset.
- *Label Aggregation:* The dataset was imbalanced (approximately a 2:1 ratio of False to True labels, 43:23) and there needs to be sufficient data for training, so for this reason labels were aggregated instead of creating separate models for each prostate region. Initially, labels were lesion-based, with each patient having 1-5 labels corresponding to individual lesions located in the same or different prostate zones.If a patient had at least one True (= clinically significant) lesion, the patient was assigned a True label. Patients with only False labels on their lesion(s) were given a False label. This way all available data was used to build the prediction model although it resulted in the loss of the spatial information.
- *Feature Extraction:* The segmented prostates were utilized to extract the radiomics features. including shape, texture and intensity, using PyRadiomics.
- *Model Development and Validation:* The extracted features were used as input for developing the classification radiomics model for clinical significance.

3.5 Model Development

Pipeline Design

The model was developed wth the Pipeline class from scikit-learn and included the following For standardization, the StandardScaler was used to scale the features with a mean of 0 and a standard deviation of 1. For feature reduction, Principal Component Analysis (PCA) was used to reduce the feature space. For feature selection , a Random Forest classifier was used to select the most relevant features for the classification task. For a classifier, after comparing multiple classifiers on their performance for the dataset, Logistic Regression was selected as the best one.

Hyperparameter Optimization

Hyperparameter tuning was performed using RandomizedSearchCV. The parameter distributions included:

- Number of PCA components.
- Number of estimators in the Random Forest.
- Regularization strength in Logistic Regression.

Cross-Validation

To ensure strong performance metrics, repeated stratified k-fold cross-validation was employed with 5-folds, in order to address class imbalance by maintaining the same proportion of each class in each fold. It was utilized during the hyperparameter tuning of the model with the training set.

Model Training and Validation

- Data Splitting: The dataset was split into training and test sets (70-30) and also into training and hold-out sets (80/20)
- Evaluation: The first model was tuned on the training set and trained also on the training set. The final evaluation was done on the test set. Optimal classification thresholds were calculated using precision-recall curves and performance metrics were calculated on both the training and test sets including accuracy, precision, recall, F1-score and AUC. For the second mode, the exact process was followed with the only difference being the hold-out set being kept as unseen data.

Chapter 4: Analysis

4.1 Introduction

In this section, the Analysis, the implementation of the previously described methodology will be presented and analysed. The first two paragraphs will compare prostate cancer and multiple cancer papers, respectively, on their preprocessing pipelines and their results. Following that, the final preprocessing proposal is explained. Then, the remaining paragraphs in this section will discuss the implementation of the pipeline and the consecutive creation of a classification model.

4.2 Grouping the studies on Prostate Cancer based on their objective

Group #	Description	# of Papers
1	**Predicting Clinically Significant Prostate Cancer** - Focused on developing and validating models to predict clinically significant prostate cancer.	9
2	**Prostate Cancer Detection and Characterization** - Concentrates on prostate cancer detection and characterization through various imaging techniques and radiomics.	10
3	**Radiomics Model Generalization** - Examines the generalizability of radiomics models across different datasets and scenarios.	5
4	**Prostate Cancer Detection and Machine Learning** - Explores the use of machine learning and imaging classifiers in prostate cancer diagnosis.	8
5	**PI-RADS 3 Lesion Characterization** - Specifically targets the characterization of PI-RADS 3 lesions, aiming to distinguish between benign and malignant cases.	7

1. Grouping of Prostate Cancer studies

Group 1 : Predicting Clinically Significant Prostate Cancer (csPCa)

#	doi	Title and Date	# of Patients	Pre-Processing	Classifier
1	10.3389/fonc.2022.918830	“Prediction of clinically significant prostate cancer with a multimodal MRI-based radiomics nomogram”, **2022**	201	normalization, resampling	LR
2	10.2147/qi	“Radiomics prediction model for the improved diagnosis of	381	normalization, gray-level quantisation,	LR

	ms.2019.12.06.	clinically significant prostate cancer on biparametric MRI", **2020**		resampling	
3	10.1002/jmri.25983	"Radiomic features on MRI enable risk categorization of prostate cancer patients on active surveillance: Preliminary Findings", **2018**	56	resampling, bias field correction, drift correction	QDA, RF and SVM
4	10.3390/app10010338	"A Hybrid End-to-End Approach Integrating Conditional Random Fields into CNNs for Prostate Cancer Detection on MRI", **2020**	344	interpolation, registration, z-score normalization	CRF-CNN
5	10.3390/jimaging7100215	"A Combined Radiomics and Machine Learning Approach to Distinguish Clinically Significant Prostate Lesions on a Publicly Available MRI Dataset", **2021**	299	resampling, z-score normalization, discretisation, Laplacian Gaussian filtering and wavelet decomposition	NB, KNN and RF
6	10.3390/cancers13246199	"Prediction of Clinically Significant Cancer Using Radiomics Features of Pre-Biopsy of Multiparametric MRI in Men Suspected of Prostate Cancer", **2021**	200	resampling, normalization, co-registration	LR
7	10.1016/j.media.2021.102155	"End-to-end prostate cancer detection in bpMRI via 3D CNNs: Effects of attention mechanisms, clinical priori and decoupled false positive reduction", **2021**	1950	b-spline interpolation, normalization	CNN
8	10.3389/fonc.2021.792456	"MRI Based Radiomics Compared With the PI-RADS V2.1 in the Prediction of Clinically Significant Prostate Cancer: Biparametric vs Multiparametric MRI", **2022**	204	normalization, b-spline interpolation, discretisation	SVM
9	10.3390/cancers14010012	"Classification of Clinically Significant Prostate Cancer on Multi-Parametric MRI: A Validation Study Comparing Deep Learning and Radiomics", **2021**	644	resampling, registration, z-score normalization	LR, SVM, RF, NB, LDA and QDA

2. Group 1 : Predicting Clinically Significant Prostate Cancer (csPCa)

Group 1 includes papers with the end goal of detecting clinically significant prostate cancer (csPCA). When examining the sample size Paper 7 stands out. It boasts the largest patient cohort, featuring 1950 individuals. This multicenter study, involving a single vendor, aimed to create a multi-stage 3D CAD model for automated localization of clinically significant prostate cancer. Paper 2 follows with 381 patients. This is a single-center, single-vendor study . Paper 4 features a substantial cohort of 344 patients, also adopting a single-center,

single-vendor approach. Paper 5 utilizes a single-center, single-vendor dataset with 299 patients. Paper 8 conducted in a single-center, single-vendor setting, included a cohort of 204 patients. Paper 1 features 201 patients within a single-center, multivendor setup. Paper 6 follows with a single-center, single-vendor approach and includes 200 patients. Finally, Paper 3 stands out with the smallest cohort of 56 patients, maintaining a single-center, single-vendor approach. Due to its smaller size its findings may be less transferable to other clinical settings due to its limited diversity.

Regarding preprocessing, papers 1 through 6 and paper 9 apply normalization and resampling. Papers 2 and 12 additionally employ gray-level quantization. Paper 5 has a more extensive approach, with z-score normalization, discretization, Laplacian Gaussian filtering, and wavelet decomposition. Paper 3 applies bias field correction and drift correction. Papers 6 and 9 include registration as a preprocessing step. Paper 7 involved b-spline interpolation and normalization as did Paper 8 along with discretisation.

In terms of results, Paper 2 and Paper 1 achieved the best metrics. Paper 2 achieves an impressive AUC of 0.98 for both the radiomics model and the clinical-radiomics combined model, performing better than the clinical model significantly. Paper 1 an AUC of 0.942, showing the nomogram's potential in performing better than subjective evaluation. Paper 5 follows closely with accuracies of 80% and an AUC of around 0.80 . Paper 3, despite having a smaller cohort, achieved a significant overall accuracy improvement of up to 80% when compared to PIRADS v2.0 alone. In Paper 6 radiomic features outperformed PIRADS and PSAD by 35.0% and 34.4% in predicting clinically significant prostate cancer. For Paper 7, the results were promising, with M1 detecting clinically significant prostate cancer with a low false positive rate the highest detection sensitivity on the testing datasets, with an AUC of 0.836. In the 8th study, both the radiomics model based on bpMRI and mpMRI signatures demonstrated high predictive efficiency, although there were no significant differences between them. (AUC = 0.975 vs 0.981 in the training cohort, and 0.953 vs 0.968 in the testing cohort, respectively) Importantly, both models outperformed the PI-RADS v2.1 scoring system in diagnosing csPCa. Finally, Paper 9, compared a radiomics and a deep learning model. The results showed notable differences between the two approaches. The radiomics model achieved AUCs of 0.88, 0.91, and 0.65 on independent test sets, while the deep-learning model achieved AUCs of 0.70, 0.73, and 0.44 on the same test sets.

Each paper in this group utilizes a single-vendor, single-center approach, except for Paper 1, which uses a multi vendor dataset and paper 7 which utilized a multicenter setting. Despite their achievements, each study has its specific limitations that can affect the generalizability of the results . Paper 1, despite notable results, used a relatively small dataset. Paper 2, although it did utilize a bigger dataset, is a single-center study. Paper 3, which focuses on active surveillance patients, leaves out cases with a PIRADS score of 3, which can impact its

ability to be broadly applied. Paper 4 introduces high variability in performance, and Paper 5, despite achieving good results, does not conclude on a clear link between Gleason grading and clinical outcomes. Additionally, Paper 6 did manual segmentations and it is a single-center study. Paper 7 although it did include a very high number of patients from various institutes, it only included data from one vendor which could affect generalizability. Paper 8, noted that they did not distinguish between PCa occurring in the peripheral zone (PZ) and transition zone (TZ) which could affect the possible applications and generalizability again. Lastly, Paper 9, did not account for clinical data and co-existing benign prostatic diseases and relied on a single clinician for reference, which could introduce biases.

Group 2 : Prostate Cancer Detection and Characterization

#	doi	Title and Date	# of Patients	Pre-Processing	Classifier
1	10.3390/cancers12082200	"Combination of Peri-Tumoral and Intra-Tumoral Radiomic Features on Bi-Parametric MRI Accurately Stratifies Prostate Cancer Risk: A Multi-Site Study", **2020**	231	resampling , interpolation, drift correction, bias field correction	QDA
2	10.1007/s13246-021-01022-1	"Bi-parametric magnetic resonance imaging based radiomics for the identification of benign and malignant prostate lesions: cross-vendor validation", **2021**	459	z-score normalization, resampling	RF, SVM and LASSO
3	10.1007/s00330-020-07227-4	"Advanced zoomed diffusion-weighted imaging vs. full-field-of-view diffusion-weighted imaging in prostate cancer detection: a radiomic features study", **2021**	136	normalization, discretisation	LASSO
4	10.1002/acm2.12992	"Voxel-based supervised machine learning of peripheral zone prostate cancer using noncontrast multiparametric MRI", **2020**	17	bias correction (N4ITK), noise reduction, standardization, co-registration	SVM
5	10.18383/j.tom.2018.00033	"Gleason Probability Maps: A Radiomics Tool for Mapping Prostate Cancer Likelihood in MRI Space", **2019**	48	normalization, alignment	N/A
6	10.26502/jrci.2809061	"Radiomic Features on Prostatic Multiparametric Magnetic Resonance Imaging Enable Progression Risk in Patients on	55	normalization	LR

		Active Surveillance: A Pilot Study", **2022**			
7	10.1109/ISBI.2019.8759217	"Classification of Prostate Cancer : Low Grade vs. High Grade using a Radiomics Approach", **2019**	40	resampling, registration	SVM
8	10.1038/s41598-019-45766-z	"Repeatability of Multiparametric Prostate MRI Radiomics Features", **2019**	15	bias field correction(N4), normalization, discretisation, registration	N/A
9	10.1002/jmri.25562	"Radiomic features for prostate cancer detection on MRI differ between the transition and peripheral zones: Preliminary findings from a multi-institutional study", **2017**	80	B-spline elastic registration, computationally analyzed at the same resolution, correction for acquisition artifacts (intensity drift and bias field correction - N3 - for endorectal coil)	LR
10	10.3389/fonc.2021.718155	"A Fully Automatic Artificial Intelligence System Able to Detect and Characterize Prostate Cancer Using Multiparametric MRI: Multicenter and Multi-Scanner Validation", **2021**	131	de-noising (Gaussian filter), N4 bias correction, intensity normalization, interpolation, discretisation	SVM

3. Group 2 : Prostate Cancer Detection and Characterization

Group 2 includes studies about the detection and characterization of prostate cancer. Paper 2, aims to develop a multi-center, multivendor radiomics model for prostate cancer risk stratification and has the largest dataset of 459 patients. Paper 1, focuses on development of a radiomics model for prostate cancer detection through a multicenter, multivendor study, involved 231 patients. In contrast, Paper 3 utilized a smaller cohort of 136 patients and was limited to a single-center, single-vendor setup. Paper 4 also has a single-center, single-vendor dataset with only 17 patients in an attempt to characterize clinically significant prostate cancer. Paper 6 includes 55 patients and a single-center , single-vendor approach while aiming to identify imaging biomarkers for prostate cancer diagnosis. Meanwhile, Paper 7, focuses on radiomics features for distinguishing between prostate cancer and benign prostatic hyperplasia, has one of the smallest cohorts, 40 patients, and also has a single-center, single-vendor design. Paper 5, has a single-center, single-vendor design with 48 patients and focuses on generating new image contrasts by learning unique image signatures associated with prostate cancer. Paper 8, wants to assess the repeatability of radiomics features utilizing the smallest dataset (15 patients) with a single-center, single-vendor approach. Paper 9 aims to assess whether radiomic features for prostate cancer detection from 3 Tesla mpMRI differ between the transition zone (TZ) and peripheral

zone (PZ). This multicenter, multivendor study involved 80 patients. Lastly, Paper 10 wants to develop and validate a fully automated computer-aided diagnosis (CAD) system for the detection and characterization of prostate cancers based on their aggressiveness. This multicenter, multivendor study included 131 patients .

The selected papers employ various preprocessing techniques to enhance the quality and consistency of their data. Normalization is the only preprocessing utilized in Paper 6, Papers 2, 3, and 5 also employ normalization but also additional steps too. Paper 2 includes resampling, while Paper 3 includes discretisation. Paper 5 includes alignment. Interpolation is employed in Paper 1 along with drift correction and bias field correction. Paper 7 employs resampling and registration. Paper 4, utilizes bias field correction , noise reduction, standardization and co-registration. Paper 8, on the other hand, performed bias field correction (N4), normalization, discretization, and registration. Paper 9 utilizes B-spline elastic registration, drift correction and bias field correction (N3 - for endorectal coil). Paper 10 , with the most extensive approach, includes de-noising (Gaussian filter), N4 bias correction, intensity normalization, interpolation and discretisation.

When considering the results, Paper 2 had the strongest performance. Its biparametric radiomics model performed better than single-parametric models, achieving an AUC of 0.833. The comprehensive diagnostic model achieved an AUC of 0.911. Paper 1 also shows robust performance, with an AUC of 0.87, especially when peri- and intra-tumoral radiomic features were combined. Paper 4 achieved an e AUROC of 0.93 for prostate cancer detection using T2WI, DWI, and DTI models, however it used a small dataset of only 17 patients. Paper 3 achieved AUCs of 0.93 and 0.94 in its mp-MRI model and mixed model, respectively, despite its relatively smaller cohort size. Paper 6 also has a limited sample size, but its radiomic shape feature extracted from DWI maps achieved an AUC of 0.76 for predicting progression to clinically significant prostate cancer. Paper 5 developed stable Gleason probability maps that outperform conventional clinical imaging (AUC = 0.79). Paper 7 achieved an AUC of 0.77 for classifying high-grade and low-grade prostate cancer lesions. Paper 8, showed significant variability in the repeatability of radiomics features. The authors suggested to not rely on prior studies for selection of radiomics features due to the impact of image type, preprocessing, and region of interest on repeatability and the study was unable to conclude on any universal features or preprocessing pipelines. Paper 9 found that a zone-aware classifier significantly improved cancer detection accuracy in the PZ compared to a zone-ignorant classifier, with AUC values ranging from 0.61 to 0.71 when evaluated on MRI data from multiple institutions. Lastly, in Paper 10, the CAD system achieved a high ROC curve with an AUC of 0.96 in distinguishing between low and high-aggressive tumors in the training set and an AUC of 0.81 in the validation set.

However, each paper has its set of drawbacks. Paper 1 points to its small validation set, which could affect the reliability of its results. Paper 2, despite its strong performance,has limitations due to its single-center and limited-vendor approach, which could affect its generalizability. Paper 3 had a small dataset for the the insignificant prostate cancer category, which could limit its predictive ability. Paper 4 had a small number of patients as well and relied only on radiologists for classification and validation. Also, it focused only on intermediate and high-grade tumors, limiting its application to all grades. Paper 6 limitations include a single-center, single-vendor design and the lack of DCE sequences. Moving to Paper 7, it relied on Gleason score as the ground truth, which can vary between pathologists. Furthermore, it was conducted in a single-center, single-vendor setting, and tumor locations are not included as features. Paper 5 had unlabeled non-cancer confounding diseases in its dataset, which may lead to errors. Additionally, its single-vendor approach and the use of an endorectal coil might affect its generalizability. Paper 8, even though it couldn't establish universally stable feature and preprocessing recommendations, it found specific features with high repeatability that could be considered for radiomics signatures, and suggested further research to assess their predictive power on different datasets. Paper 9 , due to its multicenter, multivendor design had limitations due to variations in MRI acquisition parameters and the unavailability of ground truth. Paper 10 completely left out TZ tumors. It, also, categorized all aggressive tumors as non-indolent, potentially leading to over-treatment or additional investigations.

Group 3 : Radiomics Model Generalization

#	doi	Title and Date	# of Patients	Pre-Processing	Classifier
1	10.3390/diagnostics11020369	“A Multi-Center, Multi-Vendor Study to Evaluate the Generalizability of a Radiomics Model for Classifying Prostate cancer: High Grade vs. Low Grade”, **2021**	204	registration, resampling, ComBat	WORC*
2	10.1186/s13244-021-01099-y	“Single-center versus multi-center biparametric MRI radiomics approach for clinically significant peripheral zone prostate cancer”, **2021**	262	normalization , gray level discretisation	XGBoost
3	10.21203/rs.3.rs-180726/v1	“Integration of Clinical identifications With Deep Transferrable Imaging Feature Representations Can Help Predict Prostate Cancer Aggressiveness and Outcome”, **2021**	1442	normalization	KNN, AB, RF, LR and SVM

4	10.3390/jcm12010140	"A Framework of Analysis to Facilitate the Harmonization of Multicenter Radiomic Features in Prostate Cancer", **2022**	210	normalization, b-spline interpolation (isotropic resampling), discretisation	LogitBoost, RF, KNN, and DT
5	10.1007/s13246-019-00720-1	"An inter-center statistical scale standardization for quantitatively evaluating prostate tissue on T2-weighted MRI", **2019**	51	bias correction (N4ITK), noise reduction, intensity standardization	N/A

4. Group 3 : Radiomics Model Generalization

*Workflow for Optimal Radiomics Classification (WORC) platform, an open-source machine learning software specifically designed for radiomics applications.

This group focused on generalizability. Paper 3 had the largest and most comprehensive dataset among the group, with 1442 patients across multiple centers but a single-vendor design. Its goal is to develop a generalizable machine learning platform for PCa Gleason grade and PIRADS prediction. Paper 4 also has a multicenter, single-vendor design with 210 patients to develop a framework for harmonizing radiomic features extracted from T2-weighted MRI. Paper 1, also involved multiple centers and also multiple vendors with 204 patients, and focuses on evaluating the generalizability of radiomics models for prostate cancer classification. Paper 2 had a dataset of 262 patients, and compared multi-center, multi-vendor data to single-center, single-vendor data for the classification of clinically significant PZ prostate cancer. Finally, Paper 5, comprising the smallest dataset, aims to assess different candidate biological reference tissues for standardizing T2-weighted MRI intensity distributions. It is a multicenter, single-vendor study involving 51 patients.

Turning to preprocessing, Papers 3, 2 and 4 employ normalization. On top of that, Paper 2 and Paper 4 employ discretisation, with Paper 4 also utilizing b-spline interpolation. Paper 1 uses registration, resampling and ComBat. Finally Paper 5 employs N4 bias correction, noise reduction and intensity standardization.

