UNIVERSITÀ DEGLI STUDI DI MILANO

UNIVERSITÀ DEGLI STUDI DI MILANO

CORSO DI DOTTORATO INDUSTRIALE

IN INTERSECTORAL INNOVATION – DOTTORATO INTERSETTORIALE PER L'INNOVAZIONE

XXXVIII CICLO

SCIENZE ESATTE

DIPARTIMENTO DI AFFERENZA DEL DOTTORANDO

DIPARTIMENTO DI INFORMATICA

TESI DI DOTTORATO DI RICERCA

# A Multi-Stage Machine Learning Framework for Prostate Cancer Diagnosis Based on mpMRI Imaging

SETTORE SCIENTIFICO DISCIPLINARE INF01A

Saman Fouladi
Matr. R14090
ORCID: 0000-0003-3367-4952

**TUTOR**
Prof. Gabriele Gianini

**CO-TUTOR**
Dr. Marco Alì

**COORDINATORE DEL CORSO DI DOTTORATO**
Prof. Ernesto Damiani

A.A. 2025–2026

# Summary

**Abstract**

Prostate cancer is one of the most common malignancies among men worldwide. Early and accurate diagnosis of the pathology plays a critical role in improving treatment outcomes. This thesis presents a comprehensive Machine Learning and Deep Learning ((ML)/(DL)) framework for prostate cancer diagnosis: it integrates zonal prostate segmentation, lesion segmentation, Prostate Imaging Reporting and Data System (PI-RADS) classification, and cancer detection using multiparametric Magnetic Resonance Imaging (mp-MRI). The analysis relies on data from T2-weighted imaging (T2W), Apparent Diffusion Coefficient (ADC) imaging, and Diffusion-weighted Imaging (DWI), along with morphologic and biomarker-related clinical information. The proposed multi-stage workflow aims to support radiologists, reduce diagnostic variability, and possibly serve as an educational tool for medical trainees.

The studied Deep Learning models were based on U-Net architectures. For *peripheral zone* (PZ) and *central gland* (CG) segmentation, we relied on the Prostate158 dataset of 3T mp-MRIs: we trained on 90 scans, validated on 25, and tested on 24. We considered Attention-Res-UNet, Vanilla-Net, and V-Net individually and also as an ensemble. Meta-Net and YOLO-V8 were also evaluated. We achieved high performance in providing anatomically precise delineations: YOLO-V8 achieved the best results, with a DSC of 89% for the central gland and 73% for peripheral zone segmentation.

For *lesion segmentation*, we developed a dedicated dataset of 311 mp-MRI cases collected at the Centro Diagnostico Italiano (CDI), Milan, including 58 PI-RADS

3 and 253 PI-RADS 4–5 cases. The data set comprised T2W, ADC, and DWI sequences with manually annotated masks and was preprocessed by normalization and registration. Four deep learning architectures, U-Net, Dense U-Net, Attention U-Net, and LSTM U-Net, were evaluated using both single-modality and multi-input strategies. The Dense U-Net achieved the best performance, with a DSC of 69% on ADC images for PI-RADS 4–5 and 68% for PI-RADS 3–5.

PI-RADS is a standardized scoring system used by radiologists to categorize prostate lesions based on their likelihood of clinically significant cancer. The *PI-RADS machine learning classifier* improved diagnostic consistency by extracting discriminative imaging features. Three approaches were evaluated for automated PI-RADS 3–5 classification using T2W, DWI, and ADC sequences: (1) handcrafted radiomic features from manually segmented lesions, (2) a fully automated lesion and zonal segmentation pipeline, and (3) a custom convolution neural network learning high-level features from ADC images and masks. ADC-derived features performed best, with an ensemble model achieving 77% accuracy Accuracy (Acc), 83% AUC, and 0.618 Matthews Correlation Coefficient Matthews Correlation Coefficient (MCC), with PI-RADS 5 most reliably classified (AUC  94%), while PI-RADS 3 remained the most challenging.

The *cancer detection stage* integrated imaging and clinical parameters, such as age, prostate-specific antigen (Prostate-Specific Antigen (PSA)) levels, and biopsy results, to enhance malignancy prediction. In total, 345 patients were included in the study, with data collected from Trita Hospital in Tehran, Iran. Using ADC images, the proposed model achieved the highest performance, with an accuracy of 83%, AUC of 87%, and a Matthews Correlation Coefficient of 0.638.

A key contribution of this research is the creation of an end-to-end clinically inspired workflow that connects all diagnostic stages and includes, for practical deployment, a user-friendly 3D Slicer plugin that we contributed to developing.

Although the size of the data set and the lack of multicenter validation still represent limitations of the work, the results highlight the potential of multimodal AI-driven approaches to improve diagnostic accuracy, standardize reporting, and help personalized patient care.

This work establishes a foundation for future developments in AI-assisted prostate cancer diagnosis and clinical decision support systems.

**Keywords**

Prostate Cancer; MultiParametric Magnetic Resonance Imaging (mp-MRI); Prostate Lesions; PI-RADS; Prostate Zones; Deep Neural Networks; U-NETs.

# Acknowledgments

I would like to thank all those who supported me and contributed to the successful completion of this thesis.

Firstly, I would like to sincerely thank my supervisor, Professor Gabriele Gianini, for his invaluable support, guidance, and insight into the vast field of artificial intelligence and machine learning. Beyond his expertise, his approach as a professor, with patience, encouragement, and thoughtful mentorship, taught me lessons that went far beyond the academic realm.

Thanks to the professors at the Department of Computer Science 'Giovanni degli Antoni' of the University of Milan, especially Prof. Ernesto Damiani, for his guidance from the start, and to the International Research & Innovation Center on Digital Intelligent Systems (IRIXYS) for providing an enriching experience and broadening my perspective.

I would also like to extend my heartfelt thanks to Dr. Marco Alì, my co-supervisor at CDI (Centro Diagnostico Italiano), for his generous support, providing all the necessary resources and guidance for my research, and for his trust and encouragement, which greatly motivated me to achieve my goals.

A special thanks to all the colleagues, doctors, and radiologists at CDI, whose support, collaboration, and shared insights made this interdisciplinary research possible.

A profound thank you to my family and friends, whose unwavering support throughout my life has guided me and encouraged me to overcome challenges, even from afar. Their love, encouragement, and belief in me have been a constant source of strength on this journey.

Finally, I wish to thank all the participants of this study. Their time, cooperation, and contribution to scientific progress made this research possible.

# Acronyms

| | |
|---|---|
| Acc | Accuracy. |
| ADC | Apparent Diffusion Coefficient. |
| AFMS | Anterior Fibro Muscular Stroma. |
| AI | Artificial Intelligence. |
| ANN | Artificial Neural Network. |
| AUC | Area Under the Curve. |
| | |
| bp-MRI | biparametric magnetic resonance imaging. |
| | |
| CAD | Computer-Aided Diagnostic. |
| CG | Central Gland. |
| CiS | Clinically Insignificant. |
| CM | Confusion Matrix. |
| CNNs | Convolution Neural Networks. |
| CS | Clinically Significant. |
| CsPCa | clinically significant prostate cancers. |
| CZ | Central Zone. |
| | |
| DCE | Dynamic Contrast Enhanced. |
| DL | Deep Learning. |
| DRE | Digital Rectal Examination. |
| DSC | Dice Similarity Coefficients. |
| DTL | Deep Transfer Learning. |

| | |
|---|---|
| DWI | Diffusion-weighted Imaging. |
| GLCM | Gray Level Co-occurrence Matrix. |
| GLDM | Gray Level Dependence Matrix. |
| GLRLM | Gray Level Run Length Matrix. |
| GLSZM | Gray Level Size Zone Matrix. |
| GS | Gleason Score. |
| IoU | Intersection over Union. |
| KNN | k-Nearest Neighbor. |
| LASSO | Least Absolute Shrinkage and Selection Operator. |
| LGBM | Light Gradient Boosting Machine. |
| MCC | Matthews Correlation Coefficient. |
| ML | Machine Learning. |
| MLP | Multilayer Perceptron. |
| mp-MRI | multi-parametric Magnetic Resonance Imaging. |
| NGTDM | Neighboring Gray Tone Difference Matrix. |
| NN | Neural Network. |
| PCa | Prostate Cancer. |
| PI-RADS | Prostate Imaging Reporting and Data System. |
| PoC | Point of Care. |
| PSA | Prostate-Specific Antigen. |
| PSAD | Prostate-Specific Antigen Density. |
| PZ | Peripheral Zone. |

| | |
|---|---|
| ROI | Region of Interest. |
| SHAP | SHapley Additive exPlanations. |
| SVM | Support Vector Machine. |
| T2W | T2-weighted imaging. |
| TIJ | Theory of Independent Judges. |
| TRUS | Transrectal Ultrasound. |
| TTPM | Transperineal Template Prostate-Mapping. |
| TZ | Transition Zone. |
| WG | Whole Gland. |
| YOLO | You Only Look Once. |

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 Context

Prostate cancer (Prostate Cancer (PCa)) stands as the most prevalent malignancy among men worldwide, with an estimated 1.4 million new cases and 375,000 deaths in 2020 [2]. The incidence rates exhibit significant geographical variation, with higher rates observed in Northern and Western Europe, North America, and the Caribbean, and lower rates in Asia and North Africa [3]. In the United States, the age-adjusted incidence rate is 120.2 per 100,000 men, while the mortality rate is 19.2 per 100,000 men [4].