Regarding results, Paper 3 stands out by integrating clinicians' prior identifications with deep transferable imaging feature representations, demonstrating promising performance in risk stratification for PCa Gleason grade. Paper 4's use of ComBat for harmonization achieves 70% accuracy and 78% AUC, which is notable. Paper 2 exhibits a significant performance reduction when transitioning from a single-center, single-vendor dataset (ScSv - AUC=0.82) to a multi-center, multi-vendor (McMv - AUC = 0.75)dataset, highlighting the challenges of data heterogeneity. Still the McMv model achieves a notable result. Paper 1 evaluates radiomics models against radiologists. The three single-center models obtained a mean AUC of 0.75, (which decreased to 0.54 when the model was applied to the external data), the radiologists obtained a mean AUC of 0.46. In the multicenter setting, the radiomics model

obtained a mean AUC of 0.75 while the radiologists obtained a mean AUC of 0.47.Paper 5 demonstrated that the ischioanal fossa had the highest reproducibility among the standardization methods, with a%interCV of 18.9 for center 1 and 11.2 for center 2. These findings imply that the ischioanal fossa could serve as a reference tissue for standardizing T2WI intensities.

Moving to drawbacks, Paper 1 is limited by the fact that the ground truth grading was done by a single pathologist/center and the fact that it utilized a relatively small number of patients. Additionally, it does not include clinical data or epidemiological factors. Paper 2 used different biopsy techniques resulting in a non-uniform gold standard for labeling. Paper 3 used MRI-guided biopsy as the reference standard and relied on center slices instead of the full 3D volumes. Paper 4, despite its promise, relies on a small sample size and needs external validation with more patients and cancer types. Paper 5, also has a very limited cohort size . The authors also note the exclusion of healthy subjects, and potential artifacts from patient movement as possible limitations.

Group 4 : Prostate Cancer Detection and Machine Learning

#	doi	Title and Date	# of Patients	Pre-Processing	Classifier
1	10.3389/fonc.2022.934108	"Evaluation of the Efficiency of MRI-Based Radiomics Classifiers in the Diagnosis of Prostate Lesions", **2022**	238	normalization, resampling, discretisation	DT, Gaussian NB, XGBoost, LR, RF and SVM
2	10.1002/jmri.27204	"Deep-Learning-Based Artificial Intelligence for PI-RADS Classification to Assist Multiparametric Prostate MRI Interpretation: A Development Study", **2020**	687	alignment, resampling, normalization	CNN
3	10.3390/diagnostics11040594	"Advanced Imaging Analysis in Prostate MRI: Building a Radiomic Signature to Predict Tumor Aggressiveness", **2021**	102	bias field correction(N4), intensity standardization	SVM
4	10.3389/fonc.2020.631831	"Use of Radiomics to Improve Diagnostic Performance of PI-RADS v2.1 in Prostate Cancer", **2021**	203	normalization, resampling	LR
5	10.1038/s41598-021-81272-x	"Utility of T2-weighted MRI texture analysis in assessment of peripheral zone prostate cancer aggressiveness: a single-arm, multicenter study", **2021**	128	co-registration, bias correction (N4), normalization	SVM

6	10.1002/jmri.27793	"Integrative Machine Learning Prediction of Prostate Biopsy Results From Negative Multiparametric MRI ", **2022**	230	bias field correction (N4), z-score normalization	SVM
7	10.3390/cancers12020390	"Multiparametric MRI for Prostate Cancer Detection: New Insights into the Combined Use of a Radiomic Approach with Advanced Acquisition Protocol", **2020**	65	registration, normalization, discretisation	LR
8	10.1007/s10278-018-0160-1	"A Deep Learning-Based Approach for the Detection and Localization of Prostate Cancer in T2 Magnetic Resonance Images", **2019**	19	normalization	CNN

5. Group 4 : Prostate Cancer Detection and Machine Learning

This group focuses on cancer detection. Paper 1 is a single-center, single-vendor study of 238 patients to assess various imaging classifiers for prostate disease diagnosis in the future. Paper 2, on the contrary, is a multicenter, multivendor study with 687 patients that aimed to develop an artificial intelligence (AI) solution for PI-RADS classification. Paper 3, a single-center, single-vendor study of 102 patients, wants to develop a reproducible radiomic pipeline for prostate cancer aggressiveness prediction. Paper 4 has a single-center, single-vendor design with 203 patients and aims to develop a radiomics model to improve the performance of PI-RADS v2.1. Paper 5 used a multicenter, single-vendor dataset of 128 patients to examine textural features for assessing prostate cancer aggressiveness. Paper 6, which has 230 patients, utilized a single-center, single-vendor setup to identify patients who could safely avoid prostate biopsy using a radiomics-based machine learning approach. Paper 7, with 65 patients, did a single-center, single-vendor study to compare standard and advanced radiomic models for prostate cancer detection, by taking into account 2D and 3D lesion segmentation. Paper 8 wants to create a deep convolutional encoder-decoder architecture capable of segmenting the prostate, its anatomical structures, and malignant lesions all at the same time. It was done at a single center with a single vendor, and had the smallest dataset of the group with 19 patients.

In terms of preprocessing, Paper 1 applied normalization, resampling, and discretization . Paper 2 used alignment, resampling, and normalization. Paper 3 employed bias field correction (N4) and intensity standardization. Paper 4 employed normalization and resampling. Paper 5 implemented co-registration, bias correction (N4), and normalization. Paper 6 used bias field correction (N4) and z-score normalization. Paper 7 applied registration, normalization, and discretization while Paper 8, only used normalization.

In terms of results, Paper 1 had high diagnostic abilities with the random forest classifier performing the best with an AUC of 0.88. Paper 2 compared AI PI-RADS scoring with

radiologist-assigned PI-RADS scoring and found no statistically significant differences for clinically significant prostate cancer. Paper 3 produced an accuracy of 0.88 in predicting PCa aggressiveness. Paper 4 combined Rad-score with PI-RADS and significantly improved PCa diagnosis, with AUC of 0.931 in the validation set. Paper 5 achieved the best accuracy (84%) by using ADC + T2W features. Moving to Paper 6, the machine learning radiomics model achieved 98.3% and 98.0% negative predictive values (NPVs). In Paper 7, both the standard and advanced radiomic models showed significant diagnostic accuracy with an AUC up to 0.99, with the 3D segmentation model performing the best. Finally, in the Paper 8, AUC, accuracy, and recall were 0.995, 0.894, and 0.928, respectively.

Every study had its own drawbacks. Paper 1's limitations include the single-center design and the lack of follow-up data. In Paper 2, the AI model required manual segmentation of the lesions. Furthermore, the study's retrospective approach and use of multi-center data raise questions about potential biases. Because the sample was small yet homogeneous, Paper 3 had limitations. Selection bias may have been induced by Paper 4's retrospective design, which also lacked external and prospective validation. The cohort in Paper 5 lacked comparisons with clinical readings including PI-RADS scores and had an unbalanced distribution of cancer aggressiveness classes.Paper 6's dataset was relatively small and showed an imbalance between positive and negative biopsies. Paper 7 highlighted the need for MR-guided biopsy techniques to address uncertainties in histological radiological correlations and featured a small sample size, limiting its ability to evaluate non-binary classification tasks and method reproducibility. Paper 8, as noted by the authors themselves, had a very limited cohort of only 19 patients. Although it achieved very high AUC values, the small number of patients more than likely will lead to lack of generalizability.

Group 5 : PI-RADS 3 Lesion Characterization

#	doi	Title and Date	# of Patients	Pre-Processing	
1	10.3389/fonc.2022.840786	“Utility of Clinical–Radiomic Model to Identify Clinically Significant Prostate Cancer in Biparametric MRI PI-RADS V2.1 Category 3 Lesions”, **2022**	103	histogram-based intensity standardization, resampling, registration	LR
2	10.1002/jmri.27692	“Magnetic Resonance Imaging Radiomics-Based Machine Learning Prediction of Clinically Significant Prostate Cancer in Equivocal PI-RADS 3 Lesions”, **2021**	240	normalization, discretisation, interpolation	RF
3	10.118	“Machine learning-based	463	resampling, intensity	SVM

	6/s12880-023-01002-9	radiomics model to predict benign and malignant PI-RADS v2.1 category 3 lesions: a retrospective multi-center study", **2023**		discretisation, Z-score normalization	
4	10.3389/fonc.2021.825429	"Development and Validation of a Radiomics Nomogram for Predicting Clinically Significant Prostate Cancer in PI-RADS 3 Lesions", **2022**	306	Z-score standardization	LR
5	10.1007/s00261-020-02678-1	"A radiomics machine learning-based redefining score robustly identifies clinically significant prostate cancer in equivocal PI-RADS score 3 lesions", **2020**	263	normalization, discretisation	SVM
6	10.3390/jcm11216304	"Radiomics in PI-RADS 3 Multiparametric MRI for Prostate Cancer Identification: Literature Models Re-Implementation and Proposal of a Clinical–Radiological Model", **2022**	116	(1st model): normalization, resampling, discretisation (2nd model): standardization, resampling, discretisation (3rd model): normalization, b-spline interpolation, discretisation	linear discriminant, linear, quadratic, and cubic SVM, CT, and KNN
7	10.1038/s41598-020-80749-5	"Evaluation of a multiparametric MRI radiomic-based approach for stratification of equivocal PI-RADS 3 and upgraded PI-RADS 4 prostatic lesions", **2021**	80	registration, normalization	LR

6. Group 5 : PI-RADS 3 Lesion Characterization

The goal of the five research in this group was to create radiomics models for the detection of clinically significant prostate cancer (csPCa) in PI-RADS 3 lesions. In Paper 1, 103 patients from a single center and vendor were included in a dataset that was largely used to assess clinical variables in conjunction with radiomics. Paper 2 also used a single-center, single-vendor dataset with 240 patients and focused on T2WI radiomics. Papers 3 and 4 on the other hand used multicenter, multivendor datasets. The former included 463 patients and tried to differentiate PI-RADS 3 tumors that were benign from those that were malignant. The latter created a radiomics nomogram for csPCa prediction within PI-RADS 3 lesions and comprised 306 patients. The objective of Paper 5, a 263 patient single-center, single-vendor trial, was to exclude csPCa in ambiguous PI-RADS score 3 categories.Paper 6, a single-center, single-vendor study involving 116 patients, investigated the utility of radiomics in detecting prostate cancer lesions in PI-RADS 3 lesions and peripheral PI-RADS 3 lesions

upgraded to PI-RADS 4. The purpose of Paper 7 was to develop a model to aid in the clinical management of prostate lesions classified as PI-RADS 3. This study involved 80 participants and was carried out at a single center using a single vendor.

When it comes to preprocessing, Paper 1 employed histogram-based intensity standardization,resampling and registration. Paper 2 and paper 5 utilized normalization and discretization. Paper 2 , additionally, employed interpolation. Paper 3 focused on resampling, intensity discretization, and Z-score normalization. Paper 4 applied Z-score standardization. Paper 6 employed registration and normalization. Lastly, Paper 7, included 3 models with 3 different preprocessing approaches; Model 1 : normalization, resampling, discretisation Model 2: standardization, resampling, discretisation, Model 3: normalization, b-spline interpolation, discretisation.

The studies' diagnostic performances were high. Paper 1 used a clinical-radiomics model to reach an AUC of 0.88, whereas Paper 2 used T2WI radiomics to get an AUC of 0.76.An integrated model for the prediction of csPCa was developed in Paper 3, with a mean AUC of 0.803. The radiomics nomogram generated in Paper 4 , with an AUC of 0.939, had strong calibration and discrimination abilities. The Radiomics Machine Learning (RML) model presented in Paper 5 has an AUC of 0.89. With an AUC of 80%, Paper 6 found that second-order models for PI-RADS 3 stratification outperformed first-order models. Conversely, for upPI-RADS 4 stratification, first-order models outperformed superior-order models with an AUC of 89%. In Paper 7 biopsy results were strongly associated with specific radiomic features, depending on the model used. Clinically significant cancers could be predicted with a 66% sensitivity and 71% specificity using PSA density alone. The proposed model combined PSA density and radiomic features achieved a 76% specificity and an 80% sensitivity.

Some limitations that were shared by all of the studies were single-center, single-vendor datasets, small sample numbers, and retrospective study designs. In Paper 1, only PSA and age were included as clinical variables. Paper 2 explained that the focus on PI-RADS 3 lesions, may restrict the generalizability of the findings but this is what all the papers also did in this group. Paper 3 did not consider lesion location, whereas Paper 4 depended on manual radiologist segmentation. Paper 5 used subjectively given PI-RADS values from two radiologists and introduced selection bias by eliminating non-follow-up patients. Paper 6 also has its limitations due to unbalanced datasets, the lack of a separate analysis concentrating on clinically relevant PCa lesions, the absence of features from DCE-MRI parameters, and the possibility of inter-observer heterogeneity in feature extraction.

4.3 Grouping the studies on Different Cancers based on their objective

Group	Description	# of papers
1	**Image Preprocessing Impact on Radiomics Features** - the effects of various image preprocessing methods on the stability and reliability of radiomic features in various cancer types	5
2	**Radiomic Features for Survival Prediction** - use of radiomic features to predict survival outcomes and stratify patients in various cancer types	8
3	**Radiomics for Cancer Risk Assessment -** analysis of radiomic features to establish predictive models that can identify individuals at higher risk of developing cancer.	6
4	**Radiomics for Cancer Detection -** using radiomic features for the detection of cancer in patients enhance early diagnosis and timely intervention for cancer patients	7
5	**Radiomics for Cancer Grading** - establishment of predictive models and radiomic signatures for grading cancer, facilitating precise characterization	5
6	**Radiomics for Prognosis** - assessing the prognosis of cancer patients by incorporating radiomic features and clinical factors in predictive models	7

7. Grouping of Different Cancer studies

Group 1 : Image Preprocessing Impact on Radiomics Features

#	doi	Title and Date	# of Patients	Pre-Processing	Classifier
1	10.1088/1361-6560/ab2f44	"Impact of image preprocessing on the scanner dependence of multi-parametric MRI radiomic features and covariate shift in multi-institutional glioblastoma datasets", **2019**	161	8-bit global rescaling (discretisation), bias field correction, histogram standardization, isotropic resampling	N/A
2	10.1002/acm2.12795	"Impact of image preprocessing methods on reproducibility of radiomic features in multimodal	262	co-registration, resampling, skull stripping, noise reduction, bias field	gradient boosting

		magnetic resonance imaging in glioblastoma", **2019**		correction, intensity normalization	multi-class classification
3	10.1002/mp.14368	"Repeatability of radiomic features in magnetic resonance imaging of glioblastoma: Test–retest and image registration analyses" , **2020**	17	registration,N4/N3 bias correction, wavelet transform, binning (discretisation)	N/A
4	10.3390/cancers12020518	"The Impact of Normalization Approaches to Automatically Detect Radiogenomic Phenotypes Characterizing Breast Cancer Receptors Status" , **2020**	91	7 different normalization methods	SVM, RF, and NB
5	10.1038/s41598-020-69298-z	"Standardization of brain MR images across machines and protocols: bridging the gap for MRI-based radiomics" , **2020**	20 + 243	bias field correction (N4), resampling, skull stripping, co-registration, normalization (three methods - Nyul, WhiteStripe, Z-Score), discretisation (two methods - fixed bin size and fixed bin number)	NN, RF, SVM, LR, NB

8. Group 1 : Image Preprocessing Impact on Radiomics Features

The main objective of the studies in Group 1 is to investigate different preprocessing methods and pipelines in order to establish the best possible one for radiomics. The first 3 papers study glioblastoma tumors. Paper 1 aims to assess the impact of common image preprocessing methods on MRI radiomic features in a multicenter and multivendor study involving 161 patients. Paper 2 investigates the effect of intensity inhomogeneity correction and noise filtering, on the robustness and reproducibility of radiomic features in a multicenter, multi-vendor dataset of 262 patients which is the largest of this group. On the other hand, Paper 3 with the smallest dataset, assesses the repeatability of radiomic features in MRI in a single-center, single-vendor study involving 17 patients. Paper 4 compares three normalization strategies for predicting clinical phenotypes in a 91-patient dataset from multiple centers and a single vendor. Finally, Paper 5 compares three distinct intensity normalization methods and two approaches for intensity discretization in brain MRIs for future radiomic research. It featured two datasets, the first with 20 patients and the second with 243.

Moving to preprocessing, all studies utilized a number of preprocessing methods and examined them for their impact on radiomic features. To begin, Paper 1 examined discretisation, bias field correction, histogram standardization, and isotropic resampling.

Paper 2 in comparison examined co-registration, resampling, skull stripping, noise reduction, bias field correction, and intensity normalization. Paper 3 employed registration,N3 and N4 bias correction, wavelet transform, discretization. Paper 4 compared 7 different methods of normalization. Lastly, Paper 5's different steps included N4 bias field correction, resampling, skull stripping, co-registration, normalization, and discretization.

When looking at results, beginning with Paper 1, they concluded that GLSZM features were more dependent on scanner parameters than Haralick features and magnetic field strength has a greater impact than the vendor. Image preprocessing methods had varying effects on feature dependence, with Laplacian Gaussian filtering being the most dependent feature. Covariate changes were seen in response to bin numbers and image preprocessing, with histogram standardization having the most impact. 8-bit-local-rescaling was the most effective in predicting overall survival. Paper 2 concluded that necrosis characteristics ($\bar{n}$ ~449/1461, 30%) are connected with glioma survival and mutations. Local binary pattern filtered pictures are affected the least by intensity inhomogeneity and have the most repeatable features. Also, the reproducible features increased after bias field correction. In Paper 3, the results demonstrate the highest repeatability for Laplacian of Gaussian image processing (mean 78.9%) and Full Affine transformation with 12 degrees of freedom (mean 32.4%) among registration techniques, and no differentiation for N4, N3, or no bias correction. Paper 4 demonstrated the strong relationships between non-normalized radiomic characteristics and techniques such as scaling, z-score, robust z-score, and upper quartile normalization. Conversely, the correlations of the more aggressive approaches (log-transformation, quantile normalization, and whitening normalization) are weak, indicating that they should be used with caution. Finally, the results of Paper 5 showed that the performance of classification models and the robustness of first-order features were improved by intensity normalization. The accuracy of tumor grade classification increased from 0.67 to 0.82 with Nyul, WhiteStripe, and Z-Score normalization techniques.

When it comes to limitations , Paper 1 mentions that feature repeatability and the impact of other factors, such as receiver coils, on radiomic characteristics, were not assessed. Paper 2, concluded that more validation is required to determine the best preprocessing technique for standardizing MR images. Paper 3 highlights the need for validation in a more extensive multicenter dataset. Paper 4 states that the dataset's images were acquired more than ten years ago and this could mean that the techniques were evaluated using outdated equipment and may not be relevant in the modern era, highlighting the need for additional validation. The necessity for additional validation was also mentioned in Paper 5 as a study limitation.

Group 2 : Radiomic Features for Survival Prediction

#	doi	Title and Date	# of Patients	Pre-Processing	Classifier
1	10.1148/radiol.2018180200	"Radiomic MRI Phenotyping of Glioblastoma: Improving Survival Prediction", **2018**	217	skull stripping, registration, N4 bias correction, normalization	RSF
2	10.1148/radiol.2016160845	"Radiomic Profiling of Glioblastoma: Identifying an Imaging Predictor of Patient Survival with Improved Performance over Established Clinical and Radiologic Risk Models", **2016**	119	registration, N4 bias correction, intensity normalization,discrete and stationary or undecimated wavelet transformations	Supervised Principal Component Analysis
3	10.3389/fncom.2019.00058	"A Multi-parametric MRI-Based Radiomics Signature and a Practical ML Model for Stratifying Glioblastoma Patients Based on Survival Toward Precision Oncology", **2019**	163	co-registration, smoothing, interpolation, skull-stripping, intensity standardization(MRI intensity rescaling)	Linear SVM, Gaussian SVM, Coarse Gaussian SVM, KNN, Coarse KNN, Cosine KNN, Medium, KNN, Discrimination analysis, Linear Discriminant, Ensemble Learning, Subspace Discrimination
4	10.1007/s00330-020-07089-w	"Radiomics risk score may be a potential imaging biomarker for predicting survival in isocitrate dehydrogenase wild-type lower-grade gliomas ", **2020**	117	resampled, N4 bias correction, registration, normalization	Radiomics Risk Score (RRS)
5	10.1016/j.ejrad.2019.07.010	"Improving survival prediction of high-grade glioma via machine learning techniques based on MRI radiomic, genetic and clinical risk factors", **2019**	147	N4 correction bias, skull striping resampling, isotropic resampling, intensity normalization	Radiomics Risk Score (RRS)
6	10.1016/j.ebiom.2020.103093	"Incremental prognostic value and underlying biological pathways of radiomics patterns in medulloblastoma", **2020**	172	N4 bias field distortion correction, isotropic resampling, registration, histogram matching (for intensity	*not specified*

				normalization), discretisation	
7	10.18383/j.tom.2018.00052	"Multiparameter MRI Predictors of Long-Term Survival in Glioblastoma Multiforme", **2019**	22	resampling , co-registration, intensity calibration	N/A
8	10.1093/neuonc/nox188	"Radiomic subtyping improves disease stratification beyond key molecular, clinical, and standard imaging characteristics in patients with glioblastoma ", **2018**	181	intensity normalization, registration	N/A

9. Group 2 : Radiomic Features for Survival Prediction

The papers in this group attempt to predict survival for different cancers. Paper 1 aims to investigate whether integrating radiomic features from MRI with clinical and genetic profiles could improve survival prediction in patients with glioblastoma. The study was conducted at a single-center, using data from a single vendor, and included 217 patients - the largest dataset of the group. Paper 2 looked into whether radiomic feature-based MRI signatures could predict survival and classify patients with newly diagnosed glioblastoma more accurately than established clinical and radiologic models. The study was conducted at a single-center with data from a single vendor and included 119 patients. Paper 3, a multicenter study with 163 patients, wanted to create a radiomics signature and compare several machine learning models to classify patients into groups based on overall survival using pre-operative mpMRI of patients with glioblastoma. Paper 4 aimed to evaluate whether radiomics from MRI could predict overall survival in patients with IDHwt lower-grade gliomas and investigate the added prognostic value of radiomics over clinical features. This single-center study included 117 patients. Paper 5 , a multicenter study with 147 patients, aims to develop a radiomics signature to predict overall survival in patients with high-grade glioma and create a nomogram combining radiomic, genetic, and clinical risk factors. Paper 6 aims to develop a radiomics signature for predicting overall survival (OS) and progression-free survival (PFS) in patients with medulloblastoma (MB) and investigate the prognostic value and biological pathways of radiomics patterns. It was a single-center study that included 172 patients. Following that, Paper 7, with the smallest dataset , utilized radiomic analysis of standard-of-care mpMRI scans to subdivide glioblastoma tumors into distinct regions called "habitats". The study involved only 22 patients in a single-center training cohort and a multicenter validation cohort. Finally, Paper 8, a single-center study involving 181 patients with glioblastoma, analyzed radiomic features extracted from mpMRI scans. The goal is to develop a radiomic signature for predicting progression-free and overall survival (PFS and OS) .

Regarding preprocessing, most papers had an extensive pipeline. In Papers 1 and 5 the prep-processing included skull stripping, registration, N4 bias correction, normalization, with Paper 5 also adding resampling as a step. Paper 2 employed registration, N4 bias correction, normalization, and discrete and stationary or undecimated wavelet transformations. Paper 3's preprocessing steps included co-registration, smoothing, interpolation, skull-stripping, and intensity standardization. Paper 4 utilized resampling, N4 bias correction, registration and normalization. Paper 8, with the least preprocessing steps, did intensity normalization and registration. Paper 7, also with fewer steps, included resampling, co-registration, and intensity calibration. And lastly, Paper 6 used N4 bias field correction, isotropic resampling, registration, histogram matching (normalization), and discretization.

Moving to the results , begining with Paper 1, the study suggests that adding a radiomics model to clinical and genetic profiles improves survival prediction compared to models utilizing only clinical and genetic data, with an AUC of 0.782. In Paper 2, the radiomic model outperformed both radiologic and clinical risk models in the prediction of progression-free survival (PFS) and overall survival (OS). More specifically, the Supervised Principal Component (SPC) analysis model achieved an OS Integrated Brier Score (IBS) of 0.149 and an AUC of 0.654, while the PFS IBS was 0.138 with an AUC of 0.611. When paired with clinical data, the SPC analysis model improved more, with an OS IBS of 0.142 and an AUC of 0.696, and a PFS IBS of 0.132 with an AUC of 0.637. Paper 3 produced promising results and attempted but did not manage to validate the radiomic signature they created. Nevertheless, the ensemble model showed superior performance in predicting survival classes, with an overall accuracy of 57.8% and AUC values of 0.81 for short-, 0.47 for medium-, and 0.72 for long-survivors. The radiomic signature produced in Paper 4, on the other hand, showed that radiomic signature scores independently predicted survival with hazard ratios of 9.479 and 6.148, enhancing the model's performance for predicting overall survival by increasing iAUC to 0.780–0.797 from 0.726. In Paper 5, the radiomic signature, along with IDH status and age, were proved to be independent risk factors, and the nomogram combining these factors improved overall survival estimation with AUC values of 0.764 and 0.758 in the training and test cohorts, respectively.In Paper 6, it was found that the combined radiomics and clinical signature performed better than individual signatures, with an AUC of 0.762 for predicting OS and 0.697 for PFS. Also, 9 pathways showed a strong correlation with the radiomics signature. Paper 7 discovered that the fractional tumor volume in habitat 6 around the time of diagnosis was the strongest predictor of future survival, therefore improving overall survival rates. In both the discovery and validation cohorts, the fractional tumor volume in habitat 6 was 35% ± 6.5% and 34% ± 4.8%, respectively. Finally, Paper 8 showed that the radiomic signature enhanced accuracy of predictions for PFS and OS, reducing errors by 36% and 37%, respectively.

All papers come with their respective limitations. A lot of them include a small dataset or a single-center study (or both) which reduces generalizability but these are not the only drawbacks. Paper 1 notes a lack of semi-automatic tumor outlining, lack of external validation, and points out the heterogeneous nature of GBM tumors, and their inability to consider the effects of various treatments on tumor progression. Paper 2 highlights that the final post processing workflow and statistical processing have multiple steps and require approximately 60 minutes of computation time per patient. Paper 3 has limitations as well due to its small dataset and the absence of information on tumor resection status. Additionally, they didn't manage to validate their model. Paper 4 also points out the retrospective nature of their data, and the possible variations in images acquired over several years with different acquisition parameters. For Paper 5 the drawbacks included the retrospective nature of the study as well, and the inclusion of only T1W1 and T2-FLAIR images. Paper 6 notes the lack of volumetric MRI data and Paper 7 admits to the inability to draw strong conclusions due to the very limited sample of 22 patients. Lastly, Paper 8 underlines the use of non-automatic segmentation and a long post processing time as limitations.

Group 3 : Radiomics for Cancer Risk Assessment

#	doi	Title and Date	# of Patients	Pre-Processing	Classifier
1	10.3390/jpm12111854	"Preoperative Tumor Texture Analysis on MRI for High-Risk Disease Prediction in Endometrial Cancer: A Hypothesis-Generating Study", **2022**	96	resampling	LR
2	10.3390/cancers15082209	"Prediction of Deep Myometrial Infiltration, Clinical Risk Category, Histological Type, and Lymphovascular Space Invasion in Women with Endometrial Cancer Based on Clinical and T2-Weighted MRI Radiomic Features", **2023**	413	non-uniformity correction (N4), resampling, intensity normalization	fitcauto method (an automated classifier training function in MATLAB) selected the Compact Classification Ensemble Classifier
3	10.1186/s13244-022-01156-0	"MRI-based radiomics analysis improves preoperative diagnostic performance for the depth of stromal invasion in patients with early stage cervical cancer", **2022**	234	normalization	LR

4	10.18383/j.tom.2019.00029	"Radiomic Features of Multiparametric MRI Present Stable Associations With Analogous Histological Features in Patients With Brain Cancer", **2020**	16	registration, intensity normalization	Linear Mixed-Effects model
5	10.7150/jca.50872	"MRI-Based Radiomic Model for Preoperative Risk stratification in Stage I Endometrial Cancer", **2021**	102	normalization	LR
6	10.1148/radiol.212873	"Development and Validation of Multiparametric MRIbased Radiomics Models for Preoperative Risk Stratification of Endometrial Cancer ", **2022**	157	resampling, normalization, discretisation	RF

10. Group 3 : Radiomics for Cancer Risk Assessment

The studies of this group attempt to do risk stratification for a variety of cancers. Paper 1 is a multicenter study with 96 women that aims to develop and validate an MRI-based radiomics model for preoperative prediction of high-risk endometrial cancer. The objective is to estimate deep myometrial invasion (DMI), predict lymphovascular space invasion (LVSI), and differentiate between low-risk and other risk categories. Paper 2, a multicenter and multivendor study involving 413 patients, the primary goal was to predict various clinical parameters, including risk, in women with endometrial cancer using machine learning classification methods based on clinical and image signatures extracted from T2-weighted MR images. Paper 3, aims to develop and validate a T2WI-based radiomics model for the detection of middle or deep stromal invasion in early-stage cervical cancer. It is a single-center study with 234 patients. Paper 4, is a single center study involving a multivendor dataset of 16 patients - the smallest from the group - with brain cancer. The goal is to investigate the localized relationship between MR-derived radiomic features and histology-derived "histomic" features. Paper 5 boasts a larger dataset of 102 patients in a single-center setting. The objective was to establish a risk classification model for endometrial cancer based on MRI and clinical factors. Finally, Paper 6, a multicenter, multivendor study encompassing data from 157 patients, aims to evaluate the performance of mpMRI three-dimensional radiomics-based machine learning models. The models should be able to differentiate between low- and high-risk histopathologic markers in advanced-stage endometrial carcinoma.

Most papers in this group included one step in their preprocessing . One exception is Papers 2 and 7 that included resampling and normalization. Paper 2 also employed N4 bias field correction and Paper 7, discretisation. Another exception is Paper 5 which utilized

registration and normalization. Paper 1 and Paper 4 performed only resampling. Paper 3 and 6 performed only normalization.

Moving to results, for Paper 1 the results showed that whole-tumor radiomic models achieved an AUC of 0.85 for DMI estimation, 0.92 for LVSI prediction, and 0.84 for differentiating low-risk from other risk classes. The model in Paper 2, achieved notable AUCs of 0.79, 0.82, 0.91, and 0.85 for different classifications. The corresponding 95% confidence intervals demonstrated the robustness of these results. Paper 3's radiomics model achieved an AUC of 0.879 in the validation cohort, outperforming radiologists and maximal tumor diameter (MTD), both with sensitivity and specificity of 87.9% and 84.6%. Paper 4 showed that while the overall findings were heterogeneous, several radiomic features demonstrated strong associations with their histomic counterparts, especially those derived from FLAIR and post-contrast T1W images.Paper 5 concluded that the risk-classification radiomic model performed better than the model based on clinical and conventional MRI characteristics, with an AUC of 0.946. Furthermore, the combined model (radiomic features and tumor size) showed the best predictive performance, with AUCs of 0.955 in the training and 0.889 in the validation cohorts. Lastly, in Paper 6, the radiomics models showed excellent performance, with AUCs ranging from 0.74 to 0.84 in the test set. It is important to note that radiomics outperformed radiologist readings in identifying deep myometrial invasion.

The studies have their respective limitations. A small dataset, and a single-center or single-vendor design are common limitations noted between a number of them but they are not the only ones. Paper 1 notes a potential selection bias, and data inhomogeneity due to a decade-long data collection period from 2009 to 2019. Additionally, they mention the lack of automated segmentation and not integrating other routine sequences like DWI and contrast-enhanced MRI. The latter are the same limitations underlined in Paper 3. Paper 2's team highlights the limitations through their future goals including exploring different classifiers, and incorporating additional evaluation metrics. Paper 4 was constrained by its very small sample size of only 16, focusing solely on primary brain cancer patients, and the use of a tile-based prediction method, which left out the utilization of shape- and size-based radiomic features. Paper 5 notes a number of varied limitations; the study's retrospective design with varying scanning parameters, the use of only one sequence for texture analysis, and the omission of prognostic information in the models. Lastly, Paper 6, explains that the drawbacks of the study included variations in ROIs due to them being drawn by two different radiologists, which then were compared to radiomics, and an imbalance in the distribution of histopathologic features.

Group 4 : Radiomics for Cancer Detection

#	doi	Title and Date	# of Patients	Pre-Processing	Classifier
1	10.3390/diagnostics12051085	"MRI-Based Radiomic Features Help Identify Lesions and Predict Histopathological Grade of Hepatocellular Carcinoma", **2022**	97	normalisation	LR
2	10.3390/cancers14102372	"Fully Automatic Whole-Volume Tumor Segmentation in Cervical Cancer", **2022**	131	resampling, z-score normalization	U-Net (CNN)
3	10.3390/diagnostics11060919	"Radiomics and Machine Learning with Multiparametric Breast MRI for Improved Diagnostic Accuracy in Breast Cancer Diagnosis", **2021**	93	discretisation, ComBat	Gaussian SVM
4	10.3390/app10176109	"Breast Cancer Mass Detection in DCE–MRI Using Deep-Learning Features Followed by Discrimination of Infiltrative vs. In Situ Carcinoma through a Machine-Learning Approach", **2020**	55	resampling, normalization	MLP-ANN
5	10.1016/j.ejrad.2019.108755	"Machine Learning-Based Multiparametric MRI Radiomics for Predicting the Aggressiveness of Papillary Thyroid Carcinoma ", **2020**	120	N4 bias correction, image intensity rescale, 8 image filters (Wavelet, Laplacian of Gaussian, Square, Square Root, Logarithm, Local Binary Pattern, Gradient Magnitude and Exponential)	22 tested -> best : GBC, LR, PAC, LSVC
6	10.1016/j.ejrad.2019.04.004	"Preliminary utilization of radiomics in differentiating uterine sarcoma from atypical leiomyoma: Comparison on diagnostic efficacy of MRI features and radiomic features ", **2019**	78	isotropic resampling, discretisation	LR

7	10.1007/s10278-020-00336-y	"MRI Radiomics for the Prediction of Fuhrman Grade in Clear Cell Renal Cell Carcinoma: a Machine Learning Exploratory Study", **2020**	32	isotropic resampling, normalization, discretisation, filters (Laplacian of Gaussian filters, Wavelet decomposition)	DT

11. Group 4 : Radiomics for Cancer Detection

Group 4 includes studies with the objective of detecting different types of cancer. Paper 1 is a multicenter, multivendor study involving 97 patients with the goal to develop an MRI-based radiomics approach for the preoperative detection of hepatocellular carcinoma (HCC) and the prediction of its histological grade. Paper 2 aims to train a deep learning algorithm for the automatic segmentation of primary tumors in cervical cancer patients. It is a multicenter, multivendor study with 131 patients. Paper 3, also a multicenter, multivendor study, with the goal to assess radiomics analysis in conjunction with machine learning of DCE and DWI radiomics models separately and in combination as multiparametric MRI for improved breast cancer detection. The study involved 93 patients. Paper 4 on the other hand is single-center, single-vendor study with 55 patients. It presents a prototype of a computer-aided detection/diagnosis (CAD) system aimed at assisting radiologists in discriminating between in situ and infiltrating breast cancer tumors. Paper 5, also a single-center, single-vendor study, investigates the predictive capability of machine learning-based multiparametric MRI radiomics for evaluating the aggressiveness of papillary thyroid carcinoma (PTC) preoperatively. The study comprised 120 patients. Paper 6 is a multicenter, multivendor study involving 78 patients. the objective is to explore whether MRI and radiomic features could differentiate uterine sarcoma from atypical leiomyoma and compare the diagnostic performance of a radiomic model with radiologists. Finally, Paper 7 is a single-center, single-vendor study with the smallest cohort of 32 patients. The goal is to assess a combined approach of radiomics and machine learning based on MRI for non-invasively predicting Fuhrman grade Clear Cell Renal Cell Carcinoma, more specifically distinguishing high- from low-grade tumors and assessing grade.

Preprocessing pipelines in this group, just like the previous one, are fairly simple. Papers 2 and 4 utilized resampling and normalization. Conversely Paper 6 employed resampling and discretization. Paper 1 only had one step ; normalization. Paper 3 employed discretization and ComBat. Papers 5 and 7 are the exceptions with a lot more extensive pipelines. Paper 5 included N4 bias correction, image intensity rescale and 8 different filters. Paper 7 included resampling, normalization, discretisation, and 2 different filters.

Turning to results, Paper 1 achieved promising outcomes from radiomic prediction models, with the best AUCs ranging from 71% to 96%. Radiomics based on T2 and DCE showed

potential for both HCC detection and grading. The deep learning algorithm in Paper 2 performed better than the two independent radiologists in tumor segmentation, with median dice scores of 0.60 and 0.58 compared to 0.78 for the radiologists laying the ground for its future use as a detection tool. However agreement amongst raters was better than between the deep learning method and raters. Paper 3's multiparametric radiomics model that combined DCE and DWI extracted features achieved the best AUC of 0.85 and diagnostic accuracy of 81.7%. The CAD system in Paper 4 achieved a sensitivity of 75% for mass detection and an AUC of 0.70 for distinguishing between in situ and infiltrative tumors. In Paper 5, the combination of feature selection and a Gradient Boosting Classifier achieved an AUC of 0.92 for predicting PTC aggressiveness, outperforming clinical characteristics alone which achieved an AUC of 0.56. In Paper 6, radiologists' diagnostic performance based on MRI achieved an AUC of 0.752, sensitivity of 58.6%, specificity of 91.8%, and accuracy of 79.5% while the best radiomic model achieved an AUC of 0.830, a sensitivity of 76.0%, a specificity of 73.2% on average, and an accuracy of 73.9%. Lastly, the ensemble methods in Paper 7 achieved accuracy greater than 90% in differentiating high- and low-grade tumors, with the best accuracy (84.4%) achieved by random forest.

Besides the common limitations, all Papers highlighted additional ones, Paper 1 pointed out the unbalanced patient population, a lack of standardization in radiomic investigations, and a lack of reproducibility of radiomic features. The drawbacks mentioned in Paper 2 were the extensive preprocessing, which could impact data quality, and the variability in acquisition methods and scanners. Paper 3 explained that due to the inclusion of small tumors they had to lower the data to 16 gray levels and as a result they had to exclude lesions with less than 40 pixels which could introduce selection biases. Paper 4' only highlighted drawback was its single-center nature. Paper 5 admitted to limitations regarding the lack of validation. Paper 6 mentions its retrospective nature, which could introduce selection bias, the reliance on radiologist-selected imaging features, and the use of manual segmentations. Paper 7 also mentioned its retrospective nature of the study, and manual segmentations. Additionally they pointed to the lack of reproducibility analysis.