The disease predominantly affects older men, with more than 80% of cases diagnosed in individuals over 65 years of age [5]. Risk factors for prostate cancer include age, race, family history, and certain genetic mutations. Notably, Black men have a higher incidence and mortality rate compared to other ethnic groups, with a two-fold increased risk of diagnosis and a 2.5-fold increased risk of death [6].

The clinical management of prostate cancer has evolved with advancements in imaging technologies, particularly mp-MRI, which combines T2W, DWI, and dynamic contrast-enhanced imaging. This imaging modality has significantly improved the detection and localization of prostate lesions, helping to differentiate between clinically significant Clinically Significant (CS) and indolent tumors [7]

[8].

**Organ and zonal segmentation.** A critical component of mpMRI analysis is the accurate segmentation of the prostate into its anatomical zones, including the peripheral zone, transition zone Transition Zone (TZ), and central zone Central Zone (CZ). Among these, the peripheral zone is the most common site for clinically significant prostate cancer clinically significant prostate cancers (CsPCa). However, manual delineation of these zones is time-consuming and subject to inter-observer variability. Recent studies have demonstrated the potential of DL algorithms to automate prostate zonal segmentation with high accuracy, offering a promising solution to this challenge [9].

**Lesion segmentation.** After zonal segmentation, the next key step is the identification and outlining of lesions within the prostate. Lesion segmentation enables precise assessment of tumor size, location, and spatial relationship to surrounding anatomical structures, which is critical for diagnosis, risk stratification, and treatment planning. Traditionally, this task relies on manual contouring by radiologists, which is both time-consuming and prone to inter-observer variability. Recent advances in DL have demonstrated strong potential for automating lesion segmentation, thereby reducing radiologists' workload, improving consistency, and minimizing the risk of human error [10].

**Prostate Imaging Reporting and Data System (PI-RADS) prediction.** The PI-RADS was developed to standardize the acquisition, interpretation, and reporting of prostate mpMRI [10]. The most recent version, PI-RADS v2.1, assigns each lesion a score from 1 to 5, where higher scores indicate a greater likelihood of clinically significant prostate cancer [11]. This standardized framework improves communication among radiologists and urologists, guides decisions regarding biopsy and treatment, and reduces variability in reporting. Recent research suggests that integrating AI–based lesion detection with PI-RADS scoring could further improve diagnostic accuracy and reproducibility, providing valuable support in clinical decision-making [12].

Moreover, clinical data play a crucial role in the accurate assessment and management of prostate cancer. Factors such as prostate-specific antigen levels, prostate-specific antigen Density (PSAD), and patient history are vital in guiding biopsy decisions and interpreting imaging findings. Incorporating these clinical parameters alongside PI-RADS scoring can enhance risk stratification, reduce unnecessary biopsies, and improve the detection of clinically significant prostate cancer [13].

**Prostate Cancer Prediction.** However, it's important to note that not all lesions with high PI-RADS scores are malignant. For instance, lesions categorized as PI-RADS 4 or 5 have a high probability of being prostate cancer, but false positives can occur, leading to negative biopsy results [14]. Therefore, biopsy remains essential for definitive diagnosis, even in the presence of high PI-RADS scores.

## 1.2 Formulation of the Problem

This thesis addresses the problem of improving prostate cancer detection and characterization using mp-MRI through four main steps: (1) automated prostate zonal segmentation, (2) lesion segmentation, (3) PI-RADS classification, and (4) real cancer detection by integrating imaging and clinical data. These tasks represent essential components of the diagnostic pathway, yet remain challenging due to the complexity of prostate anatomy, the heterogeneity of tumors, and limitations of current imaging interpretation practices. To address this challenge, we utilized real-world patient data collected from a diagnostic center. We applied state-of-the-art AI models tailored to each stage of the pipeline. This approach allows us to evaluate the feasibility and performance of advanced AI methods in a clinically realistic setting.

Despite the adoption of PI-RADS as a standardized framework, mpMRI interpretation is still highly dependent on radiologist expertise and suffers from inter-

and intra-observer variability. Manual segmentation of zones and lesions is labor-intensive and inconsistent, while lesion classification is limited by the fact that not all lesions with high PI-RADS scores are malignant. For this reason, a biopsy is still required to confirm malignancy, but it is invasive, costly, and carries risks such as infection or bleeding. Even then, biopsies may under-sample the prostate and miss clinically significant tumors.

Furthermore, imaging alone does not provide a complete picture of disease risk. Clinical parameters such as prostate-specific antigen, PSAD, and patient history are indispensable for accurate diagnosis and treatment planning. The lack of integration between imaging findings and clinical information reduces diagnostic accuracy and may lead to unnecessary biopsies or the missed detection of cancers.

The core problem, therefore, is the need for automated, accurate, and reproducible methods that combine advanced image analysis with clinical data to support the reliable detection and characterization of prostate cancer. This thesis formulates and addresses this problem by developing and evaluating DL–based approaches across the four key diagnostic steps, aiming to reduce diagnostic variability, assist clinical decision-making, and improve patient outcomes.

## 1.3 Research Contributions

This thesis makes several key contributions to the field of prostate cancer diagnosis and medical imaging, addressing both methodological and clinical challenges through the development of an integrated artificial intelligence framework.

- **Development of a comprehensive AI pipeline:** A multi-stage framework was designed to perform prostate zonal segmentation, lesion segmentation, PI-RADS classification, and real cancer detection. This integrated approach reflects the actual clinical workflow by linking the outputs of one stage to the next and is more complete than most previous studies that focus on only one task. By supporting multiple diagnostic stages, the framework has the potential to improve diagnostic consistency and reduce inter-observer variability.

- **Use of real-world, clinically validated data:** The models were trained and evaluated on mpMRI data collected from diagnostic centers, with manual annotations supervised by two expert radiologists. This ensures that the models are clinically relevant and that the results reflect real diagnostic challenges encountered in routine practice.

- **Multimodal image analysis:** Unlike many studies that rely on one or two MRI sequences, this thesis incorporates all three major mpMRI sequences (T2-weighted (T2W), apparent diffusion coefficient (ADC), and diffusion-weighted imaging (DWI)). The integration of complementary imaging information enables more comprehensive tissue characterization and improves segmentation, classification, and detection performance.

- **Integration of imaging and clinical data:** For real cancer detection, selected clinical features were incorporated alongside imaging data. This multimodal fusion highlights the importance of combining clinical and imaging information to achieve more accurate and reliable predictions and supports personalized risk assessment.

- **Advancement in AI model application:** State-of-the-art deep learning models were applied and customized for each stage of the pipeline. The results demonstrate the potential of these models to assist radiologists by automating time-consuming tasks and improving diagnostic efficiency and consistency.

- **Educational and practical tool development:** A user-friendly 3D Slicer plugin was developed to share the zonal segmentation model with the wider research and clinical community. The plugin enables easy application of the trained model to new patient data, supporting both clinical assistance and the training of medical residents.

- **Potential for future clinical impact:** Despite being trained on a relatively small dataset, the proposed framework demonstrated promising performance. With access to larger multicenter datasets, the pipeline could

achieve higher accuracy and broader clinical applicability, supporting improved clinical decision-making and patient outcomes.

## 1.4 Thesis Outline

To provide an overview of the thesis contributions, Chapter 2 presents the relevant terms, definitions, and key concepts in prostate cancer and AI, providing the foundational knowledge necessary to understand the subsequent methodology and analysis. Chapter 3 reviews the literature related to the four main tasks addressed in this study: prostate zonal segmentation, lesion segmentation, PI-RADS classification, and real prostate cancer detection. Existing approaches, their limitations, and gaps in the literature are discussed, highlighting the motivation for the proposed multi-stage AI framework.

Chapter 4 details the materials and methods used in the study, including dataset description, preprocessing steps, neural network Neural Network (NN) architectures, and implementation details for each of the four stages. Each stage is addressed separately, with clear explanations of how the outputs of one step are utilized in the next, reflecting a continuous and clinically relevant workflow. Chapter 5 presents the results of the experiments for each stage, including quantitative evaluations and visualizations, to assess the performance and effectiveness of the proposed models.

Chapter 6 provides an in-depth discussion of the results, comparing them with previous studies, highlighting their significance, and illustrating findings using charts and tables for clarity. The chapter concludes with a summary of the main findings, the contributions of the thesis, its limitations, and recommendations for future research, emphasizing the potential impact of the proposed AI framework on clinical practice and medical education.