Group 5 : Radiomics for Cancer Grading

#	doi	Title and Date	# of Patients	Pre-Processing	Classifier
1	10.3389/fonc.2020.00459	"Preoperative Prediction of Extramural Venous Invasion in Rectal Cancer: Comparison of the Diagnostic Efficacy of Radiomics Models and Quantitative Dynamic Contrast-Enhanced Magnetic Resonance Imaging", **2020**	106	registration, resampling	LR

2	10.3390/brainsci13060912	"Predicting Histopathological Grading of Adult Gliomas Based On Preoperative Conventional Multimodal MRI Radiomics: A Machine Learning Model", **2023**	500	registration, resampling, normalization	Gaussian NB, KNN, RF, AB, SVM, MLP-ANN
3	10.3390/jcm9061853	"Non-Invasive Assessment of Breast Cancer Molecular Subtypes with Multiparametric Magnetic Resonance Imaging Radiomics", **2020**	91	normalization	MLP-ANN
4	10.1016/j.ebiom.2019.08.059	"Tumor grading of soft tissue sarcomas using MRI-based radiomics", **2019**	122	N4 bias field correction, intensity normalization, discretisation, isotropic resampling (b-spline interpolation), reconstruction (wavelet decomposition filtering and Laplacian of Gaussian)	LR
5	10.1002/jmri.27532	"Magnetic Resonance Imaging-Based Radiomics Nomogram for Prediction of the Histopathological Grade of Soft Tissue Sarcomas: A Two-Center Study ", **2021**	181	Gray-level quantization, isotropic resampling, ComBat	LR

12. Group 5 : Radiomics for Cancer Grading

In this group the Papers focus on cancer grading. Paper 1 deals with rectal cancer and aims to develop and validate an MRI-based radiomics model to preoperatively predict high-risk endometrial cancer. It was conducted as a single-center study, with data from a single MRI vendor, and included 106 patients. Paper 2 focuses on glioma grading. This study has the objective of developing a predictive model for classifying adult gliomas into grades 2–4 based on preoperative conventional multimodal MRI radiomics. It included 500 patients, making it the largest cohort.However, it was a single-center study with data from a single MRI vendor. Paper 3 focuses on breast cancer subtypes and has the objective of classifying breast cancer molecular subtypes It was conducted as a single-center study with data from a single-vendor MRI system, involving 91 patients. Paper 4 addresses soft tissue sarcoma grading. The multicenter and multivendor study developed MRI-based radiomics grading models to differentiate between low-grade and high-grade soft tissue sarcoma (STS). The study included 122 patients and had a multicenter and multivendor design. Lastly, Paper 5 aims to assess the potential of radiomics for disease stratification beyond key molecular,

clinical, and standard imaging features in glioblastoma. It included 181 patients from a single-center study using data from one MRI vendor.

For preprocessing, in Paper 1, the preprocessing was relatively straightforward ; registration and resampling. Paper 2, also followed the same steps but also added normalization. Paper 3 only employed normalization. Paper 4, included various preprocessing steps : N4 bias field correction, intensity normalization, discretisation, isotropic resampling , and reconstruction (wavelet decomposition filtering and Laplacian of Gaussian). Lastly, Paper 5, with a simpler pipeline, utilized Gray-level quantization, isotropic resampling, and ComBat.

The results of these studies highlight the capabilities of radiomics in cancer grading. In Paper 1, radiomics models predicted extramural venous invasion with AUCs of 0.826 and 0.872 in the training cohort. Paper 2 produced an AUC of 0.81 for glioma grade categorization in the validation set. Paper 3, which focused on breast cancer subtypes, had a total AUC of 0.86 for triple-negative subtype categorization. Paper 4 achieved AUCs for soft tissue sarcoma grading ranging from 0.69 to 0.78 for various MRI sequences, demonstrating the models' potential clinical value.In Paper 5 several models for prediction of clinical outcomes were compared. The RS-Combined model was the best with an AUC of 0.829 on the external validation set while the radiomics nomogram which combined the model with risk variables achieved AUC of 0.879 (external validation).

All of the studies had drawbacks. Paper 1 noted many issues, including manual ROI segmentation, a small sample size, single-center data, and the use of only one contrast-enhanced phase for tumor image segmentation. Paper 2 pointed to its retrospective nature, single-center design, and small sample size. Paper 3 had its own limitations. These were its retrospective design, single-center data, and the manual segmentation of the images. Paper 4 highlighted the need for a larger cohort with either numerous subtypes or enough data to support a single one. Paper 5's problems were non-automatic segmentation and a long post-processing time, which might limit clinical applications.

Group 6 : Radiomics for Prognosis

#	doi	Title and Date	# of Patients	Pre-Processing	Classifier
1	10.3390/cancers14081858	“Measurement of Perfusion Heterogeneity within Tumor Habitats on Magnetic Resonance Imaging and Its Association with Prognosis in Breast Cancer Patients”, **2022**	455	original images -> b-spline interpolation ROIS -> nearest-neighbor interpolation Histogram-matching ->	*not specified*

				harmonization of MRI intensities	
2	10.3390/genes14010028	"Identifying Associations between DCE-MRI Radiomic Features and Expression Heterogeneity of Hallmark Pathways in Breast Cancer: A Multi-Center Radiogenomic Study", **2022**	174	z-score normalization, discretisation	RF
3	10.3390/cancers14225507	"Unsupervised Analysis Based on DCE-MRI Radiomics Features Revealed Three Novel Breast Cancer Subtypes with Distinct Clinical Outcomes and Biological Characteristics", **2022**	246	registration, N4 bias correction, b-spline resampling, normalization, histogram remapping	RF
4	10.3390/cancers12102958	"Baseline MRI-Radiomics Can Predict Overall Survival in Non-Endemic EBV-Related Nasopharyngeal Carcinoma Patients", **2020**	136	image denoising (Gaussian filter and bias correction N4), z-score standardization, resampling (b-spline interpolation), histogram discretisation	N/A
5	10.1038/s41416-019-0706-0	"MRI-based radiomics model for preoperative prediction of 5-year survival in patients with hepatocellular carcinoma", **2020**	201	intensity normalization, resampling	RF
6	10.1038/s41598-018-22739-2	"Radiomic MRI signature reveals three distinct subtypes of glioblastoma with different clinical and molecular characteristics, offering prognostic value beyond IDH1", **2018**	208	co-registration, smoothing, correction for magnetic field in-homogeneities, skull stripping	N/A
7	10.1002/jmri.27444	"Whole-Volume Tumor MRI Radiomics for Prognostic Modeling in Endometrial Cancer", **2020**	138	z-score normalization	LR

13. Group 6 : Radiomics for Prognosis

Group 6, contains studies that estimate cancer patients' prognosis by combining radiomic features and clinical data. Paper 1 a multicenter study involving 455 patients that aims to identify perfusional subregions sharing similar kinetic characteristics from DCE-MRI using data-driven clustering. Additionally, it attempts to evaluate the effect of perfusion heterogeneity based on these subregions on patients' survival outcomes. Paper 2 , also a multicenter study, involves 174 patients. The main goal is to investigate the relationship between DCE-MRI radiomic features and the expression activity of hallmark pathways in breast cancer. It also sought to develop prediction models of pathway-level heterogeneity. A

third multicenter study, paper 3, aims to reveal the heterogeneity of dynamic (DCE-MRI in breast cancer, identify its prognosis values, and explore its molecular characteristics. It encompasses 246 patients. Paper 4, on the other hand, is a single-center study involving 136 patients amining to train an MRI-based radiomic signature as a prognostic factor in Nasopharyngeal Carcinoma patients. Paper 5 , a multicenter study, involves 201 patients with hepatocellular carcinoma and aims to develop a radiomics model incorporating a radiomics signature and clinical risk factors to evaluate prognosis. Paper 6 , a single-center study, and utilizing a single MRI vendor, involved 208 patients with de novo glioblastoma. The primary objective is to explore the imaging heterogeneity within glioblastoma using radiomic analysis of pre-operative multiparametric MRI (mpMRI) data. The study aimed to gain insights into disease subtypes, risk stratification, and improved treatment planning. Finally, Paper 7 wants to develop MRI-based whole-volume tumor radiomic signatures to predict aggressive endometrial cancer disease for accurate preoperative staging and prognostication. It is a multicenter , single-vendor involving 138 patients.

In Paper 1 interpolation and histogram matching were used as preprocessing . Paper 2's preprocessing included z-score normalization and discretization. In Paper 3, on the other hand,it was more extensive, with registration, bias correction, resampling, normalizing, and histogram remapping. In Paper 4 it included denoising, normalization, resampling, and histogram discretization. Paper 6 included co-registration, smoothing, magnetic field inhomogeneities correction, and skull stripping. In Paper 5 preprocessing included intensity normalization and resampling. Finally, only z-score normalization was used in Paper 7 to prepare the data for analysis.

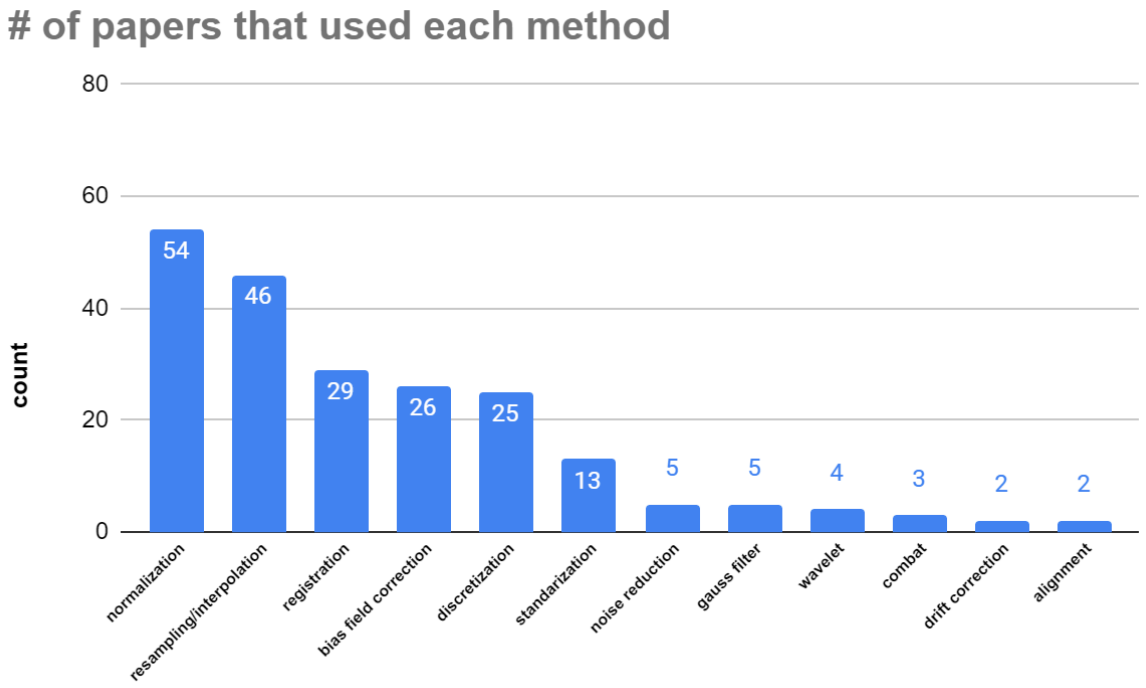
Moving on to results; Paper 1 identified five distinct habitats (ie perfusion patterns). The high-risk habitat (HRS) was found to be an independent risk factor for predicting worse disease-free survival (DFS) outcomes in both the HRS-only risk model and combined habitat risk model. In the validation cohort, the combined habitat risk model (hazard ratio = 4.128, p = 0.003, AUC = 0.760) outperformed the other five risk models. In Paper 2, the prediction model for the mTORC1 signaling pathway obtained the best results, with mean absolute errors of 27.29% and 28.61% in internal and external test sets, respectively. Paper 3 discovered Three imaging subtypes that showed high repeatability. The tumor sizes and enhancement patterns varied significantly between subtypes, with significant outcomes in the discovery cohort (p = 0.024) and prognosis datasets (p ranged from 0.0001 to 0.0071). The poorest outcomes were typically seen in tumors with large diameters and fast growth. This study gives valuable insights into the heterogeneity of breast cancer as indicated by DCE-MRI, which has important prognostic implications. Paper 4 showed that the radiomics-based signature exhibited good predictive potential for overall survival and loco-regional recurrence-free survival, with AUC values of 0.68 and 0.72, respectively.In all cases, combining radiomics with clinical characteristics improved prognostic performance. In

Paper 5, the top 30 survival-related radiomics traits were chosen for the radiomics signature. The model showed excellent calibration and discrimination, with a mean AUC of 0.9804 in the training set and 0.7578 in the validation set. Paper 6 , through an extensive analysis of 267 radiomic features extracted from various glioblastoma subregions, was able to find three distinct clusters: rim-enhancing, irregular, and solid. Remarkably, they achieved an 88% clustering assignment reproducibility. Of significant clinical importance, they found that the rim-enhancing subtype exhibited the longest survival, even surpassing existing molecular estimates. Paper 7, demonstrated that whole-tumor radiomic signatures achieve AUCs of 0.84/0.76 (training/validation) for DMI, 0.73/0.72 for LNM, 0.71/0.68 for FIGO III + IV, 0.68/0.74 for NE histology, and 0.79/0.63 for E3 tumor. Conversely, single-slice radiomics achieve similar training AUCs, but worse validation AUCs for LNM and FIGO III + IV. Additionally, tumor volume achieves comparable training AUC to the whole-tumor radiomic signatures, but worse for E3 tumors.

The studies all have their own limitations though. In Paper 1, there are concerns regarding the accuracy of the results due to the lack of robust pathological connections with image-based segmentations. Confounding factors may also be introduced if an inhomogeneous patient cohort is used. In Paper 2 the possible problems with the existing dataset's representativeness are highlighted, which underlines the need for better preprocessing and larger datasets for validation.While the study in Paper 3 reveals that imaging features have promising predictive value, it is obvious that these findings need to be validated in bigger datasets, and the lack of specific metrics like C-indices or hazard ratios does not allow for evaluation of the findings.The study in Paper 4, exclusively included patients with N-positive diseases and did not include an independent validation cohort. Paper 5, although it achieved notable results, had a retrospective nature and lacked genetic traits, which would offer a more reliable result. Finally, Paper 6 is limited by the lack of genetic and histopathologic data, which restricts the applications of the research, and Paper 7 by the omission of some picture sequences, which may affect how generalizable the results are.

4.4 Proposal for a Preprocessing Pipeline in Prostate Cancer Radiomics Research

In the following graph the occurrences of each preprocessing step are presented with the most common methods being normalization, resampling, registration, bias field correction and discretization.



1. Histogram of number of papers that included each preprocessing method

There were also a few papers that stood out for their results along with their study design or sizable cohort or both. In the 2021 "*Bi-parametric magnetic resonance imaging based radiomics for the identification of benign and malignant prostate lesions: cross-vendor validation*", Xuefu Ji et al. using a multi-center and multi-vendor dataset of 459 patients, performed normalization and resampling and achieved an AUC of 0.833 for their prostate cancer risk stratification model. In another 2021 study "*Integration of Clinical identifications With Deep Transferrable Imaging Feature Representations Can Help Predict Prostate Cancer Aggressiveness and Outcome*", Jie Bao et al. did normalization on their impressive single-vendor, 1442 patients dataset derived from multiple institutions. They achieved an average AUC of 0.85 for models on the external testing set when trying to predict PIRADS scores. In the 2022 study "*Evaluation of the Efficiency of MRI-Based Radiomics Classifiers in the Diagnosis of Prostate Lesions*'', Linghao Li et al. conducted a single-vendor, single-center study with 238 patients to test the performance of various classifiers for prostate cancer diagnosis. After applying normalization, resampling and discretisation and testing a number of classifiers they conclude that the random forest classifier performed the best with an AUC of 0.88. In the 2021 study "*Use of Radiomics to Improve Diagnostic Performance of PI-RADS*

v2.1 in Prostate Cancer", Mou Li et al. utilized a single-center, single-vendor design with 203 patients to develop a radiomics model to better the performance of PI-RADS v2.1. They performed normalization and resampling and concluded that combining the radiomics score with the PIRADS score produced the best results with an AUC of 0.931 in the validation set. In the 2022 study"*Development and Validation of a Radiomics Nomogram for Predicting Clinically Significant Prostate Cancer in PI-RADS 3 Lesions*" Tianping Li et al. , used a multi-center and multivendor dataset consisting of 306 patients to create radiomics nomogram for clinically significant PCa prediction in PI-RADS 3 lesions. They performed normalization as the sole pre-processing step and the final nomogram achieved an AUC of 0.939.

In an older study in 2018 "*Radiomic MRI Phenotyping of Glioblastoma: Improving Survival Prediction*" Sohi Bae et al. investigated whether integrating radiomic features from MRI with clinical and genetic profiles could improve survival prediction in patients with glioblastoma. To do this they chose a single-center, single-vendor dataset of 217 patients and performed skull stripping, registration, N4 bias correction and normalization. The results showed that the combined model improves survival prediction compared to models utilizing only clinical and genetic data, with an AUC of 0.782. A more recent study from 2023 titled "*Prediction of Deep Myometrial Infiltration, Clinical Risk Category, Histological Type, and Lymphovascular Space Invasion in Women with Endometrial Cancer Based on Clinical and T2-Weighted MRI Radiomic Features*", Xingfeng Li et al. in a multicenter and multivendor study involving 413 patients to predict various clinical outcomes in women with endometrial cancer using machine learning classification methods based on clinical and radiomics features. They performed bias field correction, resampling, normalization and achieved an AUC of 0.879 in the validation cohort, outperforming radiologists. The final study, also in 2023, titled *"Predicting Histopathological Grading of Adult Gliomas Based On Preoperative Conventional Multimodal MRI Radiomics: A Machine Learning Model"*, Peng Du et al. had the goal of developing a predictive model for classifying adult gliomas into grades 2–4 based. It included 500 patients from one center and one vendor. Registration, resampling and normalization were employed on the MRI images. The final model produced an AUC of 0.81 for glioma grade categorization in the validation set.

From the above we can derive the conclusion that most studies, including the most notable ones, perform normalization and resampling with bias field correction and registration following close by. Normalization was chosen as one of the preprocessing steps to ensure that the intensities of all images are standardized. Resampling was chosen as a necessary step due to the difference in size between the MRI images and their corresponding segmentations. Since the dataset images were acquired at the same center, using one vendor and modality, registration was deemed unnecessary to minimize preprocessing steps. Finally bias field correction was selected over registration due the importance of

removing vendor-derived artifacts which could introduce biases or problems during radiomics feature extraction.

4.5 Pipeline Implementation

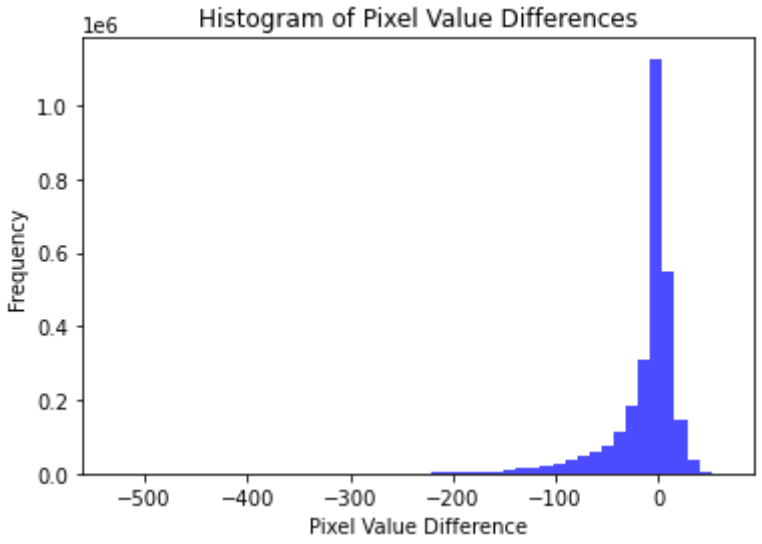
Bias Field Correction

The Bias Field correction script was implemented with the simpleITK. After testing different iteration counts, it was determined that 10 iterations provided the best results as more than 10 iterations led to a very harsh over-correction. The shrink factor was selected to be 1 (default) as there was no benefit to lessening the image sizes since the dataset is small and causes no computational concerns. The number of fitting levels was also selected as its default value of 4.

```
shrink_factor = 1 # default=1
num_iterations = 10 # default = 50
num_levels = 4 #default=4
```

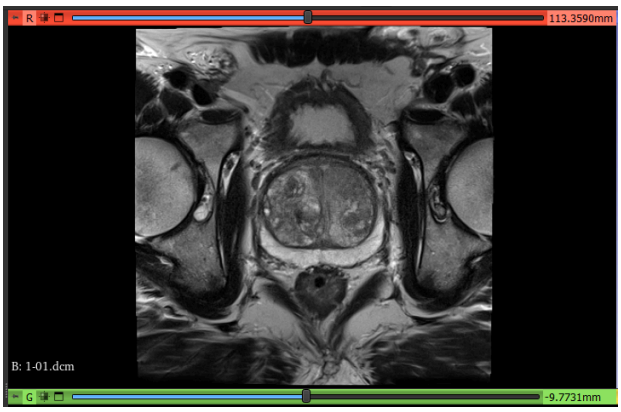
The corrected images are saved in a separate file in .nrrd format.

In the following figure, the bias field is represented in a histogram as the difference of the before and after correction images.

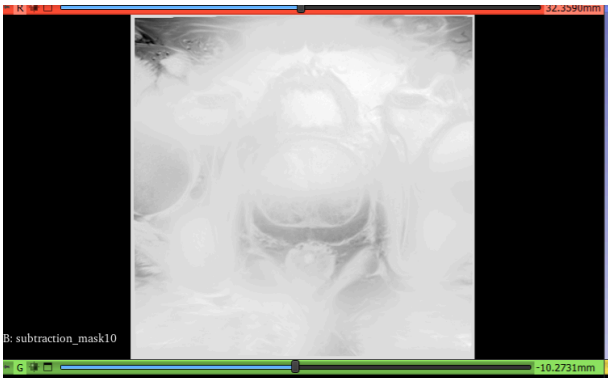


2. The bias field histogram

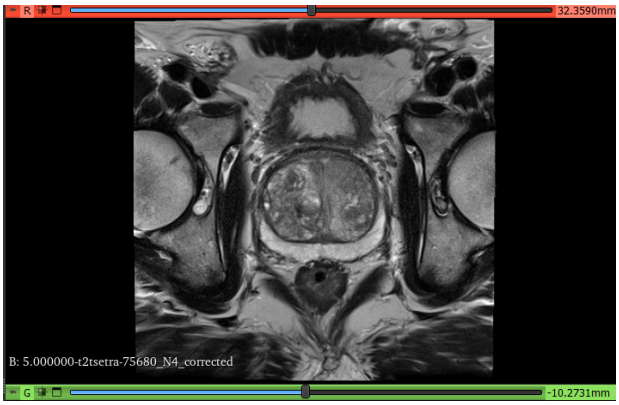
Due to the higher quality of the MRI images the bias field cannot be easily visually inspected, however images of before and after N4 bias field correction as well as the bias field itself are provided below. The software used for viewing the images was Slicer 3D.



3. Image Before bias field correction



4. Image of The bias field



5. Image After Bias field correction

Appendix A & B contains the python scripts for the N4 correction and the subtraction respectively

Normalization, Resampling & Radiomics Feature Extraction

The radiomics feature extraction was performed with the pyradiomics library. Before the extraction two functions were used to normalize and resample the images respectively. Steps:

- Files: The function checks for the existence of the N4-corrected image and the segmentation mask files.

- Image compatibility: The segmentation mask is a DICOM file, but the N4-corrected image is an NRRD file so it is converted to an image format in order for them to be compatible.
- Normalization: Z-score normalization is applied to normalize the intensity values of the N4-corrected image using simpleITK and numpy according to the formulα:

$$x' = \frac{x-\mu}{\sigma}$$

where x' is the normalized data, x is the original, is the μ mean intensity and σ the standard deviation.

- Resampling: The N4-corrected image is resampled so that it matches the segmentation mask's dimensions in terms of spacing, origin, and direction.
- Mask Application: The resampled image is masked to focus on the prostate region.
- Feature Extraction: The N$ corrected images were nrrd files and had to be transformed into image files to be able to be used alongside the segmentations which were DICOM files. Radiomic features were extracted using the pyradiomics library, with all features enabled and images processed in 3D.

The extracted radiomic features were 116, and included:

First Order Statistics: Mean, Median, Minimum, Maximum, Standard Deviation, Variance, Skewness, Kurtosis, Energy, Entropy

Shape-based Features: Volume, Surface Area, Sphericity, Compactness, Maximum 3D Diameter, Major Axis Length, Minor Axis Length, Elongation

Texture Features:

Gray Level Co-occurrence Matrix (GLCM): Contrast, Dissimilarity, Homogeneity, Energy, Correlation, ASM (Angular Second Moment)

Gray Level Run Length Matrix (GLRLM): Short Run Emphasis (SRE), Long Run Emphasis (LRE), Gray Level Non-Uniformity (GLN), Run Length Non-Uniformity (RLN), Run Percentage (RP)

Gray Level Size Zone Matrix (GLSZM): Small Area Emphasis (SAE), Large Area Emphasis (LAE), Gray Level Non-Uniformity (GLN), Zone Percentage (ZP)

Neighborhood Gray Tone Difference Matrix (NGTDM): Coarseness, Contrast, Busyness, Complexity, Strength

Appendix C contains the python script that performs Normalisation and Resampling as well as Radiomic Feature Extraction

Model Development

Data Preprocessing:

The csv was loaded and preprocessed by removing non-feature columns and by spliting 3-dimensional data columns into separate columns for size and spacing. This way the dataframe could be utilized for model training.

Classifier Selection:

Several classifiers were compared : SVM, Random Forest, Logistic Regression, KNN, and Gradient Boosting, using stratified k-fold cross-validation with 5 folds. Performance metrics were calculated for each and included accuracy, precision, recall, F1 score, AUC, and F-beta score . The best classifier, Logistic Regression, was selected based on these metrics and used for model development.

Hyperparameter Tuning:

The pipeline included a standard scaler, PCA, feature selection with Random Forest and the Logistic Regression classifier. In order to find the best prarameters for each, the RandomizedSearchCV was utilized with repeated stratified k-fold cross-validation of 5 folds and it concluded on the following :

```
Best parameters train-holdout : {'PCA__n_components':
0.9921607745818374, 'classifier__C': 8.09397348116461,
'feature_selection__estimator__n_estimators': 50}

Best parameters train-test: {'PCA__n_components':
0.9834529727696003, 'classifier__C': 0.894925020519195,
'feature_selection__n_estimators': 200}
```

Model Training and Evaluation:

For the train test model the dataset was split into 70% training and 30% test set. The training set was used for hyperparameter tuning with repeated stratified 5-fold cross validation and then used to fit the model while the test set was used for final evaluation of the model. Optimal thresholds were calculated using precision-recall curves and for both training and test and metrics were calculated, including accuracy, precision, recall, F1 score, and AUC. The model was then saved.

For the train-holdout model the data were split into 80% training and 20% holdout set. The training set was utilized exactly in the same way as previously, while the hold-out set was

kept as unseen data to assess the model's performance. The same metrics were calculated as before and the model was also saved.

Performance Metrics of the two models:

Set	Accuracy	Precision	Recall	F1	AUC
Training (80%)	0.827	0.8	0.667	0.727	0.922
Hold-out (20%)	0.643	0.5	0.4	0.444	0.689
Training (70%)	0.804	0.769	0.625	0.690	0.867
Test (30%)	0.750	0.600	0.857	0.706	0.791

14. Final Classification Models Metrics

Appendix D contains the python scripts for the data preprocessing, the classifier selection, the hyperparameter tuning and the model development

Chapter 5: Discussion and analysis of findings

In the the train-test model, in the training set the model achieved an overall AUC of 0.867, an accuracy of 80.4%, precision of 76.9%, which is important for minimizing false positives, a sensitivity(recall) of 62.5% and an F1 score, showing the balance between sensitivity and precision , of 0.690. This shows the models ability to successfully classify most patients and avoid a lot of false positives. In the test set, most metrics, besides recall and F1, were slightly lower which is to be expected. In this case, the model achieved an overall AUC of 0.791, an accuracy of 75%, precision of 60%, a sensitivity(recall) of 85.7% and an F1 score of 0.706. This indicates that the model maintains an overall good performance with a slightly lowered ability to avoid false positives and slightly better ability to identify true positives.

However, when looking at the train-holdout model, although the metrics for the training show a strong performance, the metrics in the hold-out set show a significant drop. Precision went from 80% to 50% and AUC went from 92,2% to 68,9%. This could be explained by the fact that the dataset itself is very small and therefore when creating subsets, they will be even smaller. It is possible that 80% or even 100% of the dataset is not enough to capture the intricate patterns in the radiomics data that would give the model better classification abilities when faced with unseen data. Also in such a small hold-out set each incorrect prediction has a larger impact on the overall metrics compared to a larger dataset.

When comparing this work to existing literature, the examples are limited and typically correlate to only one part of this study.

The closest work to this one is a study published in 2023 called "*Enhancing Prostate Cancer Classification by Leveraging Key Radiomics Features and Using the Fine-Tuned Linear SVM Algorithm*" by Metin Varan et al. which utilized the ProstateX dataset to create a classification radiomics model. The main focus, however, was to use an SVM classifier and then, find the best feature selector and compare them to no-feature selection. No preprocessing is specified to have been performed on the dataset. The sole metric provided as an evaluation of the final models is accuracy. All feature selectors achieved 90-95% accuracy whereas using no feature selection resulted in 43.64% accuracy. These results, while impressive, are hard to interpret due to the lack of other relevant metrics, however if we specifically focus on accuracy of the model developed in this thesis, it is lower than the accuracy achieved by this research paper. However, due to the different end-goals and lack of further metrics a direct comparison is challenging.
Another notable study is the 2023 study "*Weakly Supervised MRI Slice-Level Deep Learning Classification of Prostate Cancer Approximates Full Voxel- and Slice-Level Annotation: Effect of Increasing Training Set Size*" by Cedric Weißer MD et al. In this study the PostateX was utilized as part of the training cohort. The study did not use radiomic features but instead

developed a CNN model with the MRI images to classify prostate cancer. The goal was to compare slice-level labeling and patient-level labeling along with different training cohort sizes. The preprocessing included normalization and registration. Slice-level annotation produced the best results with an AUC of 0.75,0.80,0.83 with a test set of 200,500,998 respectively. Patient-level annotation produced an AUC of 0.64 , 0.72 and 0.78 for the respective testing sets. This paper, although used a different approach, highlights the value of location-based labels as the final model had better results with this additional information. It is also a bigger study with more patients which makes the results even more reliable compared to the significantly smaller set of 66 patients utilized in this thesis. The AUCs achieved in the paper are a bit higher although comparable. It is not easy however to conclusively compare the two as this work utilized radiomics features as input for the final model while the paper used the MRI images themselves.
A last honorable mention is the recently published study called "*Prediction of Clinically Significant Prostate Cancer Using Radiomics Models in Real-World Clinical Practice*" by Jie Bao et al. Although this study did not use the ProstateX dataset, they provide an interesting look into the potential of radiomics for predicting clinically significant cancer and PIRADS scores, especially when using such a large dataset of 1616 patients as they did. A portion of those patients came from external hospitals and were kept as external test sets. For preprocessing they mention resampling, as was done in this thesis, but they also included denoising and discretization. They used the FeatureExplorer software (which is based on pyradiomics) to extract the radiomics and then they compared a number of classifiers, namely random forest (RF), support vector machine (SVM), logistic regression (LR), and linear discriminant analysis (LDA) to conclude on the best one. The best one, random forest, produced an AUC of 0.874, in an internal testing cohort and, 0.876 and 0.893 in external testing cohorts. This study although did not use the same dataset is an important one as it used a significantly larger multi-center cohort and achieved impressive results that could serve as a step towards effective assistance to medical staff for PIRADS scoring and predicting cancer aggressiveness.

Circling back to original research questions, the purpose of this research was, first, identify an effective preprocessing pipeline for MRI images that could used for radiomics analysis afterwards and second, to implement this pipeline as the first step to developing a radiomics features classification model that could classify patients into clinically and not-clinically significant. The results of the study showcase the final model's ability to successfully differentiate between these two categories despite the dataset only consisting of 66 patients. The preprocessing pipeline utilized was based on an extensive literature review and comparison of about 80 papers in order to propose a robust pipeline. The final proposal was simple, involving only 3 steps, bias field correction, normalization and resampling, ensuring easier reproducibility and lower chances of preprocessing related artifacts being added to the original images. The results on the final model highlight the pipeline's effectiveness as the metrics show a strong performance. The model developed was also a simple approach, utilizing a Logistic Regression classifier, and it successfully identified most patients that had

prostate cancer on the train and test set model. However, the train-holdout model highlights the need for further tuning and validation on a larger dataset in order for generalizability and practical applications.

Chapter 6: Conclusion and recommendations

The purpose of this research was to establish a preprocessing pipeline for prostate cancer MRI radiomics analysis and then utilize this pipeline to develop a classification model that would distinguish clinically significant from non-clinically significant prostate cancer. In order to achieve this ~ 80 papers were compared on their preprocessing pipeline as well as their study design (single/multi center, single/multi vendor, number of patients), the resulting metrics and their limitations. The final proposed pipeline was applied on a 66 patient dataset and then radiomic features were extracted to be used for the development of a classification model. The entire method was implemented with the Python programming language.

Through this process 3 steps (bias field correction, normalization and interpolation) were identified as the most frequently used steps among all papers and also among the most notable studies. After implementing this pipeline on the dataset, the radiomics features were extracted and utilized to create two final models.

The train-test model achieved an accuracy of 80.4% and an AUC of 0.867 on the training set while on the test set it achieved an accuracy of 75% and an AUC of 0.791. This difference is to be expected when making predictions on the test set as it was unseen data however the performance was still strong and showed no signs of overfitting. Of note is the precision which is a measure of the model's ability to avoid false negatives. On the training set the model achieved a precision of 76.9% which is a good value, however then it dropped to 60% on the test set which might mean that the model struggles a little more to currently identify some cases as negative for cancer.

The train-holdout model on the other hand, performed worse due to the "unseen data" aspect. Precision dropped from 80% to 50% and AUC from 92,2% to 68,9% from training to hold-out. This result truly highlights the need for bigger publicly available datasets to be used to develop more robust models that could potentially be introduced in clinical practice. This work remains however as an important step in this direction and could open the door for future research that focuses on generalizability and clinical applications.

The use of the proposed pipeline will aid in removing artifacts and the standardization of the images which will result in more reliable radiomics features. The final model could be used in a clinical setting to minimize the need for invasive biopsies and as a supplementary tool to help avoid false positives and unnecessary treatments. This makes the whole process easier and safer for both the doctor and the patient. The patient would only need to get an MRI

scan and then the doctor, after segmenting the prostate, could utilize the code provided in this research to determine whether the patient has prostate cancer. Even in cases were a blood test or biopsy might have been conducted, it could still be useful to take advantage of the model as it could serve as an additional confirmation (or not) of the results, ensuring that the patient is not going to be administered with chemotherapy or radiation therapy and possibly lose their lives when they never even had cancer to begin with.

However, there are several limitations that require discussion. Although the test-train model achieved a strong performance, the dataset used was small (66 patients). This means that without further testing on a larger cohort the generalizability of the results cannot be ensured as shown with the hold-out set. Furthermore, the aggregation of the labels resulted in the loss of spatial information which could have a useful tool for minimizing the need to biopsy every single lesion on a patient as it could exclude some of them for being cancer positive. Also, although the entire process is done with python instead of utilizing a number of tools, which helps simplify the process, it still is not ideal for a clinical setting as most practitioners would not be equipped with sufficient programming understanding to integrate it easily and successfully in their practice. It is also important to mention that this research utilized pre-existing segmentations and did not develop any way to automate this step. Therefore in order to use this process in a clinical setting, there would be a need to either train existing staff or hire new staff that would be responsible for utilizing the code along with an experienced radiologist to take on the image segmentations.

In order to address these limitations, there are some future research suggestions that could aid in that process. First, further validation is needed on a larger dataset and re-tuning of the model might be necessary. A larger dataset that also has lesion-based labels, just like the ProstateX, could allow for the creation of region-based models which could point out exactly which lesion(s) are cancer positive. Furthermore, including images from a number of institutions and different vendors could also be useful as it would mean that the final model could be applied to a lot more settings around the world as well as for bigger research incentives. Finally, the creation of a website or a software that runs the python scripts in the background would be the final goal as it would make this process a lot easier for physicians and erase the need for additional staff. Incorporating an automated segmentation tool in this software would, although it would have its own limitations, completely automate this process. This way the MRI scans would be the only prerequisite as input for the software and the output would be the model's prediction on the clinical significance and would ideally also include the specific location of the cancerous lesion(s).

References

1. Rebello RJ, Oing C, Knudsen KE, et al. Prostate cancer. *Nature Reviews Disease Primers*. 2021;7(1). doi:https://doi.org/10.1038/s41572-020-00243-0
2. Bilal M, Javaid A, Amjad F, Youssif TA, Afzal S. An overview of prostate cancer (PCa) diagnosis: Potential role of miRNAs. *Translational Oncology*. 2022;26:101542. doi:https://doi.org/10.1016/j.tranon.2022.101542
3. Madej A, Wilkosz J, Różański W, Lipiński M. Complication rates after prostate biopsy according to the number of sampled cores. *Central European Journal of Urology*. 2012;65(3):116-118. doi:https://doi.org/10.5173/ceju.2012.03.art3
4. Cheung D, Finelli A. Magnetic resonance imaging diagnosis of prostate cancer: promise and caution. *Canadian Medical Association Journal*. 2019;191(43):E1177-E1178. doi:https://doi.org/10.1503/cmaj.190568
5. Loeb S, Bjurlin MA, Nicholson J, et al. Overdiagnosis and Overtreatment of Prostate Cancer. *European Urology*. 2014;65(6):1046-1055. doi:https://doi.org/10.1016/j.eururo.2013.12.062
6. 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. The Lancet Oncology. 2019;20(1):100-109. doi:https://doi.org/10.1016/s1470-2045(18)30569-2
7. Ghafoor S, Burger IA, Vargas AH. Multimodality Imaging of Prostate Cancer. *Journal of Nuclear Medicine*. 2019;60(10):1350-1358. doi:https://doi.org/10.2967/jnumed.119.228320
8. Sarkar S, Das S. A Review of Imaging Methods for Prostate Cancer Detection. *Biomedical Engineering and Computational Biology*. 2016;7s1(1-15):BECB.S34255. doi:https://doi.org/10.4137/becb.s34255
9. Pedler K, Kitzing YX, Varol C, Arianayagam M. The Current Status of MRI in Prostate Cancer. *Australian Journal for General Practitioners*. 2015;44(4):225-230. https://www.racgp.org.au/afp/2015/april/the-current-status-of-mri-in-prostate-cancer
10. Tempany CMC, Carroll PR, Leapman MS. UpToDate. www.uptodate.com. Published December 2022. https://www.uptodate.com/contents/the-role-of-magnetic-resonance-imaging-in-prostate-cancer
11. Hricak H, White S, Vigneron D, et al. Carcinoma of the prostate gland: MR imaging with pelvic phased-array coils versus integrated endorectal--pelvic phased-array coils. *Radiology*. 1994;193(3):703-709. doi:https://doi.org/10.1148/radiology.193.3.7972810
12. Steinkohl F, Pichler R, Junker D. Short Review of Biparametric Prostate MRI. *Memo*. 2018;11(4):309-312. doi:https://doi.org/10.1007/s12254-018-0458-1
13. Czarniecki M. Prostate Imaging-Reporting and Data System (PI-RADS) | Radiology Reference Article | Radiopaedia.org. Radiopaedia. Published 2014. Accessed 2023. https://radiopaedia.org/articles/prostate-imaging-reporting-and-data-system-pi-rads-1
14. Mayerhoefer ME, Materka A, Langs G, et al. Introduction to Radiomics. *Journal of Nuclear Medicine*. 2020;61(4):488-495. doi:https://doi.org/10.2967/jnumed.118.222893
15. Court LE, Rao A, Krishnan S. Radiomics in cancer diagnosis, cancer staging, and prediction of response to treatment. *Translational Cancer Research*. 2016;5(4):337-339. doi:https://doi.org/10.21037/tcr.2016.07.14
16. Shur JD, Doran SJ, Kumar S, et al. Radiomics in Oncology: A Practical Guide. *RadioGraphics*. 2021;41(6):1717-1732. doi:https://doi.org/10.1148/rg.2021210037

17. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology*. 2016;278(2):563-577. doi:https://doi.org/10.1148/radiol.2015151169
18. 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:https://doi.org/10.1148/radiol.2020191145
19. Moore CM. Image registration | Radiology Reference Article | Radiopaedia.org. Radiopaedia. Published August 2022. Accessed November 24, 2023. https://radiopaedia.org/articles/image-registration
20. Stamoulou E, Spanakis C, Manikis GC, et al. Harmonization Strategies in Multicenter MRI-Based Radiomics. *Journal of Imaging*. 2022;8(11):303. doi:https://doi.org/10.3390/jimaging8110303
21. Juntu J, Sijbers J, Dyck D, Gielen J. Bias Field Correction for MRI Images. *Advances in Soft Computing*. 2005;30:543-551. doi:https://doi.org/10.1007/3-540-32390-2_64
22. Zhao W, Hu Z, Kazerooni AF, et al. Physics-Informed Discretization for Reproducible and Robust Radiomic Feature Extraction Using Quantitative MRI. *Investigative Radiology*. Published online September 11, 2023:10.1097/RLI.0000000000001026. doi:https://doi.org/10.1097/RLI.0000000000001026
23. Leng J, Xu G, Zhang Y. Medical image interpolation based on multi-resolution registration. *Computers & Mathematics with Applications*. 2013;66(1):1-18. doi:https://doi.org/10.1016/j.camwa.2013.04.026
24. Shinohara RT, Sweeney EM, Goldsmith J, et al. Statistical normalization techniques for magnetic resonance imaging. *NeuroImage Clinical*. 2014;6(6):9-19. doi:https://doi.org/10.1016/j.nicl.2014.08.008
25. Duron L, Balvay D, Vande Perre S, et al. Gray-level discretization impacts reproducible MRI radiomics texture features. Fan Y, ed. *PLOS ONE*. 2019;14(3):e0213459. doi:https://doi.org/10.1371/journal.pone.0213459
26. Moore C, Bell D, Noise reduction. Reference article, Radiopaedia.org (Accessed on 25 Nov 2023) https://doi.org/10.53347/rID-72577
27. Moreno López M, Frederick JM and Ventura J (2021) Evaluation of MRI Denoising Methods Using Unsupervised Learning. *Front. Artif. Intell.* 4:642731. doi: 10.3389/frai.2021.642731
28. V U, Srinivasan R, Bell D, et al. Noise. Reference article, Radiopaedia.org (Accessed on 25 Nov 2023) https://doi.org/10.53347/rID-12937
29. Zhang X, Feng Y, Chen W, Li X, Faria AV, Feng Q and Mori S (2019) Linear Registration of Brain MRI Using Knowledge-Based Multiple Intermediator Libraries. *Front. Neurosci.* 13:909. doi: 10.3389/fnins.2019.00909
30. Chilaca-Rosas MF, Garcia-Lezama M, Moreno-Jimenez S, Roldan-Valadez E. Diagnostic Performance of Selected MRI-Derived Radiomics Able to Discriminate Progression-Free and Overall Survival in Patients with Midline Glioma and the H3F3AK27M Mutation. *Diagnostics*. 2023;13(5):849. doi:https://doi.org/10.3390/diagnostics13050849
31. Mirza-Aghazadeh-Attari M, Ambale Venkatesh B, Aliyari Ghasabeh M, et al. The Additive Value of Radiomics Features Extracted from Baseline MR Images to the Barcelona Clinic Liver Cancer (BCLC) Staging System in Predicting Transplant-Free Survival in Patients with Hepatocellular Carcinoma: A Single-Center Retrospective Analysis. *Diagnostics*. 2023;13(3):552. doi:https://doi.org/10.3390/diagnostics13030552

32. Feng C, Zhou Z, Huang Q, Meng X, Li Z, Wang Y. Radiomics Nomogram Based on High-b-Value Diffusion-Weighted Imaging for Distinguishing the Grade of Bladder Cancer. *Life*. 2022;12(10):1510. doi:https://doi.org/10.3390/life12101510
33. Fiset S, Welch ML, Weiss J, et al. Repeatability and reproducibility of MRI-based radiomic features in cervical cancer. *Radiotherapy and Oncology*. 2019;135:107-114. doi:https://doi.org/10.1016/j.radonc.2019.03.001
34. Bernatz S, Ackermann J, Mandel P, et al. Comparison of machine learning algorithms to predict clinically significant prostate cancer of the peripheral zone with multiparametric MRI using clinical assessment categories and radiomic features. *European Radiology*. 2020;30(12):6757-6769. doi:https://doi.org/10.1007/s00330-020-07064-5
35. Tixier F, Um H, Bermudez D, et al. Preoperative MRI-radiomics features improve prediction of survival in glioblastoma patients over MGMT methylation status alone. *Oncotarget*. 2019;10(6):660-672. doi:https://doi.org/10.18632/oncotarget.26578
36. Metin Varan, Jahongir Azimjonov, Bilgen MaÇal. Enhancing Prostate Cancer Classification by Leveraging Key Radiomics Features and Using the Fine-Tuned Linear SVM Algorithm. *IEEE access*. 2023;11:88025-88039. doi:https://doi.org/10.1109/access.2023.3306515
37. Weißer C, Netzer N, Görtz M, et al. Weakly SupervisedMRISlice-Level Deep Learning Classification of Prostate Cancer Approximates Full Voxel- and Slice-Level Annotation: Effect of Increasing Training Set Size. *Journal of magnetic resonance imaging*. 2023;59(4):1409-1422. doi:https://doi.org/10.1002/jmri.28891
38. Bao J, Qiao X, Song Y, et al. Prediction of clinically significant prostate cancer using radiomics models in real-world clinical practice: a retrospective multicenter study. *Insights into imaging*. 2024;15(1). doi:https://doi.org/10.1186/s13244-024-01631-w
39. Jing G, Xing P, Li Z, et al. Prediction of clinically significant prostate cancer with a multimodal MRI-based radiomics nomogram. *Frontiers in Oncology*. 2022;12. doi:https://doi.org/10.3389/fonc.2022.918830
40. Li M, Chen T, Zhao W, et al. Radiomics prediction model for the improved diagnosis of clinically significant prostate cancer on biparametric MRI. *Quantitative Imaging in Medicine and Surgery*. 2020;10(2):368-379. doi:https://doi.org/10.21037/qims.2019.12.06
41. Algohary A, Viswanath S, Rakesh Shiradkar, et al. Radiomic features on MRI enable risk categorization of prostate cancer patients on active surveillance: Preliminary findings. *Journal of Magnetic Resonance Imaging*. 2018;48(3):818-828. doi:https://doi.org/10.1002/jmri.25983
42. Lapa P, Castelli M, Gonçalves I, Sala E, Rundo L. A Hybrid End-to-End Approach Integrating Conditional Random Fields into CNNs for Prostate Cancer Detection on MRI. *Applied Sciences*. 2020;10(1):338. doi:https://doi.org/10.3390/app10010338
43. Donisi L, Cesarelli G, Castaldo A, et al. A Combined Radiomics and Machine Learning Approach to Distinguish Clinically Significant Prostate Lesions on a Publicly Available MRI Dataset. *Journal of Imaging*. 2021;7(10):215. doi:https://doi.org/10.3390/jimaging7100215
44. Ogbonnaya CN, Zhang X, Alsaedi BSO, et al. Prediction of Clinically Significant Cancer Using Radiomics Features of Pre-Biopsy of Multiparametric MRI in Men Suspected of Prostate Cancer. *Cancers*. 2021;13(24):6199. doi:https://doi.org/10.3390/cancers13246199
45. Saha A, Hosseinzadeh M, Huisman H. End-to-end prostate cancer detection in bpMRI via 3D CNNs: Effects of attention mechanisms, clinical priori and decoupled false positive reduction. *Medical Image Analysis*. 2021;73(1361-8415):102155. doi:https://doi.org/10.1016/j.media.2021.102155

46. Chen T, Zhang Z, Tan S, et al. MRI Based Radiomics Compared With the PI-RADS V2.1 in the Prediction of Clinically Significant Prostate Cancer: Biparametric vs Multiparametric MRI. *Frontiers in Oncology*. 2022;11. doi:https://doi.org/10.3389/fonc.2021.792456
47. Castillo T. JM, Arif M, Starmans MPA, et al. Classification of Clinically Significant Prostate Cancer on Multi-Parametric MRI: A Validation Study Comparing Deep Learning and Radiomics. *Cancers*. 2021;14(1):12. doi:https://doi.org/10.3390/cancers14010012
48. Algohary A, Rakesh Shiradkar, Pahwa S, et al. Combination of Peri-Tumoral and Intra-Tumoral Radiomic Features on Bi-Parametric MRI Accurately Stratifies Prostate Cancer Risk: A Multi-Site Study. *Cancers*. 2020;12(8):2200-2200. doi:https://doi.org/10.3390/cancers12082200
49. Ji X, Zhang J, Shi W, et al. Bi-parametric magnetic resonance imaging based radiomics for the identification of benign and malignant prostate lesions: cross-vendor validation. *Physical and Engineering Sciences in Medicine*. 2021;44(3):745-754. doi:https://doi.org/10.1007/s13246-021-01022-1
50. Hu L, Da Wei Zhou, Cai Xia Fu, et al. Advanced zoomed diffusion-weighted imaging vs. full-field-of-view diffusion-weighted imaging in prostate cancer detection: a radiomic features study. *European radiology*. 2020;31(3):1760-1769. doi:https://doi.org/10.1007/s00330-020-07227-4
51. Gholizadeh N, Simpson J, Ramadan S, et al. Voxel-based supervised machine learning of peripheral zone prostate cancer using noncontrast multiparametric MRI. *Journal of applied clinical medical physics*. 2020;21(10):179-191. doi:https://doi.org/10.1002/acm2.12992
52. D McGarry S, D Bukowy J. Gleason Probability Maps: A Radiomics Tool for Mapping Prostate Cancer Likelihood in MRI Space. *Tomography*. 2019;5(1):127-134. doi:https://doi.org/10.18383/j.tom.2018.00033
53. Totaro A, Di Paola V, Campetella M, Scarciglia E, Boldrini L, Manfredi, R. Radiomic Features on Prostatic Multiparametric Magnetic Resonance Imaging Enable Progression Risk in Patients on Active Surveillance: A Pilot Study. *Journal of Radiology and Clinical Imaging*. 2022;05(04). doi:https://doi.org/10.26502/jrci.2809061
54. Castillo JM, Martijn P. A. Starmans, Niessen WJ, Ivo Schoots, Klein S, Veenland JF. Classification Of Prostate Cancer: High Grade Versus Low Grade Using A Radiomics Approach. *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy*. 2019;(1319-1322). doi:https://doi.org/10.1109/isbi.2019.8759217
55. Schwier M, Griethuysen van, Vangel M, et al. Repeatability of Multiparametric Prostate MRI Radiomics Features. *Scientific Reports*. 2019;9(1). doi:https://doi.org/10.1038/s41598-019-45766-z
56. Ginsburg SB, Algohary A, Pahwa S, et al. Radiomic features for prostate cancer detection on MRI differ between the transition and peripheral zones: Preliminary findings from a multi-institutional study. *Journal of Magnetic Resonance Imaging*. 2017;46(1):184-193. doi:https://doi.org/10.1002/jmri.25562
57. Giannini V, Mazzetti S, Defeudis A, et al. A Fully Automatic Artificial Intelligence System Able to Detect and Characterize Prostate Cancer Using Multiparametric MRI: Multicenter and Multi-Scanner Validation. *Frontiers in oncology*. 2021;11. doi:https://doi.org/10.3389/fonc.2021.718155
58. Castillo T. JM, Starmans MPA, Arif M, et al. A Multi-Center, Multi-Vendor Study to Evaluate the Generalizability of a Radiomics Model for Classifying Prostate cancer: High Grade vs. Low Grade. *Diagnostics*. 2021;11(2):369. doi:https://doi.org/10.3390/diagnostics11020369

59. Jeroen Bleker, Derya Yakar, Bram van Noort, et al. Single-center versus multi-center biparametric MRI radiomics approach for clinically significant peripheral zone prostate cancer. *Insights into imaging*. 2021;12(1). doi:https://doi.org/10.1186/s13244-021-01099-y
60. Bao J, Hou ying, Zhi R, et al. Integration of Clinical Identifications With Deep Transferrable Imaging Feature Representations Can Help Predict Prostate Cancer Aggressiveness and Outcome. *Research Square (Research Square)*. Published online February 11, 2021. doi:https://doi.org/10.21203/rs.3.rs-180726/v1
61. Castaldo R, Brancato V, Cavaliere C, et al. A Framework of Analysis to Facilitate the Harmonization of Multicenter Radiomic Features in Prostate Cancer. *Journal of clinical medicine*. 2022;12(1):140-140. doi:https://doi.org/10.3390/jcm12010140
62. Gholizadeh N, Fuangrod T, Greer PB, Lau P, Ramadan S, Simpson J. An inter-centre statistical scale standardisation for quantitatively evaluating prostate tissue on T2-weighted MRI. *Australasian Physical & Engineering Sciences in Medicine*. 2019;42(1):137-147. doi:https://doi.org/10.1007/s13246-019-00720-1
63. Li L, Gu L, Kang B, et al. Evaluation of the Efficiency of MRI-Based Radiomics Classifiers in the Diagnosis of Prostate Lesions. *Frontiers in oncology*. 2022;12. doi:https://doi.org/10.3389/fonc.2022.934108
64. Sanford T, Harmon S, Evrim Türkbey, et al. Deep-Learning-Based Artificial Intelligence for PI-RADS Classification to Assist Multiparametric Prostate MRI Interpretation: A Development Study. *Journal of Magnetic Resonance Imaging*. 2020;52(5):1499-1507. doi:https://doi.org/10.1002/jmri.27204
65. Damascelli A, Gallivanone F, Cristel G, et al. Advanced Imaging Analysis in Prostate MRI: Building a Radiomic Signature to Predict Tumor Aggressiveness. *Diagnostics*. 2021;11(4):594. doi:https://doi.org/10.3390/diagnostics11040594
66. Li M, Yang L, Yue Y, Xu J, Huang C, Song B. Use of Radiomics to Improve Diagnostic Performance of PI-RADS v2.1 in Prostate Cancer. *Frontiers in Oncology*. 2021;10(631831). doi:https://doi.org/10.3389/fonc.2020.631831
67. Nketiah GA, Elschot M, Scheenen TW, Maas MC, Bathen TF, Selnæs KM. Utility of T2-weighted MRI texture analysis in assessment of peripheral zone prostate cancer aggressiveness: a single-arm, multicenter study. *Scientific Reports*. 2021;11(1):2085. doi:https://doi.org/10.1038/s41598-021-81272-x
68. Zheng H, Miao Q, Liu Y, Raman SS, Scalzo F, Sung K. Integrative Machine Learning Prediction of Prostate Biopsy Results From Negative Multiparametric MRI. *Journal of Magnetic Resonance Imaging*. 2021;55(1):100-110. doi:https://doi.org/10.1002/jmri.27793
69. Monti S, Brancato V, Di Costanzo G, et al. Multiparametric MRI for Prostate Cancer Detection: New Insights into the Combined Use of a Radiomic Approach with Advanced Acquisition Protocol. *Cancers*. 2020;12(2):390. doi:https://doi.org/10.3390/cancers12020390
70. Alkadi R, Taher F, El-baz A, Werghi N. A Deep Learning-Based Approach for the Detection and Localization of Prostate Cancer in T2 Magnetic Resonance Images. *Journal of Digital Imaging*. 2018;32(5):793-807. doi:https://doi.org/10.1007/s10278-018-0160-1
71. Jin P, Yang L, Qiao X, et al. Utility of Clinical–Radiomic Model to Identify Clinically Significant Prostate Cancer in Biparametric MRI PI-RADS V2.1 Category 3 Lesions. *Frontiers in oncology*. 2022;12(840786). doi:https://doi.org/10.3389/fonc.2022.840786
72. Hectors SJ, Chen C, Chen J, et al. Magnetic Resonance Imaging Radiomics-Based Machine Learning Prediction of Clinically Significant Prostate Cancer in Equivocal PI-RADS 3 Lesions.

Journal of Magnetic Resonance Imaging. 2021;54(5)(1466–1473). doi:https://doi.org/10.1002/jmri.27692

73. Jin P, Shen J, Yang L, et al. Machine learning-based radiomics model to predict benign and malignant PI-RADS v2.1 category 3 lesions: a retrospective multi-center study. *BMC Medical Imaging*. 2023;23(1). doi:https://doi.org/10.1186/s12880-023-01002-9
74. Li T, Sun L, Li Q, et al. Development and Validation of a Radiomics Nomogram for Predicting Clinically Significant Prostate Cancer in PI-RADS 3 Lesions. *Frontiers in oncology*. 2022;11(825429). doi:https://doi.org/10.3389/fonc.2021.825429
75. Hou Y, Bao M, Wu CJ, Zhang J, Zhang Y, Shi H. A radiomics machine learning-based redefining score robustly identifies clinically significant prostate cancer in equivocal PI-RADS score 3 lesions. *Abdominal Imaging*. 2020;45(12):4223-4234. doi:https://doi.org/10.1007/s00261-020-02678-1
76. Corsi A, Elisabetta De Bernardi, Pietro Andrea Bonaffini, et al. Radiomics in PI-RADS 3 Multiparametric MRI for Prostate Cancer Identification: Literature Models Re-Implementation and Proposal of a Clinical–Radiological Model. *Journal of Clinical Medicine*. 2022;11(21):6304-6304. doi:https://doi.org/10.3390/jcm11216304
77. Brancato V, Aiello M, Basso L, et al. Evaluation of a multiparametric MRI radiomic-based approach for stratification of equivocal PI-RADS 3 and upgraded PI-RADS 4 prostatic lesions. *Scientific Reports*. 2021;11(1). doi:https://doi.org/10.1038/s41598-020-80749-5
78. Um H, Tixier F, Bermudez D, Deasy JO, Young RJ, Veeraraghavan H. Impact of image preprocessing on the scanner dependence of multi-parametric MRI radiomic features and covariate shift in multi-institutional glioblastoma datasets. *Physics in Medicine & Biology*. 2019;64(16):165011. doi:https://doi.org/10.1088/1361-6560/ab2f44
79. Moradmand H, Aghamiri SMR, Ghaderi R. Impact of image preprocessing methods on reproducibility of radiomic features in multimodal magnetic resonance imaging in glioblastoma. *Journal of Applied Clinical Medical Physics*. 2019;21(1):179-190. doi:https://doi.org/10.1002/acm2.12795
80. Shiri I, Hajianfar G, Sohrabi A, et al. Repeatability of radiomic features in magnetic resonance imaging of glioblastoma: Test–retest and image registration analyses. *Medical Physics*. 2020;47(9):4265-4280. doi:https://doi.org/10.1002/mp.14368
81. Castaldo R, Pane K, Nicolai E, Salvatore M, Franzese M. The Impact of Normalization Approaches to Automatically Detect Radiogenomic Phenotypes Characterizing Breast Cancer Receptors Status. *Cancers*. 2020;12(2):518. doi:https://doi.org/10.3390/cancers12020518
82. Carré A, Klausner G, Edjlali M, et al. Standardization of brain MR images across machines and protocols: bridging the gap for MRI-based radiomics. *Scientific Reports*. 2020;10(1). doi:https://doi.org/10.1038/s41598-020-69298-z
83. Bae S, Choi YS, Ahn SS, et al. Radiomic MRI Phenotyping of Glioblastoma: Improving Survival Prediction. *Radiology*. 2018;289(3):797-806. doi:https://doi.org/10.1148/radiol.2018180200
84. Kickingereder P, Burth S, Wick A, et al. Radiomic Profiling of Glioblastoma: Identifying an Imaging Predictor of Patient Survival with Improved Performance over Established Clinical and Radiologic Risk Models. *Radiology*. 2016;280(3):880-889. doi:https://doi.org/10.1148/radiol.2016160845
85. Osman AFI. A Multi-parametric MRI-Based Radiomics Signature and a Practical ML Model for Stratifying Glioblastoma Patients Based on Survival Toward Precision Oncology. *Frontiers in Computational Neuroscience*. 2019;13(58). doi:https://doi.org/10.3389/fncom.2019.00058