## 1.5 Research Outputs

Published conference manuscripts:

- Publication and presentation of the conference paper as the first author in the 16th International Conference On Management of Digital Ecosystems (MEDES 2024), Italy, "Advanced Prostate MRI Analysis: UNET-Based Models for Zonal and Lesion Segmentation", Saman Fouladi, Gabriele Gianini, Deborah Fazzini, Alessandro Maiocchi, Ernesto Damiani, Sergio Papa, Marco Ali.Communications in Computer and Information Science, vol 2518. Springer, Cham. https://doi.org/10.1007/978-3-031-93598-$5_1$3

- Publication and presentation of the conference paper as the last author in the 4th Italian Conference On Big Data and Data Science (ITADATA), "A Federated Learning Architecture for Prostate MRI Image Segmentation", A.Bovio, Mario Barile, Francesco Pallotta, Lorenzo Pede, Alessandro.Maiocchi, Marco Ali, Fatemeh Darvizeh, Deborah Fazzini, Francesco Lacavalla, Massimo Banzi, Gabriele Gianini, Filippo Berto, Ruslan Bondaruc, Ernesto Damiani, S.Fouladi,2025

Published journal manuscripts:

- Publication of the journal paper as first author: "Neural Network Models for Prostate Zones Segmentation in Magnetic Resonance Imaging. Information.", Saman Fouladi, Luca di Palma, Fatemeh Darvizeh, Deborah Fazzini, Alessandro Maiocchi, Alessandro Maiocchi, Gabriele Gianini, Marco Ali. Information 2025, 16(3), 186; https://doi.org/10.3390/info16030186

Already accepted, to be published soon in conference proceedings:

- Publication and presentation of the conference paper as the first author in the 17th International Conference On Management of Digital Ecosystems (MEDES 2025), Vietnam, "A hybrid LSTM-UNet architecture for segmentation of MRI prostatic lesion images", Saman Fouladi, Fatemeh Darvizeh, Rosario Di Meo, Isa Bossi Zanetti, Gabriele Gianini, Corrado Mio, Ernesto Damiani, Eros Cambie, Antonino Licata, Alessandro Maiocchi, Marco Ali, and Deborah Fazzini, 2025.

Under review (Journal manuscript)

- Publication of the journal paper as first author: "Automated Prostate Lesion Segmentation in MRI: Exploring UNet-Based Models Across Multi-Sequence Imaging (T2W, ADC, and DWI)", Saman Fouladi, Fatemeh Darvizeh, Gabriele Gianini, Rosario Di Meo, Luca Di Palma, Ernesto Damiani, Alessandro Maiocchi, Deborah Fazzini, Marco Alì, 2025.

- Publication of the journal paper as first author: "Automated PI-RADS 3–5 Classification Using Multiparametric MRI: A Comparative Study of Radiomics and DL Approaches", Saman Fouladi, Isa Bossi Zanetti, Fatemeh Darvize, Rosario Di Me, Luca Di Palm, Eros Cambie, Antonino Licata, Alessandro Maiocchi, Ernesto Damiani, Marco Ali, Gabriele Gianini and Deborah Fazzini, 2025.

In preparation (Journal manuscript)

- "Integrating Multiparametric MRI and Clinical Data for Prostate Cancer Detection Using Machine Learning Classifiers."

## 1.6 Multilayered Urban Sustainability Action (MUSA) Project

This thesis is part of the MUSA project. MUSA is an Innovation Ecosystem funded by the Italian Ministry of University and Research under the National Recovery and Resilience Plan (PNRR), specifically Mission 4, Component 2, Investment 1.5. Coordinated by the University of Milano-Bicocca, MUSA unites leading institutions including Politecnico di Milano, Bocconi University, and Università degli Studi di Milano, alongside numerous public and private partners.

Established in Milan, MUSA addresses the complex challenges of urban sustainability by fostering a new model of public-private collaboration. The initiative focuses on three interconnected domains: environmental, economic, and social

sustainability. It leverages the Lombardy region as a dynamic testing ground for integrated innovations and interdisciplinary responses across six thematic areas, known as "spokes":

1) Urban Regeneration – City of Tomorrow: Focuses on sustainable urban development and revitalization.

2) Big Data and Open Data in Life Sciences: Utilizes data-driven approaches to improve public health.

3) Deep Tech: Entrepreneurship and Technology Transfer: Promotes innovation and the commercialization of advanced technologies.

4) Economic Impact and Sustainable Finance: Develops models for sustainable economic growth and financial systems.

5) Sustainable Fashion, Luxury, and Design: Integrates sustainability into the fashion and design industries.

6) Innovation for Sustainable and Inclusive Societies: Ensures social inclusion and community engagement in sustainability efforts.

MUSA aims to create scalable solutions that can be replicated nationally and internationally, transforming Milan into a model for urban regeneration and sustainability.

Within this framework, the present thesis is situated in Spoke 2: Healthcare of the Future, which focuses on advancing precision medicine, digital health technologies, and AI-driven diagnostic tools to improve patient care and healthcare sustainability. Specifically, this work contributes to Work Package 1 (WP1): Artificial Intelligence for Precision Medicine, under Pilot 1.2, which addresses the development of AI-based methods for medical image analysis and decision support systems.

# Chapter 2

# Terms and Definitions

## 2.1 Prostate Cancer

Prostate cancer is the second most commonly diagnosed cancer among men [15] and the fifth leading cause of cancer-related deaths worldwide [16]. In 2018, there were approximately 1.3 million new cases, and this number is projected to rise to nearly 1.7 million new cases annually by 2030, largely due to global population aging and increased life expectancy [17]. Once prostate cancer has metastasized beyond the prostate gland, treatments like chemotherapy and immunotherapy are no longer curative [18]. Therefore, early detection is critical, as it allows for intervention while the tumor is still localized and potentially more treatable [19]. Despite the importance of timely and accurate diagnosis, current prostate cancer screening methods remain controversial, with ongoing debate about their effectiveness and lack of conclusive evidence demonstrating clear clinical benefits [18] [20] [21].

### 2.1.1 Prostate anatomy and function

The prostate is a secondary gland in the male reproductive system [22]. It sits just below the bladder neck, above the external urethral sphincter, and next to the rectum. Its main role is to produce proteolytic enzymes that help break down clotting agents in semen. During ejaculation, the prostate's muscles also contract

to help propel seminal fluid into the urethra. The structure of the prostate can be described using regional and zonal divisions, as shown in Figure 2.1. These, along with standard anatomical directions, help accurately pinpoint the location of prostate-related conditions. The regional division splits the prostate into three parts: the base (the upper part near the bladder), the apex (the lower part near the external urethral sphincter), and the midgland (the central portion in between). The zonal division, on the other hand, identifies five main zones: the peripheral zone, the anterior fibromuscular stroma Anterior Fibro Muscular Stroma (AFMS), the central zone, the transition zone, and the capsule, a thin outer layer of connective tissue and muscle. The peripheral zone is located at the back of the prostate, whereas the anterior fibromuscular stroma is found at the front. The transition zone surrounds the prostatic urethra, and the central zone lies behind the transition zone, encircling the ejaculatory ducts. Anterior fibromuscular Stroma, central zone, and transition zone are often collectively referred to as the central gland.

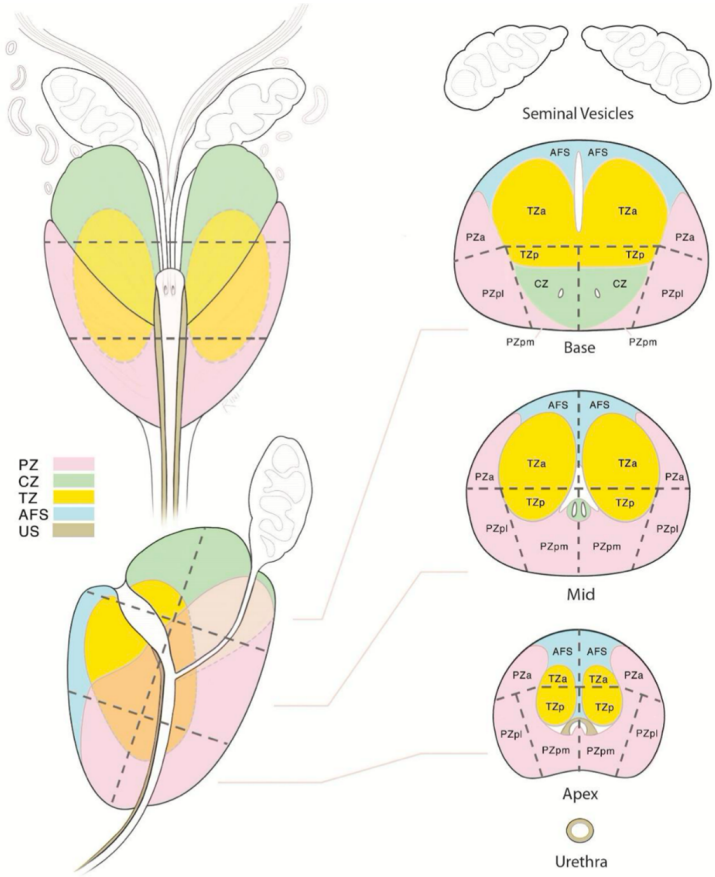



Figure 2.1: *Sectoral Map of the Prostate According to PI-RADS v2. Peripheral zone; central zone; transition zone; US: Urethral Stroma; AFS: Anterior Fascial Stroma. Adapted from the American College of Radiology, MR PI-RADS version 2.0. Image obtained from www.acr.org/Quality-Safety/Resources/PIRADS/*