86. Chae Jung Park, Han K, Kim H, et al. Radiomics risk score may be a potential imaging biomarker for predicting survival in isocitrate dehydrogenase wild-type lower-grade gliomas. *European radiology*. 2020;30(12):6464-6474. doi:https://doi.org/10.1007/s00330-020-07089-w
87. Tan Y, Mu W, Wang X, Yang G, Robert James Gillies, Zhang H. Improving survival prediction of high-grade glioma via machine learning techniques based on MRI radiomic, genetic and clinical risk factors. *European Journal of Radiology*. 2019;120(108609):108609-108609. doi:https://doi.org/10.1016/j.ejrad.2019.07.010
88. Yan J, Zhang S, Li KKW, et al. Incremental prognostic value and underlying biological pathways of radiomics patterns in medulloblastoma. *EBioMedicine*. 2020;61(103093):103093. doi:https://doi.org/10.1016/j.ebiom.2020.103093
89. Stringfield O. Multiparameter MRI Predictors of Long-Term Survival in Glioblastoma Multiforme. *Tomography*. 2019;5(1):135-144. doi:https://doi.org/10.18383/j.tom.2018.00052
90. Philipp Kickingereder, Neuberger U, Bonekamp D, et al. Radiomic subtyping improves disease stratification beyond key molecular, clinical, and standard imaging characteristics in patients with glioblastoma. *Neuro-oncology*. 2018;20(6):848-857. doi:https://doi.org/10.1093/neuonc/nox188
91. Miccò M, Gui B, Russo L, et al. Preoperative Tumor Texture Analysis on MRI for High-Risk Disease Prediction in Endometrial Cancer: A Hypothesis-Generating Study. *Journal of personalized medicine*. 2022;12(11):1854-1854. doi:https://doi.org/10.3390/jpm12111854
92. Li X, Dessi M, Marcus D, et al. Prediction of Deep Myometrial Infiltration, Clinical Risk Category, Histological Type, and Lymphovascular Space Invasion in Women with Endometrial Cancer Based on Clinical and T2-Weighted MRI Radiomic Features. *Cancers*. 2023;15(8):2209-2209. doi:https://doi.org/10.3390/cancers15082209
93. Ren J, Li Y, Yang J, et al. MRI-based radiomics analysis improves preoperative diagnostic performance for the depth of stromal invasion in patients with early stage cervical cancer. *Insights into Imaging*. 2022;13(1). doi:https://doi.org/10.1186/s13244-022-01156-0
94. A Bobholz S. Radiomic Features of Multiparametric MRI Present Stable Associations With Analogous Histological Features in Patients With Brain Cancer. *Tomography*. 2020;6(2):160-169. doi:https://doi.org/10.18383/j.tom.2019.00029
95. Chen J, Gu H, Fan W, et al. MRI-Based Radiomic Model for Preoperative Risk stratification in Stage I Endometrial Cancer. *Journal of Cancer*. 2021;12(3):726-734. doi:https://doi.org/10.7150/jca.50872
96. Lefebvre TL, Ueno Y, Dohan A, et al. Development and Validation of Multiparametric MRI–based Radiomics Models for Preoperative Risk Stratification of Endometrial Cancer. *Radiology*. 2022;305(2):375-386. doi:https://doi.org/10.1148/radiol.212873
97. Brancato V, Garbino N, Salvatore M, Cavaliere C. MRI-Based Radiomic Features Help Identify Lesions and Predict Histopathological Grade of Hepatocellular Carcinoma. *Diagnostics*. 2022;12(5):1085. doi:https://doi.org/10.3390/diagnostics12051085
98. Hodneland E, Kaliyugarasan S, Wagner-Larsen KS, et al. Fully Automatic Whole-Volume Tumor Segmentation in Cervical Cancer. *Cancers*. 2022;14(10):2372. doi:https://doi.org/10.3390/cancers14102372
99. Daimiel Naranjo I, Gibbs P, Reiner JS, et al. Radiomics and Machine Learning with Multiparametric Breast MRI for Improved Diagnostic Accuracy in Breast Cancer Diagnosis. *Diagnostics*. 2021;11(6):919. doi:https://doi.org/10.3390/diagnostics11060919

100. Conte L, Tafuri B, Portaluri M, Galiano A, Maggiulli E, De Nunzio G. Breast Cancer Mass Detection in DCE–MRI Using Deep-Learning Features Followed by Discrimination of Infiltrative vs. In Situ Carcinoma through a Machine-Learning Approach. *Applied Sciences*. 2020;10(17):6109. doi:https://doi.org/10.3390/app10176109
101. Wang H, Song B, Ye N, et al. Machine learning-based multiparametric MRI radiomics for predicting the aggressiveness of papillary thyroid carcinoma. *European Journal of Radiology*. 2020;122(108755):108755-108755. doi:https://doi.org/10.1016/j.ejrad.2019.108755
102. Xie H, Hu J, Zhang X, Ma S, Liu Y, Wang X. Preliminary utilization of radiomics in differentiating uterine sarcoma from atypical leiomyoma: Comparison on diagnostic efficacy of MRI features and radiomic features. *European Journal of Radiology*. 2019;115(39–45):39-45. doi:https://doi.org/10.1016/j.ejrad.2019.04.004
103. Stanzione A, Ricciardi C, Cuocolo R, et al. MRI Radiomics for the Prediction of Fuhrman Grade in Clear Cell Renal Cell Carcinoma: a Machine Learning Exploratory Study. *Journal of Digital Imaging*. 2020;33(4):879-887. doi:https://doi.org/10.1007/s10278-020-00336-y
104. Yu X, Song W, Guo D, et al. Preoperative Prediction of Extramural Venous Invasion in Rectal Cancer: Comparison of the Diagnostic Efficacy of Radiomics Models and Quantitative Dynamic Contrast-Enhanced Magnetic Resonance Imaging. *Frontiers in oncology*. 2020;10(459). doi:https://doi.org/10.3389/fonc.2020.00459
105. Du P, Liu X, Wu X, Chen J, Cao A, Geng D. Predicting Histopathological Grading of Adult Gliomas Based On Preoperative Conventional Multimodal MRI Radiomics: A Machine Learning Model. *Brain sciences*. 2023;13(6):912-912. doi:https://doi.org/10.3390/brainsci13060912
106. Leithner D, Mayerhoefer ME, Martinez DF, et al. Non-Invasive Assessment of Breast Cancer Molecular Subtypes with Multiparametric Magnetic Resonance Imaging Radiomics. *Journal of Clinical Medicine*. 2020;9(6):1853. doi:https://doi.org/10.3390/jcm9061853
107. Peeken JC, Spraker MB, Knebel C, et al. Tumor grading of soft tissue sarcomas using MRI-based radiomics. *EBioMedicine*. 2019;48(332–340):332-340. doi:https://doi.org/10.1016/j.ebiom.2019.08.059
108. Yan R, Hao D, Li J, et al. Magnetic Resonance Imaging-Based Radiomics Nomogram for Prediction of the Histopathological Grade of Soft Tissue Sarcomas: A Two-Center Study. *Journal of Magnetic Resonance Imaging*. 2021;53(6):1683-1696. doi:https://doi.org/10.1002/jmri.27532
109. Cho H, Kim H, Sang Yu Nam, et al. Measurement of Perfusion Heterogeneity within Tumor Habitats on Magnetic Resonance Imaging and Its Association with Prognosis in Breast Cancer Patients. *Cancers*. 2022;14(8):1858-1858. doi:https://doi.org/10.3390/cancers14081858
110. Ming W, Zhu Y, Li F, et al. Identifying Associations between DCE-MRI Radiomic Features and Expression Heterogeneity of Hallmark Pathways in Breast Cancer: A Multi-Center Radiogenomic Study. *Genes*. 2022;14(1):28-28. doi:https://doi.org/10.3390/genes14010028
111. Ming W, Li F, Zhu Y, et al. Unsupervised Analysis Based on DCE-MRI Radiomics Features Revealed Three Novel Breast Cancer Subtypes with Distinct Clinical Outcomes and Biological Characteristics. *Cancers*. 2022;14(22):5507. doi:https://doi.org/10.3390/cancers14225507
112. Bologna M, Corino V, Calareso G, et al. Baseline MRI-Radiomics Can Predict Overall Survival in Non-Endemic EBV-Related Nasopharyngeal Carcinoma Patients. *Cancers*. 2020;12(10):2958-2958. doi:https://doi.org/10.3390/cancers12102958

113. Wang XH, Long LH, Cui Y, et al. MRI-based radiomics model for preoperative prediction of 5-year survival in patients with hepatocellular carcinoma. *British Journal of Cancer*. 2020;122(7):978-985. doi:https://doi.org/10.1038/s41416-019-0706-0
114. Rathore S, Akbari H, Rozycki M, et al. Radiomic MRI signature reveals three distinct subtypes of glioblastoma with different clinical and molecular characteristics, offering prognostic value beyond IDH1. *Scientific Reports*. 2018;8(1). doi:https://doi.org/10.1038/s41598-018-22739-2
115. Fasmer KE, Hodneland E, Dybvik JA, et al. Whole-Volume Tumor MRI Radiomics for Prognostic Modeling in Endometrial Cancer. *Journal of Magnetic Resonance Imaging*. 2020;53(3):928-937. doi:https://doi.org/10.1002/jmri.27444

Appendices

A. N4 Bias Field Correction

```
import SimpleITK as sitk
import sys
import os
import pydicom
import numpy as np

def resample_mask(input_image_path, mask_image_path):
    # Read the input and mask images
    input_image = sitk.ReadImage(input_image_path)
    mask_image = sitk.ReadImage(mask_image_path)

    # Create a resampler
    resampler = sitk.ResampleImageFilter()

    # Set the resampler parameters
    resampler.SetSize(input_image.GetSize())
    resampler.SetOutputOrigin(input_image.GetOrigin())
    resampler.SetOutputSpacing(input_image.GetSpacing())
    resampler.SetOutputDirection(input_image.GetDirection())
    resampler.SetInterpolator(sitk.sitkNearestNeighbor)

    # Resample the mask image
    resampled_mask_image = resampler.Execute(mask_image)

    # Now, resampled_mask_image should have the same size as
input_image
    return resampled_mask_image

# Function to find the path of the mask corresponding to an
image with the same ID
def find_segmentation_mask_path(image_path,
segmentations_root_directory):

    # Get the patient ID from the image file path
    patient_id =
os.path.basename(os.path.dirname(os.path.dirname(image_path)))
    print(patient_id)
    # Search for the segmentation mask file within the
segmentations root directory
    for root, dirs, files in
os.walk(os.path.join(segmentations_root_directory,
patient_id)):
        for file in files:
            if file.endswith(".dcm"):  # Check for DICOM files
```

```
                # Return the full path of the segmentation
mask file
                return os.path.join(root, file)

    # If no matching segmentation mask is found, return None
    return None

# Function to perform N4 corection for a DICOM series of one
patient
def process_series(dicom_series_path, output_folder_path,
shrink_factor, mask_image_path, num_iterations, num_levels):

    print(f"Processing series in: {dicom_series_path}")
    # List all files in the specified directory
    all_files = os.listdir(dicom_series_path)

    # Filter DICOM files
    dicom_files = [file for file in all_files if
pydicom.dcmread(os.path.join(dicom_series_path, file),
stop_before_pixels=True).SOPClassUID ==
'1.2.840.10008.5.1.4.1.1.4']

    if not dicom_files:
        print(f"Error: No DICOM files found in
{dicom_series_path}. Skipping this series.")
        return

    # Sort DICOM files by Number
    dicom_files.sort(key=lambda x:
pydicom.dcmread(os.path.join(dicom_series_path,
x)).InstanceNumber)

    # Read the first DICOM file to get metadata
    first_dicom =
pydicom.dcmread(os.path.join(dicom_series_path,
dicom_files[0]))

    # Read all DICOM slices to create a 3D volume
    reader = sitk.ImageSeriesReader()
    dicom_series =
reader.GetGDCMSeriesFileNames(dicom_series_path)
    reader.SetFileNames(dicom_series)

    # Use float32 pixel type for conversion
    pixel_type = sitk.sitkFloat32
    input_volume = sitk.Cast(reader.Execute(), pixel_type)


    # Shrink if necessary
```

```
    if shrink_factor > 1:
        input_volume = sitk.Shrink(input_volume,
[shrink_factor] * input_volume.GetDimension())

    # Create a mask if specified, or use an Otsu threshold
    if mask_image_path and os.path.isfile(mask_image_path):
        maskImage = sitk.ReadImage(mask_image_path,
sitk.sitkUInt8)
    else:
        maskImage = sitk.OtsuThreshold(input_volume, 0, 1,
200)

    # Create a corrector instance
    corrector = sitk.N4BiasFieldCorrectionImageFilter()

    # Resolution levels at which the image is processed
    numberFittingLevels = 4

    if num_levels > 0:
        numberFittingLevels = num_levels

    if num_iterations > 0:

corrector.SetMaximumNumberOfIterations([num_iterations] *
numberFittingLevels)

    # Execute bias field correction on the 3D volume
    print("Execute bias field correction on the 3D volume")
    try:
        corrected_volume = corrector.Execute(input_volume,
maskImage)
        print("finished bias field correction")
    except Exception as e:
        print(f"Error during bias field correction: {str(e)}")


    # Convert the corrected volume back to the original pixel
type
    corrected_volume = sitk.Cast(corrected_volume,
input_volume.GetPixelID())

     # Specify the output path for corrected images
    output_image_path = os.path.join(output_folder_path,
f"{os.path.basename(dicom_series_path)}_N4_corrected.nrrd")
    os.makedirs(output_folder_path, exist_ok=True)  # Create
the output folder if it doesn't exist

    # Write the corrected volume to the specified output file
    sitk.WriteImage(corrected_volume, output_image_path)
```

```
    if shrink_factor > 1:
        shrunk_output_path =
os.path.splitext(output_image_path)[0] + "-shrunk.nrrd"
        sitk.WriteImage(corrected_volume, shrunk_output_path)

    print(f"Processing completed for {dicom_series_path}.
Output saved to {output_image_path}")

#Function to iterate through all patients and perform N4 by
utilizing the previous function
def main(data_path, output_path, shrink_factor=1,
masks_root_directory=None, num_iterations=50, num_levels=4):
    # List all patient directories
    patient_directories = [d for d in os.listdir(data_path) if
os.path.isdir(os.path.join(data_path, d))]

    for patient_dir in patient_directories:
        # Construct the full path for the patient directory
        patient_full_path = os.path.join(data_path,
patient_dir)

        # Find DICOM image series folder within the patient
directory
        dicom_series_folder = next((f for f in
os.listdir(patient_full_path) if
os.path.isdir(os.path.join(patient_full_path, f))), None)

        if dicom_series_folder:
            # Construct the full path for the DICOM image
series folder
            dicom_series_path =
os.path.join(patient_full_path, dicom_series_folder)

            # Specify the output path for corrected images for
this patient
            output_patient_path = os.path.join(output_path,
patient_dir)
            os.makedirs(output_patient_path, exist_ok=True)  #
Create a folder for each patient

            # Specify the output path for corrected images
            output_image_path =
os.path.join(output_patient_path,
f"{patient_dir}_N4_corrected.nrrd")

            image_mask =
find_segmentation_mask_path(dicom_series_path,
masks_root_directory)
```

```
                # Perform bias field correction
                process_series(dicom_series_path,
    output_image_path, shrink_factor, image_mask, num_iterations,
    num_levels)

        print("Finished processing all images")

#Data
patient_folder_path = r"/content/drive/MyDrive/data/PROSTATEx"
output_base_folder_path =
r"/content/drive/MyDrive/data/N4_corrected"
mask_images_path = r"/content/drive/MyDrive/data/Segmentations"
shrink_factor = 1 # default=1
num_iterations = 10 # default = 50
num_levels = 4 #default=4

#Run
main(patient_folder_path, output_base_folder_path, shrink_factor,
mask_images_path, num_iterations, num_levels)
```

B. Before and After N4 image subtraction & Histogram

```
import os
import SimpleITK as sitk
import matplotlib.pyplot as plt

def subtract_images(before_dicom_series_path, after_nrrd_path,
output_path):
    # Load the before DICOM series
    before_reader = sitk.ImageSeriesReader()
    before_dicom_series =
before_reader.GetGDCMSeriesFileNames(before_dicom_series_path)
    before_reader.SetFileNames(before_dicom_series)
    before_image = before_reader.Execute()

    after_image = sitk.ReadImage(after_nrrd_path)

    # Cast pixel type of before_image to match after_image
    before_image = sitk.Cast(before_image, after_image.GetPixelID())

    # Subtract the after image from the before image
    difference_image = sitk.Subtract(after_image, before_image)

    # Write the difference image to disk
    sitk.WriteImage(difference_image, output_path)
```

```
# Paths to the DICOM series before bias field correction and the
NRRD image after correction
before_dicom_series_path =
"/content/drive/MyDrive/data/PROSTATEx/ProstateX-0004/5.000000-t2tse
tra-75680"
after_nrrd_path =
"/content/drive/MyDrive/data/5.000000-t2tsetra-75680_N4_corrected.nr
rd"
# Output path for the difference image
output_path = "/content/drive/MyDrive/data/try2.nrrd"

# Perform subtraction
subtract_images(before_dicom_series_path, after_nrrd_path,
output_path)

def plot_histogram(image):
    # Flatten the image pixel values
    pixel_values = sitk.GetArrayViewFromImage(image).flatten()

    # Plot histogram
    plt.hist(pixel_values, bins=50, color='blue', alpha=0.7)
    plt.xlabel('Pixel Value Difference')
    plt.ylabel('Frequency')
    plt.title('Histogram of Pixel Value Differences')
    plt.show()

# Create histogram
subtracted_image = sitk.ReadImage(output_path)
plot_histogram(subtracted_image)
```