### 2.1.2 Prostate Cancer Prevalence in Prostate Zones

Different regions of the prostate exhibit different histological characteristics. The transition zone contains well-differentiated glands with epithelial cells of varying sizes, arranged in a high columnar pattern with clear cytoplasm. The transition zone has a denser stroma and lower nerve density compared to the peripheral zone. In contrast, the peripheral zone is marked by regularly sized epithelial cells and sparse stroma, arranged in a corrugated pattern with a higher nerve density than

the transition zone and the central zone. The central zone is histologically distinguished by large, irregular glandular structures, with larger nuclei and less defined cell membranes [23] [24] [25] [26]. In addition, different cell types are predominant in the peripheral zone and transition zone. In the transition zone, epithelial cells are mainly composed of club cells, hillock cells, and PSCA-positive epithelial cells, while fibroblasts are the primary stromal cell type. In contrast, the peripheral zone is mainly populated by luminal epithelial cells [27]. Numerous pathomorphological studies have confirmed that prostate cancer and prostatitis are more frequently found in the peripheral zone, while benign prostatic hyperplasia (BPH) develops almost exclusively in the transition zone. McNeal [28] noted that BPH does not occur in the peripheral zone. The central zone is rarely affected by carcinoma or hyperplasia, with very few cases of tumors originating in the central zone reported worldwide. These significant differences in disease distribution across prostate regions have increasingly drawn the attention of researchers, although the underlying pathological mechanisms remain unclear [27].

**Digital Rectal Examination (DRE):**

Historically, prostate cancer diagnosis has relied on DRE, a procedure in which a physician inserts a gloved finger into the rectum to palpate the prostate for any hard, irregular, or abnormal regions (Figure 2.2). However, the accuracy of DRE is highly dependent on the clinician's expertise and is significantly limited in its ability to detect tumors, as certain regions of the prostate are not accessible through this method [29].

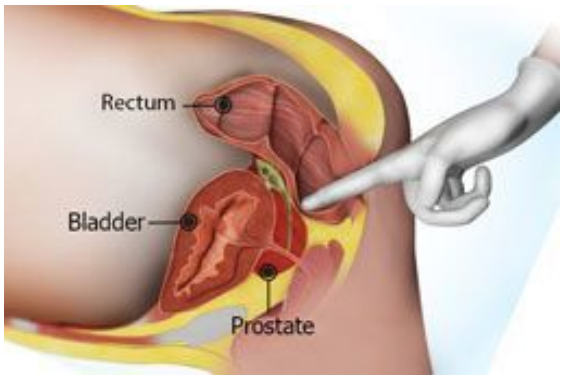



Figure 2.2: *Digital rectum examination process.*
*Image obtained from www.analfissuretreatment.co.uk*

**Prostate-Specific Antigen (PSA) Test:** Prostate-specific antigen testing has been a cornerstone in prostate cancer detection since its introduction in the early 1990s. While it has significantly improved early diagnosis and patient outcomes, prostate-specific antigen testing is not without limitations. Prostate-specific antigen is produced by both benign and malignant prostate cells, leading to potential false positives and negatives. This has raised concerns about the overdiagnosis and overtreatment of indolent cancers that may not pose significant health risks [30]. To address these challenges, recent advancements aim to refine prostate-specific antigen screening. The PROBASE trial, involving over 12,500 men aged 45-50, suggests that low-risk individuals with prostate-specific antigen levels under 1.5 ng/ml could safely extend screening intervals to five years, reducing unnecessary testing without compromising cancer detection [31]. Additionally, point-of-care (PoC) prostate-specific antigen testing is emerging as a promising approach to enhance accessibility and efficiency in prostate cancer diagnostics. These PoC tests aim to provide accurate, low-cost, and rapid results at the time and place of patient care, addressing limitations of traditional laboratory-based assays [32]. Despite these advancements, prostate-specific antigen testing remains controversial due to variability in screening practices and guidelines across regions. Inconsistent recommendations contribute to disparities in early detection and treatment outcomes [33]. In summary, while prostate-specific antigen testing has transformed prostate cancer detection, efforts continue to enhance its accuracy

and applicability, aiming for more personalized and effective screening strategies.

**Transrectal Ultrasound:**

Transrectal ultrasound Transrectal Ultrasound (TRUS) is a diagnostic imaging technique used in prostate cancer detection, primarily for guiding biopsies. It is performed by inserting a small ultrasound probe into the rectum to visualize the prostate gland. Transrectal ultrasound is particularly effective for assessing prostate volume and identifying abnormalities, but it has limitations, especially in detecting small or anterior tumors. It is less accurate in identifying prostate cancer compared to advanced imaging modalities like mp-MRI [34]. Transrectal ultrasound is often combined with mpMRI in a process known as fusion biopsy. This fusion enhances the precision of the biopsy, improving the likelihood of detecting clinically significant prostate cancer by accurately targeting suspicious areas identified on mpMRI [35].

**Prostate biopsy:**

Patients suspected of having prostate cancer are typically recommended for a biopsy, which is considered the final step in current screening protocols. Biopsies are primarily used to confirm the cancer stage and evaluate the tumor's aggressiveness. The procedure involves obtaining tissue samples from the prostate by inserting needles through the patient's rectum [36], to subsequently evaluate and classify the tissue samples using the Gleason Score (GS) system [37]. Essentially, the GS is a grading system that evaluates the architectural pattern of cancer cells in prostate biopsy samples, assigning a score ranging from 3 to 5 based on how abnormal the tissue appears (Figure 2.3).

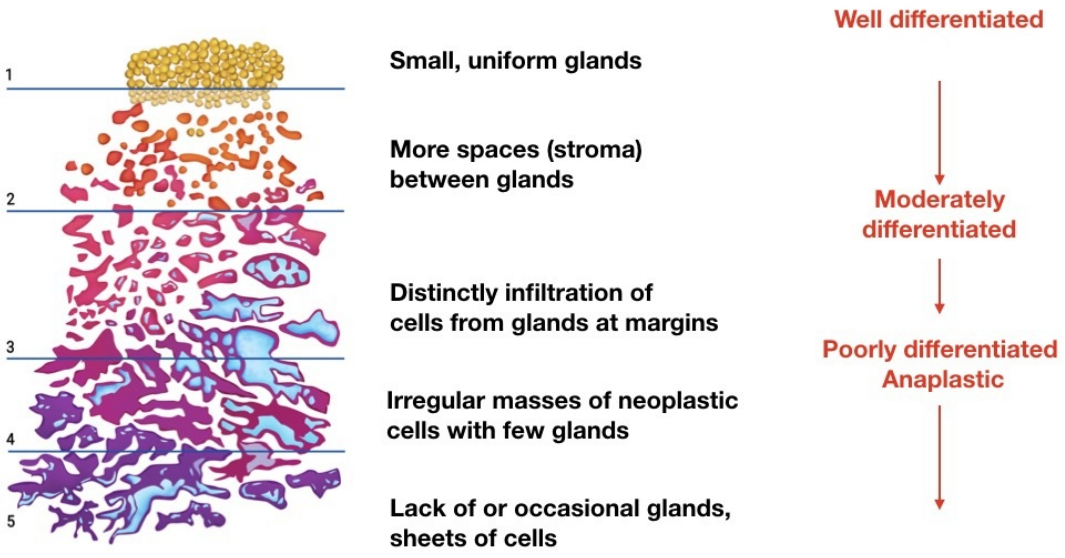



Figure 2.3: *Gleason scoring system for biopsy samples.*
*Image obtained from www.altaklinik.com/prostate/prostate-cancer/gleason-score/*

To calculate the final score, the two most predominant patterns of tumor growth are identified, each given a score, and then combined (e.g., 3+4 or 4+3). This total score is then used to classify the cancer into a specific grade group, which reflects the patient's risk level (Table 2.1).

Table 2.1: *Risk groups, grade groups, and Gleason scores.*

| **Risk Group** | **Grade Group** | **Gleason Score** |
|---|---|---|
| Low | Grade group 1 | Score $\leq 6$ |
| Intermediate (favourable) | Grade group 2 | Score $= 7$ (3+4) |
| Intermediate (unfavourable) | Grade group 3 | Score $= 7$ (4+3) |
| High | Grade group 4 | Score $= 8$ |
| Very high | Grade group 5 | Score $= 9$ or 10 |

There are two primary approaches to performing a prostate biopsy: transrectal ultrasound-guided biopsy and transperineal biopsy. The key distinction between them lies in the needle insertion route, either through the rectum or through the skin between the scrotum and anus [38]. In the first approach, when no advanced

imaging guidance is employed, typically 10 to 12 tissue samples are taken from various regions of the prostate. However, studies have indicated that despite multiple sampling attempts, certain clinically significant tumors may still go undetected [39] [40] [41] [42]. Moreover, biopsies may lead to patient discomfort or pain and can be associated with complications such as infections or rectal bleeding [42]. Overall, the existing diagnostic and screening approaches for prostate cancer highlight the need for alternative methods that can more accurately detect and characterize clinically significant tumors while also minimizing the issue of overdiagnosis that is prevalent in current practices [43] [44]. In this context, recent advancements in imaging technology and interpretation have positioned magnetic resonance imaging (MRI) as a powerful tool in the detection, staging, treatment planning, and management of prostate cancer [45] [46]. Specifically, multi-parametric MRI (mp-MRI) is increasingly being used in everyday clinical practice to manage prostate cancer, showing promising outcomes [47] [48].