C. Normalization, Resampling & Radiomics Feature Extraction

```
import os
import SimpleITK as sitk
import pandas as pd
from radiomics import featureextractor
import nrrd
import numpy as np

#Function to get all paths of the N4 corrected images
def get_image_paths(root_directory):
    image_paths = []

    # Iterate through patient folders
    for patient_folder in os.listdir(root_directory):
        # Construct the full path to the patient's directory
```

```
        patient_directory = os.path.join(root_directory,
patient_folder)

        # Check if the patient's directory exists
        if os.path.exists(patient_directory) and
os.path.isdir(patient_directory):
            # List all files in the patient's directory
            files_in_patient_directory =
os.listdir(patient_directory)

            # Iterate through files in the patient's directory
            for file_name in files_in_patient_directory:
                # Check if the file is a directory
                if os.path.isdir(os.path.join(patient_directory,
file_name)):
                    # Construct the full path to the subdirectory
containing the image file
                    subdirectory_path =
os.path.join(patient_directory, file_name)

                    # List all files in the subdirectory
                    files_in_subdirectory =
os.listdir(subdirectory_path)

                    # Iterate through files in the subdirectory
                    for subdirectory_file_name in
files_in_subdirectory:
                        # Check if the file is an image file
                        if subdirectory_file_name.endswith((".nrrd",
".dcm")):
                            # Construct the full path to the image
file
                            image_file_path =
os.path.join(subdirectory_path, subdirectory_file_name)
                            image_paths.append(image_file_path)

    return image_paths

# Function to find the path of the mask corresponding to an image
with the same ID
def find_segmentation_mask_path(image_path,
segmentations_root_directory):

    # Get the patient ID from the image file path
```

```
    patient_id =
os.path.basename(os.path.dirname(os.path.dirname(image_path)))
    print(patient_id)
    # Search for the segmentation mask file within the segmentations
root directory
    for root, dirs, files in
os.walk(os.path.join(segmentations_root_directory, patient_id)):
        for file in files:
            if file.endswith(".dcm"):  # Check for DICOM files
                # Return the full path of the segmentation mask file
                return os.path.join(root, file)

    # If no matching segmentation mask is found, return None
    return None

# Function for Z-score normalization of one image
def z_score_normalize(image):
    # Extract image data as a numpy array
    image_data = sitk.GetArrayFromImage(image)

    # Calculate the mean and standard deviation of intensity
    mean_intensity = np.mean(image_data)
    std_intensity = np.std(image_data)

    # Apply Z-score normalization
    normalized_image_data = (image_data - mean_intensity) /
std_intensity

    # Convert normalized data back to SimpleITK image
    normalized_image = sitk.GetImageFromArray(normalized_image_data)
    normalized_image.CopyInformation(image)  # Copy metadata from
the original image

    return normalized_image

#Function to resample a single image
def resample_image(image, reference_image):

    # Create an instance of the ResampleImageFilter
    resampler = sitk.ResampleImageFilter()

    # Set the parameters for resampling the input image to match the
properties of the reference_image
    resampler.SetSize(reference_image.GetSize())
    resampler.SetOutputSpacing(reference_image.GetSpacing())
```

```
    resampler.SetOutputOrigin(reference_image.GetOrigin())
    resampler.SetOutputDirection(reference_image.GetDirection())

    # Execute the resampling process using the configured
ResampleImageFilter
    resampled_image = resampler.Execute(image)

    # Return the resampled image
    return resampled_image

# Function to extract radiomics features from a single image
def extract_radiomic_features_one(corrected_image_path,
segmentation_mask_path, output_csv):

    # Check if both files exist
    if os.path.exists(corrected_image_path) and
os.path.exists(segmentation_mask_path):
        # Print paths the image and the mask
        print(f"Corrected Image Path: {corrected_image_path}")
        print(f"Segmentation Mask Path: {segmentation_mask_path}")

        try:
            # Load DICOM segmentation mask
            segmentation_mask =
sitk.ReadImage(segmentation_mask_path)

            # Load NRRD corrected array and turn it to an image
            corrected_image_array, corrected_header =
nrrd.read(corrected_image_path)
            corrected_image =
sitk.GetImageFromArray(corrected_image_array)

            # Apply Z-score normalization
            normalized_image = z_score_normalize(corrected_image)

            # Resample the corrected image to match the spacing,
origin, and direction of the segmentation mask
            resampled_normalized_image =
resample_image(normalized_image, segmentation_mask)

            # Ensure the image and mask have the same dimensions
after resampling
            if resampled_normalized_image.GetSize() ==
segmentation_mask.GetSize():
                # Apply the mask to the resampled image
```

```
                masked_image = sitk.Mask(resampled_normalized_image,
segmentation_mask)

                # Feature extraction
                extractor =
featureextractor.RadiomicsFeatureExtractor()
                extractor.enableAllFeatures()
                extractor.settings['label'] = 255  # Set the correct
label
                extractor.settings['disableAll2D'] = True # All
images are 3D

                # Extract radiomic features
                features =
extractor.execute(resampled_normalized_image, segmentation_mask)

                # Convert the features to a DataFrame
                features_df = pd.DataFrame(list(features.items()),
columns=['Feature', 'Value'])

                print("Radiomic features extracted successfully.")

                # Return the features DataFrame
                return features_df
            else:
                raise Exception("Error: Resampled image and mask
dimensions do not match.")


        except Exception as e:
            raise Exception(f"Error during radiomic feature
extraction: {str(e)}")

    else:
        raise Exception("Error: One or both files do not exist.")

# Function to iterate through all images and extract the
corresponding radiomic features
def extract_radiomic_features_for_all_images(patient_root_directory,
segmentations_root_directory, output_csv_path):

    # Get a list of image paths using the previously defined
function
    features_df = pd.DataFrame()
    image_paths = get_image_paths(patient_root_directory)
```

```
    # Create an empty DataFrame to store results
    results_df = pd.DataFrame()

    # Iterate through image paths
    for image_path in image_paths:
        # Call the function to find the segmentation mask path for
the current image
        segmentation_mask_path =
find_segmentation_mask_path(image_path,
segmentations_root_directory)

        if segmentation_mask_path is not None:
            # Call the function to extract radiomic features for the
current image and segmentation mask
            features_df = extract_radiomic_features_one(image_path,
segmentation_mask_path, output_csv_path)

            # Add patient information to the DataFrame
            patient_id =
os.path.basename(os.path.dirname(image_path))
            features_df['PatientID'] = patient_id

            # Append the results to the main DataFrame
            results_df = pd.concat([results_df, features_df],
ignore_index=True)
        else:
            print(f"Warning: No segmentation mask found for image
{image_path}")

    # Save the results to a single CSV file
    results_df.to_csv(output_csv_path, index=False)

# Directory containg patients N4 corrected image files
patient_root_directory = "/content/drive/MyDrive/data/N4_corrected"
# Directory containing patients corresponding segmentation files
segmentations_root_directory =
"/content/drive/MyDrive/data/Segmentations"
# Radiomics features output CSV file
output_csv_path =
"/content/drive/MyDrive/data/with_Norm_After_N4/Final-Radiomics.csv"

# run :
extract_radiomic_features_for_all_images(patient_root_directory,
segmentations_root_directory, output_csv_path)
```

```
# Open the CSV to reshape
df =
pd.read_csv("/content/drive/MyDrive/data/with_Norm_After_N4/Final-Ra
diomics.csv")

# Pivot the DataFrame to reshape it
df_pivoted = df.pivot(index='PatientID', columns='Feature',
values='Value').reset_index()

# Save the pivoted DataFrame to a new CSV file. Now each row
corresponds to one patient.
df_pivoted.to_csv("/content/drive/MyDrive/data/with_Norm_After_N4/Fi
nal-Radiomics_pivoted_data.csv", index=False)
```

D. Logistic Regression Classification model

```
import numpy as np
import pandas as pd
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split,
RepeatedStratifiedKFold, RandomizedSearchCV
from sklearn.pipeline import Pipeline
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
from sklearn.feature_selection import SelectFromModel
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import precision_recall_curve, precision_score,
recall_score, f1_score, roc_auc_score, accuracy_score
from scipy.stats import uniform
import joblib
import ast

# Load the dataset
df =
pd.read_csv('/content/drive/MyDrive/data/noregion_merged_radiomics_d
ataframe.csv')
df.head()

df = df.drop(columns=['PatientID',
'diagnostics_Configuration_EnabledImageTypes',
'diagnostics_Configuration_Settings',
'diagnostics_Image-original_Dimensionality',
'diagnostics_Image-original_Hash',
'diagnostics_Mask-original_BoundingBox',
```

```
'diagnostics_Mask-original_CenterOfMass',
'diagnostics_Mask-original_CenterOfMassIndex',
'diagnostics_Mask-original_Hash', 'diagnostics_Mask-original_Size',
'diagnostics_Mask-original_Spacing',
'diagnostics_Mask-original_VolumeNum',
'diagnostics_Mask-original_VoxelNum', 'diagnostics_Versions_Numpy',
'diagnostics_Versions_PyRadiomics',
'diagnostics_Versions_PyWavelet', 'diagnostics_Versions_Python',
'diagnostics_Versions_SimpleITK'])

df.head()

# It apperas that 2 columns are 3 - dimensional.
# Check the length of the tuples in
'diagnostics_Image-original_Size'
df['diagnostics_Image-original_Size'].apply(lambda x:
len(ast.literal_eval(x))).value_counts()

# We can break it down into 3 columns and then drop the original
df[['Size_dim1', 'Size_dim2', 'Size_dim3']] =
pd.DataFrame(df['diagnostics_Image-original_Size'].apply(ast.literal
_eval).tolist(), index= df.index)
df = df.drop(columns=['diagnostics_Image-original_Size'])

df.head()

# Now we do the same for the column
'diagnostics_Image-original_Spacing'
# First we check that all the data is 3 dimensional
df['diagnostics_Image-original_Spacing'].apply(lambda x:
len(ast.literal_eval(x))).value_counts()

# We can break it down into 3 columns as well and then drop the
original
df[['Spacing_dim1', 'Spacing_dim2', 'Spacing_dim3']] =
pd.DataFrame(df['diagnostics_Image-original_Spacing'].apply(ast.lite
ral_eval).tolist(), index= df.index)
df = df.drop(columns=['diagnostics_Image-original_Spacing'])

df.head()

#Finding the best classifier
# Define classifiers
classifiers = [
    ('SVM', SVC(probability=True)),
```

```
    ('Random Forest', RandomForestClassifier()),
    ('Logistic Regression', LogisticRegression()),
    ('KNN', KNeighborsClassifier()),
    ('Gradient Boosting', GradientBoostingClassifier())
]

# Split data into features (X) and target variable (y)
X = df.drop(['ProxID', 'ClinSig', 'label'], axis=1)
y = df['label']

# Define the number of splits for stratified k-fold cross-validation
n_splits = 5

# Initialize lists to store performance metrics across classifiers
classifier_performance = []

for classifier_name, classifier in classifiers:
    # Initialize stratified k-fold cross-validation
    skf = StratifiedKFold(n_splits=n_splits, shuffle=True,
random_state=42)

    # Lists to store performance metrics across folds
    cv_mean_accuracy = []
    cv_std_accuracy = []
    cv_mean_precision = []
    cv_std_precision = []
    cv_mean_recall = []
    cv_std_recall = []
    cv_mean_f1 = []
    cv_std_f1 = []
    cv_mean_auc = []
    cv_std_auc = []
    cv_mean_fbeta = []
    cv_std_fbeta = []

    for train_index, test_index in skf.split(X, y):
        # Split data into train and test sets for this fold
        X_train, X_test = X.iloc[train_index], X.iloc[test_index]
        y_train, y_test = y.iloc[train_index], y.iloc[test_index]

        # Create pipeline
        pipeline = Pipeline([('scaler', StandardScaler()),
                             ('PCA', PCA(n_components=0.95)),
                             ('feature_selection',
SelectFromModel(RandomForestClassifier(n_estimators=100))),
```

```
                            (classifier_name, classifier)])

        # Fit the pipeline to the training data
        pipeline.fit(X_train, y_train)

        # Predict the labels of the test set
        y_pred = pipeline.predict(X_test)
        y_proba = pipeline.predict_proba(X_test)[:, 1]

        # Calculate performance metrics for this fold
        accuracy = accuracy_score(y_test, y_pred)
        precision = precision_score(y_test, y_pred, zero_division=0)
        recall = recall_score(y_test, y_pred)
        f1 = f1_score(y_test, y_pred)
        auc = roc_auc_score(y_test, y_proba)
        fbeta = fbeta_score(y_test, y_pred, beta=0.5)

        cv_mean_accuracy.append(accuracy)
        cv_mean_precision.append(precision)
        cv_mean_recall.append(recall)
        cv_mean_f1.append(f1)
        cv_mean_auc.append(auc)
        cv_mean_fbeta.append(fbeta)

    # Calculate mean and standard deviation of metrics across folds
    mean_accuracy = np.mean(cv_mean_accuracy)
    std_accuracy = np.std(cv_mean_accuracy)
    mean_precision = np.mean(cv_mean_precision)
    std_precision = np.std(cv_mean_precision)
    mean_recall = np.mean(cv_mean_recall)
    std_recall = np.std(cv_mean_recall)
    mean_f1 = np.mean(cv_mean_f1)
    std_f1 = np.std(cv_mean_f1)
    mean_auc = np.mean(cv_mean_auc)
    std_auc = np.std(cv_mean_auc)
    mean_fbeta = np.mean(cv_mean_fbeta)
    std_fbeta = np.std(cv_mean_fbeta)

    # Store performance metrics for this classifier
    classifier_performance.append({
        'Classifier': classifier_name,
        'Mean Accuracy': mean_accuracy,
        'Std Accuracy': std_accuracy,
        'Mean Precision': mean_precision,
        'Std Precision': std_precision,
```

```
        'Mean Recall': mean_recall,
        'Std Recall': std_recall,
        'Mean F1 Score': mean_f1,
        'Std F1 Score': std_f1,
        'Mean AUC': mean_auc,
        'Std AUC': std_auc,
        'Mean F-beta Score': mean_fbeta,
        'Std F-beta Score': std_fbeta
    })

# Print performance metrics for each classifier
for perf in classifier_performance:
    print(f"Performance metrics for {perf['Classifier']}:")
    print(f"Mean Cross-validation Accuracy: {perf['Mean Accuracy']}
(Std: {perf['Std Accuracy']})")
    print(f"Mean Cross-validation Precision: {perf['Mean
Precision']} (Std: {perf['Std Precision']})")
    print(f"Mean Cross-validation Recall: {perf['Mean Recall']}
(Std: {perf['Std Recall']})")
    print(f"Mean Cross-validation F1 Score: {perf['Mean F1 Score']}
(Std: {perf['Std F1 Score']})")
    print(f"Mean Cross-validation AUC: {perf['Mean AUC']} (Std:
{perf['Std AUC']})")
    print(f"Mean Cross-validation F-beta Score: {perf['Mean F-beta
Score']} (Std: {perf['Std F-beta Score']})")

#We choose Logistic Regression and we move forward

#Best test-holdout model

#Split the dataframe into features, X, and labels, y
X = df.drop(['ProxID', 'ClinSig', 'label'], axis=1)
y = df['label']

# Split the data into training and holdout sets
X_train, X_holdout, y_train, y_holdout = train_test_split(X, y,
test_size=0.2, stratify=y, random_state=42)

# Check class distribution in the training and holdout sets
print("Class distribution in training set:", np.bincount(y_train))
print("Class distribution in holdout set:", np.bincount(y_holdout))

# Define the pipeline
pipeline = Pipeline([
    ('scaler', StandardScaler()),
```

```
    ('PCA', PCA(random_state=42)),
    ('feature_selection',
SelectFromModel(RandomForestClassifier(random_state=42))),
    ('classifier', LogisticRegression(random_state=42))
])

# Define parameter distributions for RandomizedSearchCV
param_dist = {
    'PCA__n_components': uniform(0.8, 0.199),  # Ensure values are
between 0.8 and 0.999
    'feature_selection__estimator__n_estimators': [50, 100, 200],
    'classifier__C': uniform(0.01, 10)
}

# Setup RandomizedSearchCV
rkf = RepeatedStratifiedKFold(n_splits=5, n_repeats=3,
random_state=42)
random_search = RandomizedSearchCV(pipeline,
param_distributions=param_dist, n_iter=50, cv=rkf, scoring='f1',
random_state=42, n_jobs=-1, error_score='raise')
random_search.fit(X_train, y_train)

# Print the best parameters and the best score
best_params = random_search.best_params_
best_score = random_search.best_score_

print(f"Best parameters found: {best_params}")
print(f"Best cross-validation score: {best_score}")

# Access the best estimator
best_model = random_search.best_estimator_

# Train the best model on the entire training set
best_model.fit(X_train, y_train)

# Predict probabilities on the training set
y_train_proba = best_model.predict_proba(X_train)[:, 1]

# Predict probabilities on the holdout set
y_holdout_proba = best_model.predict_proba(X_holdout)[:, 1]

# Calculate precision-recall curve for training set
precision_train, recall_train, thresholds_train =
precision_recall_curve(y_train, y_train_proba)
```

```
# Set desired precision level
desired_precision = 0.8

# Find the optimal threshold for the desired precision
optimal_threshold_train = thresholds_train[np.argmax(precision_train
== desired_precision)]

# Predict with the optimal threshold for training set
y_train_pred_optimal = (y_train_proba >=
optimal_threshold_train).astype(int)

# Calculate performance metrics for training set
train_accuracy = accuracy_score(y_train, y_train_pred_optimal)
train_precision = precision_score(y_train, y_train_pred_optimal)
train_recall = recall_score(y_train, y_train_pred_optimal)
train_f1 = f1_score(y_train, y_train_pred_optimal)
train_roc_auc = roc_auc_score(y_train, y_train_proba)

print("Training Set Performance Metrics with Optimal Threshold:")
print(f"Accuracy: {train_accuracy}")
print(f"Precision: {train_precision}")
print(f"Recall: {train_recall}")
print(f"F1 Score: {train_f1}")
print(f"AUC: {train_roc_auc}")

# Predict with the optimal threshold for holdout set
y_holdout_pred_optimal = (y_holdout_proba >=
optimal_threshold_train).astype(int)

# Calculate performance metrics for holdout set
holdout_accuracy = accuracy_score(y_holdout, y_holdout_pred_optimal)
holdout_precision = precision_score(y_holdout,
y_holdout_pred_optimal)
holdout_recall = recall_score(y_holdout, y_holdout_pred_optimal)
holdout_f1 = f1_score(y_holdout, y_holdout_pred_optimal)
holdout_roc_auc = roc_auc_score(y_holdout, y_holdout_proba)

print("Holdout Set Performance Metrics with Optimal Threshold:")
print(f"Accuracy: {holdout_accuracy}")
print(f"Precision: {holdout_precision}")
print(f"Recall: {holdout_recall}")
print(f"F1 Score: {holdout_f1}")
print(f"AUC: {holdout_roc_auc}")

# Save the model to a file
```

```
model_filename =
'/content/drive/MyDrive/data/best_logistic_regression_model_holdout.
joblib'
joblib.dump(best_model, model_filename)
```

#Best test-train model

```
#Divide dataset into test and train
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.3, stratify=y, random_state=42)

# Check the class distribution in the train and test sets
print("Class distribution in training set:", np.bincount(y_train))
print("Class distribution in test set:", np.bincount(y_test))
# Define the pipeline
pipeline = Pipeline([
    ('scaler', StandardScaler()),
    ('PCA', PCA(random_state=42)),
    ('feature_selection',
SelectFromModel(RandomForestClassifier(random_state=42))),
    ('classifier', LogisticRegression(random_state=42))
])

# Define parameter distributions for RandomizedSearchCV
param_dist = {
    'PCA__n_components': uniform(0.8, 0.199),
    'feature_selection__estimator__n_estimators': [50, 100, 200],
    'classifier__C': uniform(0.01, 10)
}
# Perform RandomizedSearchCV with repeated stratified k-fold
cross-validation
rkf = RepeatedStratifiedKFold(n_splits=5, n_repeats=3,
random_state=42)
random_search = RandomizedSearchCV(pipeline,
param_distributions=param_dist, n_iter=50, cv=rkf, scoring='f1',
random_state=42, n_jobs=-1, error_score='raise')
random_search.fit(X_train, y_train)

# Best parameters and best score
best_params = random_search.best_params_
best_score = random_search.best_score_

print(f"Best parameters found: {best_params}")
print(f"Best cross-validation score: {best_score}")
```

```
# Access the best estimator
best_model = random_search.best_estimator_

# Train the best model on the entire training set
best_model.fit(X_train, y_train)

# Predict probabilities on the training set
y_train_proba = best_model.predict_proba(X_train)[:, 1]

# Calculate precision-recall curve for training set
precision_train, recall_train, thresholds_train =
precision_recall_curve(y_train, y_train_proba)
optimal_idx_train = np.argmax(precision_train[recall_train >= 0.6])
optimal_threshold_train = thresholds_train[optimal_idx_train]

# Predict with the optimal threshold for training set
y_train_pred_optimal = (y_train_proba >=
optimal_threshold_train).astype(int)

# Calculate performance metrics for training set
train_accuracy = accuracy_score(y_train, y_train_pred_optimal)
train_precision = precision_score(y_train, y_train_pred_optimal)
train_recall = recall_score(y_train, y_train_pred_optimal)
train_f1 = f1_score(y_train, y_train_pred_optimal)
train_roc_auc = roc_auc_score(y_train, y_train_proba)

print("Training Set Performance Metrics with Optimal Threshold:")
print(f"Accuracy: {train_accuracy}")
print(f"Precision: {train_precision}")
print(f"Recall: {train_recall}")
print(f"F1 Score: {train_f1}")
print(f"AUC: {train_roc_auc}")

# Predict probabilities on the test set
y_test_proba = best_model.predict_proba(X_test)[:, 1]

# Calculate precision-recall curve for test set
precision_test, recall_test, thresholds_test =
precision_recall_curve(y_test, y_test_proba)
optimal_idx_test = np.argmax(precision_test[recall_test >= 0.6])
optimal_threshold_test = thresholds_test[optimal_idx_test]

# Predict with the optimal threshold for test set
```

```
y_test_pred_optimal = (y_test_proba >=
optimal_threshold_test).astype(int)

# Calculate performance metrics for test set
test_accuracy = accuracy_score(y_test, y_test_pred_optimal)
test_precision = precision_score(y_test, y_test_pred_optimal)
test_recall = recall_score(y_test, y_test_pred_optimal)
test_f1 = f1_score(y_test, y_test_pred_optimal)
test_roc_auc = roc_auc_score(y_test, y_test_proba)

print("Test Set Performance Metrics with Optimal Threshold:")
print(f"Accuracy: {test_accuracy}")
print(f"Precision: {test_precision}")
print(f"Recall: {test_recall}")
print(f"F1 Score: {test_f1}")
print(f"AUC: {test_roc_auc}")

# Save the model to a file
model_filename =
'/content/drive/MyDrive/data/best_logistic_regression_model_test_tra
in.joblib'
joblib.dump(best_model, model_filename)
```