## 2.2 Magnetic Resonance Imaging (MRI)

Multi-parametric MRI (mp-MRI) is a non-invasive method that combines different MRI techniques, including T2W, diffusion-weighted (DW), dynamic contrast-enhanced (DCE), and, if needed, magnetic resonance spectroscopy (MRS) [43]. Specifically, T2W images are usually taken in three perpendicular planes, axial, coronal, and sagittal, resulting in three separate sequences for this technique. In addition, ADC maps are typically generated automatically by software during DWI, providing an overall measure of how water molecules move within a specific voxel. While many centers rely on this combination of imaging methods, there is still no universally accepted standard for this approach. Recent studies suggest that bi-parametric MRI (bp-MRI), which includes only T2W and DW sequences (or ADC maps), may offer diagnostic performance comparable to that of full multi-parametric MRI (mp-MRI) for detecting and diagnosing prostate cancer [49]. Figure 2.4 demonstrates all of the sequences of mp-MRI.

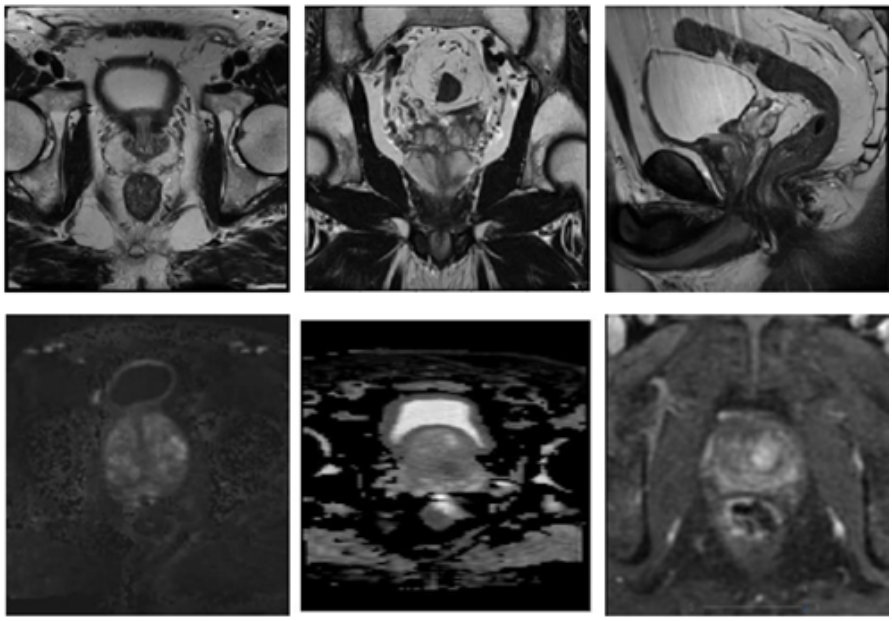


Figure 2.4: *MRI sequences typically included in prostate mp-MRI. Top row: T2W images captured in different planes (from left to right: axial, coronal, and sagittal). Bottom row: DWI with a b-value of 1400, ADC map, and DCE imaging* [49].

**T2-weighted Imaging (T2W):**

T2W provides detailed anatomical and morphological information about the prostate and any pathological structures within it. Capturing images in three perpendicular planes, axial, sagittal, and coronal, highlights the prostate's zonal anatomy and its relationship with surrounding tissues. T2W imaging helps distinguish between the high signal peripheral zone of the prostate, the mixed signal transition zone, and the central low signal zone (Figure 2.5 ). It also plays a key role in precisely locating lesions and evaluating their shape, form, and size, due to its high resolution across different planes [50].

Figure 2.5: *A slice of an T2W image*

**Diffusion-weighted Imaging (DWI):**

In DWI, the image contrast is created by the movement of water molecules within the tissue. This movement, or diffusion, can happen inside cells, in the spaces between cells, or across both. Special gradients, called diffusion-sensitizing gradients, are used to generate images that show how water moves in each voxel. The strength and timing of these gradients are controlled by what's known as the "b-value." A higher b-value makes the scan more sensitive to water movement but also reduces the signal-to-noise ratio (SNR).

By adjusting the level of diffusion weighting, it is possible to get a numerical value that gives insight into the microstructure of the tissue, which is known as the ADC. Calculating the ADC for each voxel creates an ADC map. In healthy prostate tissue, ADC values tend to be relatively high because of the glandular tubules found in the prostate's glandular areas. However, prostate cancer can disrupt these normal structures, reducing water movement. Tumors also tend to have a higher cell density than healthy tissue, further limiting water diffusion. Because of this, unusually low ADC values can be a sign of prostate cancer [51]. Radiologist reporting guidelines suggest that ADC values below the range of 750–900 $mm^2$/sec may indicate the presence of clinically significant prostate cancer [52]. An example of DWI and the corresponding ADC map for a patient with a tumor located in the peripheral zone is presented in Figure 2.6.

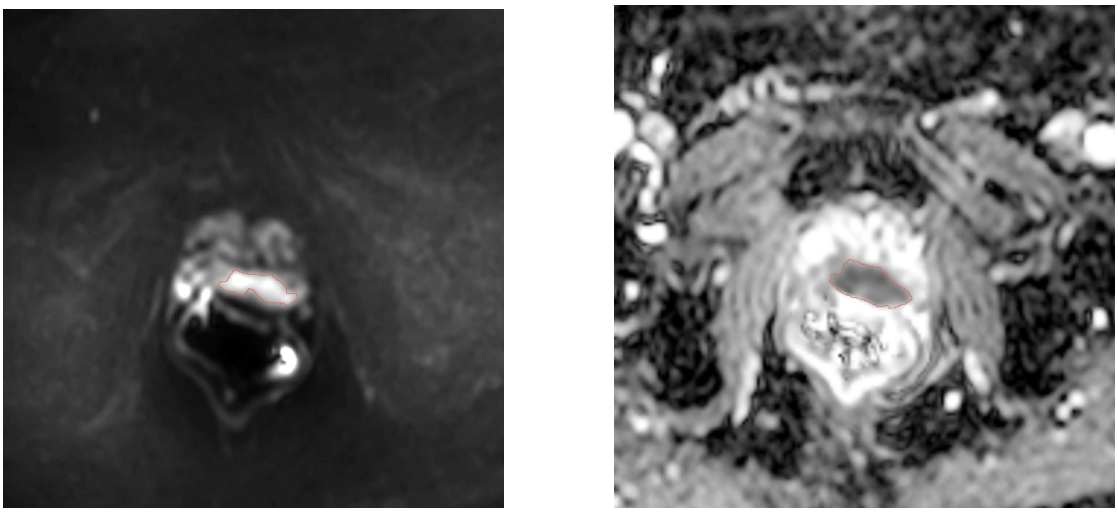


Figure 2.6: Example of a DWI image and its corresponding ADC map, showing a lesion in the peripheral zone highlighted with a red border

Alongside ADC maps, DWI with a high b-value ($b \geq 1400$ [40]) is commonly acquired. In these images, tumors show up as bright spots (hyperintensities) against a darker background because areas with restricted diffusion retain more signals. Recent studies suggest that high b-value DWI is especially useful for detecting key tumors, including those near or involving the anterior fibromuscular stroma, as well as tumors located at the base or apex of the prostate [53].

Although high b-value DWI and ADC maps are important parts of mpMRI, they do have some limitations. Because DWI is typically acquired using fast echo-planar imaging sequences, it tends to have lower spatial resolution compared to T2W imaging. Additionally, these fast sequences are highly sensitive to magnetic field variations, which can lead to distortion or artifacts, especially near areas where tissue meets air, such as near the rectum [51]. A specific challenge with high b-value DWI is the reduced signal-to-noise ratio, which can lead to image distortion and ghosting artifacts [53].

One way to work around the image quality issues of directly acquired high b-value images is to generate them by extrapolating from lower b-value images [52], a technique that has been shown to result in higher signal-to-noise ratios [54]. However, a key limitation of relying solely on DWI for diagnosing prostate cancer

is that certain benign conditions, especially prostatitis in the peripheral zone, can appear similar to cancer on DWI scans [52].

**Dynamic Contrast-Enhanced Imaging (DCE):**

Dynamic contrast-enhanced MRI of the prostate highlights tissue vascularization by capturing the enhancement that occurs after injecting an MR contrast agent. DCE-MRI is particularly useful for detecting local cancer recurrences, such as after radiotherapy or radical prostatectomy. In patients who have not received prior treatment, DCE-MRI (Figure 4, bottom row) can help identify potential cases of prostatitis and is valuable for clarifying uncertain findings in the peripheral zone [50].

**Spectroscopy:** MR spectroscopy is occasionally included in prostate mp-MRI protocols, depending on specific institutional or regional guidelines. It is particularly useful for evaluating the malignancy risk of a region of interest Region of Interest (ROI). However, its use is mostly limited to research settings, as acquiring and analyzing spectroscopy data is a complex and time-consuming process [43].

## 2.3 PI-RADS score

With the introduction of mp-MRI into the diagnostic and management process for prostate cancer, there was a need for a standardized method and terminology to consistently interpret mp-MRI findings in clinical practice. To address this, the Prostate Imaging-Reporting and Data System (PI-RADS) was created in 2013 to standardize the evaluation of prostate MRI [53], with updates released in 2015 (PI-RADS v2) and 2019 (PI-RADS v2.1) [11]. The most recent update introduced a sectoral map of the prostate (Figure 2.7), refined the scoring system to address earlier conceptual inconsistencies, and reduced the role of DCE imaging to a secondary classification function, behind T2W and DW sequences [43] [55]. The PI-RADS v2 scoring system can be summarized as follows (Figure 2.7): Different imaging sequences are obtained, and a score between 1 and 5 is assigned based on various factors such as the homogeneity and encapsulation of the detected lesion. The reference sequence and its contribution to the overall evaluation depend on

whether the lesion is located in the transition zone or the peripheral zone of the prostate.

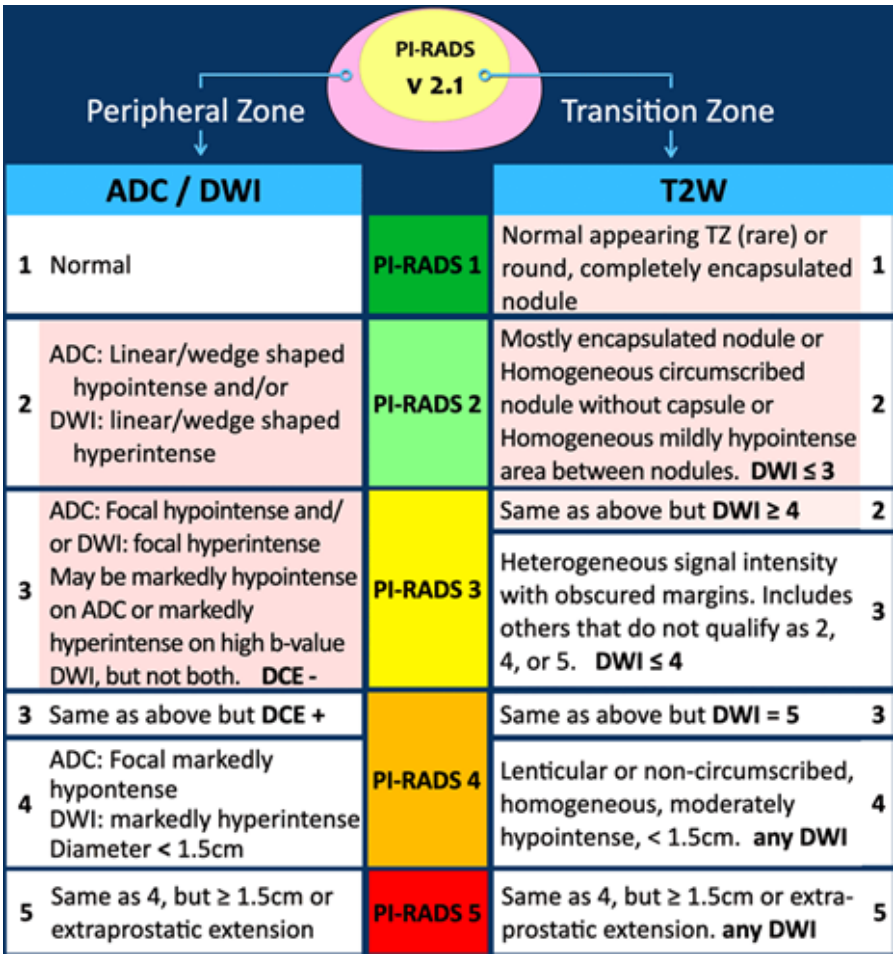



Figure 2.7: *PI-RADS v2 chart for assigning a category and evaluating the grade of the case.*
*Available at: https://radiologyassistant.nl/abdomen/prostate/prostate-cancer-pi-rads-v2-1*

Figure 2.8 illustrates how to determine the PI-RADS score of lesions in different zones, considering factors such as shape, size, and other characteristics.

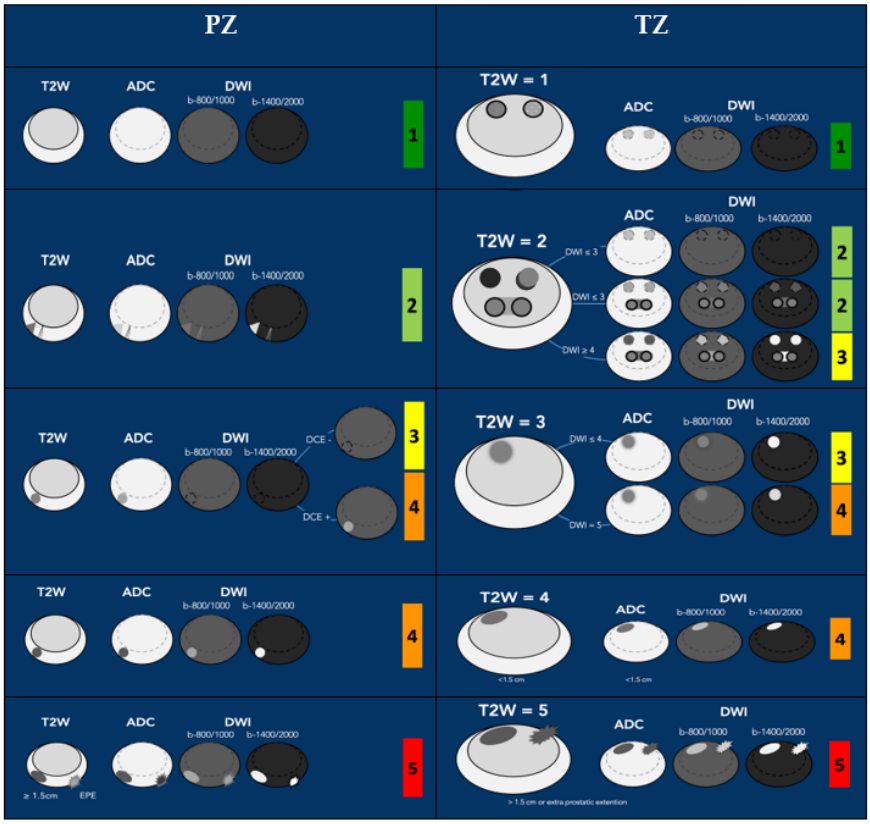



Figure 2.8: *Determining the PI-RADS score of the lesion in the peripheral zone and TZ, based on the lesion's appearance.*
*Available at: https://radiologyassistant.nl/abdomen/prostate/prostate-cancer-pi-rads-v2-1*

## 2.4 Prostate biopsy and Gleason scoring

Biopsy remains the only non-surgical method for confirming a prostate cancer diagnosis. However, due to the potential for sampling error, a negative biopsy does not completely rule out the presence of cancer. The NICE guidelines in the UK recommend MRI-guided biopsy for lesions with a Likert score of 3 or higher, and suggest skipping the biopsy for patients with mpMRI results scored as Likert 1 or 2 [56]. In contrast, PI-RADS v2.1 advises MRI-guided biopsy for lesions scored as PI-RADS 4 or 5, recommends skipping the biopsy for lesions scored PI-RADS 1 or 2, and suggests using additional clinical information to decide whether to biopsy

lesions scored as PI-RADS 3 [52].

MRI-guided biopsies are usually done under local anesthesia. Another method is transperineal template prostate-mapping (TTPM) biopsy, which is typically performed under general anesthesia. This technique involves taking multiple samples from several sites using a grid system, with around 20 different areas sampled and two or three cores taken from each. TTPM biopsy is known for its accuracy, as it eliminates the image bias found in MRI-guided biopsies and reduces the sampling errors seen in random or systematic transrectal ultrasound-guided biopsies. However, TTPM biopsy has a higher risk of side effects because it requires general anesthesia [57]. As a result, it is not used routinely in clinical practice but is often included in clinical trials involving mpMRI to provide a reliable reference standard [58].

A GS is assigned to tissue samples obtained from a biopsy to assess the aggressiveness of prostate cancer through microscopic examination [59]. Gleason grades range from 1 to 5, with grade 5 representing the most aggressive tumor cells. By convention, the two most common patterns seen across all biopsy samples from a tumor are assigned grades. The GS is the sum of these two grades. For example, a GS of 3 + 4 means that most of the cancer cells are grade 3, and a smaller proportion are grade 4, resulting in a total score of 7. Gleason scores typically range from 6 (3 + 3) to 10 (5 + 5) for prostate cancer. Additionally, Epstein et al. [60] introduced a categorization into Gleason Grade Groups with specific histopathological definitions:

- Group 1 (Gleason score $\leq 6$): Only individual, discrete, well-formed glands.
- Group 2 (Gleason score $3 + 4 = 7$): Mainly well-formed glands with a smaller portion of poorly-formed, fused, or cribriform glands.
- Group 3 (Gleason score $4 + 3 = 7$): Mainly poorly-formed, fused, or cribriform glands with a smaller portion of well-formed glands.
- Group 4 (Gleason score $4 + 4 = 8$, $3 + 5 = 8$, $5 + 3 = 8$): Only poorly-formed, fused, or cribriform glands, or predominantly well-formed glands with a lesser portion of glands lacking or predominantly lacking glands, with a lesser amount of well-formed glands.

• Group 5 (Gleason scores 9-10): No gland formation (or with necrosis), with or without poorly-formed, fused, or cribriform glands.

## 2.5 Radiomic Features in Medical Imaging and the PyRadiomics Package

Radiomics is an emerging field that aims to convert standard medical images into high-dimensional quantitative data that can be analyzed to uncover disease characteristics not visible to the human eye. By extracting a large number of features from medical images, radiomics provides a bridge between imaging and personalized medicine, offering the potential to improve diagnosis, prognosis, and evaluation of treatment response [61].

Radiomic features can generally be divided into several categories. First-order statistical features describe the distribution of voxel intensities within the region of interest, such as mean, variance, skewness, and kurtosis. Shape features capture the geometric properties of lesions, including volume, surface area, and sphericity. Texture features quantify spatial relationships between voxels and provide information about heterogeneity within the lesion. These are derived from statistical matrices such as the Gray Level Co-occurrence Matrix (GLCM), Gray Level Run Length Matrix (GLRLM), and Gray Level Size Zone Matrix (GLSZM) [62] [63].

In addition, higher-order features can be obtained by applying image filters (e.g., wavelet, Laplacian of Gaussian) to highlight specific image patterns and enhance discriminatory power [64].

To ensure standardization and reproducibility in radiomic feature extraction, the Image Biomarker Standardization Initiative (IBSI) has proposed detailed guidelines for feature definitions [63]. One of the most widely used open-source tools that adheres to these standards is the PyRadiomics package.

PyRadiomics is a Python-based software developed to provide a standardized and reproducible framework for radiomic feature extraction from medical images [65]. It supports a wide range of imaging modalities, including CT, MRI, and PET, and allows feature computation from both 2D and 3D segmentation. PyRadiomics

extracts hundreds of features across multiple classes, including first-order, shape-based, and texture features, and it offers options for preprocessing steps such as resampling, normalization, and discretization, which are critical for harmonizing data across scanners and protocols. The package also integrates with the 3D Slicer platform, enabling user-friendly workflows for researchers without extensive programming expertise.

The flexibility of PyRadiomics makes it particularly suitable for both research and clinical applications. It can be integrated into machine learning Machine Learning (ML) pipelines for predictive modeling and has been applied in numerous oncological studies, including those focused on prostate cancer, lung cancer, and brain tumors [62] [66]. By providing a robust and reproducible tool for quantitative image analysis, PyRadiomics plays a central role in the advancement of radiomics toward clinical translation.

## 2.6 Machine Learning in Radiology

AI has recently made significant progress in perception, which is the interpretation of sensory information, enabling machines to better understand and process complex data. This progress has resulted in major advancements in areas such as web search, self-driving vehicles, natural language processing, and computer vision, tasks that were previously only possible for humans. DL, a subset of ML based on a neural network structure inspired by the human brain, allows machines to automatically learn discriminative features from data, enabling them to model complex nonlinear relationships. While earlier AI methods often resulted in performance below human capabilities, recent DL algorithms have been able to match or even exceed human performance in specific tasks [67] [68] [69] [70].

This progress is largely due to advances in AI research, the vast amounts of digital data available for training algorithms, and the development of powerful computational hardware. DL techniques have even managed to outperform humans in the strategy game Go, an achievement that was once thought to be decades away due to the game's complexity and vast number of possible moves [71]. Looking

ahead, researchers predict that AI will automate many tasks, including translating languages, writing best-selling books, and performing surgery, all within the next few decades [72]. In healthcare, AI is increasingly becoming an essential part of various applications, such as drug discovery, remote patient monitoring, medical diagnostics and imaging, risk management, wearables, virtual assistants, and hospital management [73].

In radiology, physicians undergo extensive specialized training to analyze medical images and report findings to detect, characterize, and monitor diseases. This assessment often relies on experience, which, despite years of training, can sometimes be subjective. In contrast, AI algorithms can identify complex patterns in imaging data and automate certain tasks or assessments, reducing the level of subjectivity (depending on the quality of the training data). Moreover, when properly implemented and integrated into the clinical workflow, AI can enhance the reproducibility of results, serving as a valuable tool to support physicians [73].

## 2.7 Deep Learning for Medical Imaging

There are mainly two types of AI approaches commonly used in radiology today: traditional ML and DL. The first approach focuses on extracting handcrafted features, which are mathematically defined (like image texture), and can be quantified automatically or semi-automatically by software [74]. This process is typically followed by a feature selection step and an ML algorithm [75]. While these extracted features are considered useful for the tasks at hand, they often depend on expert definitions, which limit their effectiveness. As a result, the features may not always represent the most optimal way to quantify the task. Additionally, these features are usually "static," meaning they are specific to a particular imaging modality and cannot easily adapt or achieve the same level of success with different modalities that have varying signal-to-noise characteristics [73]. DL methods can automatically learn feature representations without the need for human expert intervention. Due to their data-driven nature, DL can extract more general and informative features. It also eliminates manual steps, such as defining a region

of interest, which typically involves experts manually delineating diseased tissues [76]. With enough data, DL is robust to variations like inter-reader variability, which can occur among experts. DL algorithms have achieved remarkable success in various healthcare applications, with a growing number of software products gaining approval from the U.S. Food & Drug Administration (FDA) [77]. Unlike traditional ML, DL can follow a process similar to that of radiologists, identifying parameters and features as it analyzes data, assessing their relevance, and making clinical decisions based on these factors. When comparing DL models to ML-based methods, several studies have shown significant improvements with DL techniques [78] [79]. Additionally, DL typically has a faster development time, relying on curated data instead of domain expertise to extract useful features. In contrast, ML methods have largely plateaued in terms of performance in recent years and often fail to meet the minimum requirements for clinical application, resulting in few systems transitioning to routine clinical use [80].

## 2.8 Applications of Deep Learning for Imaging and Clinical Data Analysis in Prostate Cancer

Reading radiographic images, like MRIs, involves recognizing intricate patterns, a task that computers can be trained to perform efficiently, consistently, and quickly. DL presents an alternative to traditional human-based analysis for various prostate cancer imaging applications. These applications can be divided into two categories: low-level processing methods and high-level image analysis. Low-level processing focuses on basic image tasks, such as pixel classification, segmentation, and registration.

### 2.8.1 Classification

Classification is one of the fundamental tasks in AI and ML, where the goal is to assign data samples into predefined categories based on learned patterns and

features. In this process, algorithms learn from labeled datasets to recognize distinguishing characteristics and apply this knowledge to predict the class of new, unseen data [81]. Such techniques are widely applied across domains, including natural language processing, computer vision, and bioinformatics, where accurate categorization is essential for decision-making.

In the medical domain, classification has emerged as a powerful tool for analyzing complex data, particularly in medical imaging, where it supports the differentiation of healthy and pathological tissues, the grading of disease severity, and the prediction of clinical outcomes [82]. In prostate cancer, classification is especially important for distinguishing between benign and malignant lesions and for identifying clinically significant prostate cancer. With the rise of DL, and particularly convolutional neural networks Convolution Neural Networks (CNNs), classification models have achieved high accuracy in analyzing mp-MRI scans by automatically learning discriminative features without the need for manual feature engineering [83].

Recent studies demonstrate that AI-driven classification can reach diagnostic performance levels comparable to expert radiologists, with pooled sensitivities and specificities above 0.80 for clinically significant prostate cancer detection [83]. Furthermore, integrating imaging with clinical data such as prostate-specific antigen levels, age, and biopsy history enhances the reliability of classification models, enabling a more comprehensive assessment of disease status [84]. These advances highlight the clinical potential of classification in prostate cancer, not only for assisting radiologists in routine diagnosis but also for reducing variability in interpretation and improving personalized patient care.

### 2.8.2 Segmentation

Segmentation is a crucial task in medical imaging, involving the delineation of structures, organs, and lesions within medical images. Numerous methods have been developed for both automatic and interactive segmentation of these images [82]. Prostate segmentation and the precise identification of the deformable

prostate capsule are essential for various applications, including radiation treatment planning, volume measurements, fusion-targeted biopsies, and monitoring prostate disease progression over time [85].

The zonal segmentation of the prostate, which involves the distinct identification of the peripheral zone and central gland, is also essential in clinical practice. This segmentation plays a key role in the location of prostate cancer and the planning of surgeries [86], as well as in standardizing the calculation of PSAD [87]. In addition, calculating the volumes of the prostate zones helps to understand urinary obstructive symptoms [88] better.

Prostate zonal segmentation has traditionally been done manually by radiologists on T2W imaging. However, this manual approach is both time-consuming and prone to variation between observers [89]. Automating prostate zonal segmentation is quite challenging for a few key reasons. First, the prostate itself can vary a lot in shape and internal structure, and it often has low contrast compared to nearby tissues, making it hard to clearly define the different zones. On top of that, it's tricky to ensure these automated methods work well across different hospitals or imaging centers because of the wide range of technical differences in how MRIs are taken; things like signal strength, scanning protocols, equipment models, and coil setups can all affect how the images look [90].

Accurate segmentation of the prostate and its lesions is vital for the accurate diagnosis and effective treatment planning of prostate cancer, as it allows for precise localization and detailed characterization of cancerous tissue [91].

Also, manual tumor segmentation is a time-consuming process that depends largely on the radiologist's skill, which has driven extensive research into automated segmentation techniques to improve both speed and consistency [92].

Recent progress in DL has significantly improved segmentation in bpMRI, making it more accurate and efficient for clinical use. convolution neural networks, particularly models like U-Net [93], have become central to semantic segmentation tasks in medical imaging. Numerous adaptations of the U-Net have been developed for use in this field. Thanks to their ability to learn hierarchical features directly from data, DL methods have shown strong performance in automatically

segmenting the prostate and cancerous lesions in MRI scans [94].

### 2.8.3 Ensemble learning

Ensemble learning is a powerful framework in ML, offering considerable advantages across different domains. It involves multiple individual models working together simultaneously, with their outputs combined using a specific decision fusion strategy to deliver a unified solution for a given problem [95].

The fundamental idea behind ensemble learning is that the collective generalization ability of an ensemble usually exceeds that of a single model. Over the years, researchers have explored why ensemble methods tend to outperform individual models [96]. Technically, ensemble learning involves two main processes: training several weak learners and strategically merging these learners to create a more robust model [40]. To form an effective ensemble model, various selection strategies have been developed to identify the best component models [97]. Below, we outline some common strategies used in ensemble learning:

1. Average: This method computes the average of outputs from various classifiers, selecting the class associated with the highest average value. It is often used when the output from each classifier is numerical.

2. Weighted Average: Unlike simple averaging, this approach assigns weights to individual classifiers' outputs based on their importance, aiming to reduce errors in the ensemble's output, typically derived from an error correlation matrix.

3. Nash Vote: In this strategy, each classifier assigns a score between zero and one for each potential output, which contributes to the final decision-making process.

4. Dynamically Weighted Average: Here, the weights adjust dynamically according to the confidence levels of each classifier's outputs.

5. Weighted Average with Data-Dependent Weights: This variation calculates different weights for each partition of the input space, determined using techniques like the FSL algorithm.

6. Majority Vote: Each classifier votes for the class with the highest output, and the class receiving the most votes is selected as the final class.

7. Winner Takes All (WTA): This method selects the class with the highest output from all classifiers as the final class.

8. Bayesian Combination: This probabilistic method estimates the likelihood that a sample belongs to a specific class based on belief values.

These strategies enable ensemble learning to effectively combine the outputs of multiple classifiers, thereby enhancing performance [98]. Ensemble learning has proven to be highly effective in segmenting anatomical structures from medical images, improving the accuracy and robustness of segmentation models. In this study, we employed average ensemble learning using three U-Net-based models, Att-R-Net, Vanilla-Net, and V-Net, for prostate anatomical segmentation. Each network was trained using 5-fold cross-validation, resulting in 15 models in total. An average ensemble model was then created from these 15 models to predict prostate zones in the validation dataset. Detailed descriptions of these U-Net-based network architectures are provided in the following sections.

## 2.9 Data Handling and Model Validation Techniques

Proper data handling and validation are essential to ensure the robustness and generalizability of ML models in medical imaging. In this thesis, we employed several key techniques to enhance model performance. Data augmentation was applied to artificially increase dataset diversity and reduce overfitting, particularly important given the limited availability of annotated prostate MRI data. K-fold cross-validation was used to provide a reliable estimate of model performance by systematically training and testing across multiple subsets of the dataset, minimizing bias from data partitioning. Additionally, MRI normalization methods were implemented to standardize image intensities across patients and scanners, ensuring consistent input for DL models and improving comparability of results.

### 2.9.1 Data Augmentation

Data augmentation is a vital technique in medical image analysis, addressing challenges like limited datasets and class imbalances. By applying transformations such as rotations, flips, and translations, we can synthetically increase dataset diversity, enhancing model generalization and robustness. This is particularly crucial in medical imaging, where acquiring large, annotated datasets is often constrained by privacy concerns and resource limitations. Recent studies have demonstrated the effectiveness of data augmentation in improving DL model performance across various medical imaging tasks. For instance, a study by Cirillo et al. [99] evaluated the impact of different augmentation techniques, including rotation, flipping, cropping, and translation, on prostate DWI data. The findings indicated that these augmentations significantly enhanced the accuracy of convolution neural networks in detecting prostate cancer. In our prostate lesion segmentation project, we employed data augmentation techniques such as rotation, horizontal flipping, and translation. These augmentations were applied to both zonal segmentation and lesion segmentation tasks, aiming to improve model robustness and accuracy.

### 2.9.2 K-fold cross-validation

K-fold cross-validation is a widely used resampling technique in ML and medical image analysis, particularly valuable when working with limited datasets. In this approach, the dataset is divided into k equally sized folds. The model is trained on k–1 folds and validated on the remaining one, repeating this process k times so that each fold serves once as the validation set (Figure 2.9). The final performance is averaged over all folds, offering a more stable and unbiased estimate compared to a single train-test split. This method is especially beneficial in medical imaging applications, where data scarcity and class imbalance are common challenges. It enables the efficient use of all available data for both training and evaluation, reduces variance in performance estimates, and provides a more accurate indication of the generalizability of the model to unseen data. According to a study in Medical Image Analysis, cross-validation not only enhances model evaluation but

also contributes to selecting robust models for clinical tasks [100]. In our prostate lesion segmentation project, we applied 5-fold cross-validation for both the zonal and lesion segmentation tasks. This allowed the model to be trained and validated on diverse subsets of data, improving its robustness and reducing overfitting. For the final evaluation on the test set, we averaged the predictions from the five models trained during cross-validation to obtain a more consistent and accurate result.

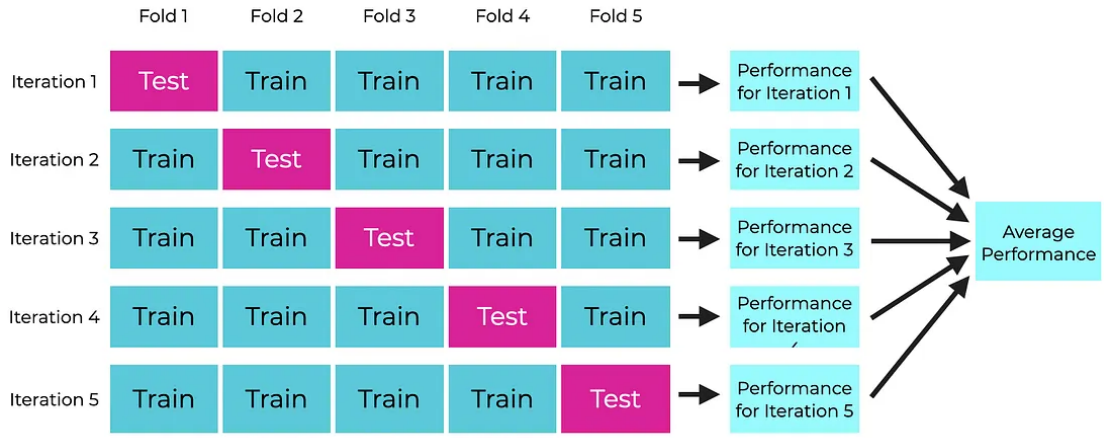



Figure 2.9: *The mechanism of 5-fold cross-validation. Image obtained from https://python.plainenglish.io/cross-validation-in-machine-learning-enhancing-model-performance-with-confidence-39402f6ca60f*

### 2.9.3 MRIs Normalization

MRI intensity normalization is a critical preprocessing step in medical imaging, aimed at reducing variability in image intensities caused by differences in scanner hardware, acquisition protocols, and patient-specific factors [101]. This process enhances the reliability of subsequent analyses such as segmentation, registration, and radiomic feature extraction [102]. There are several normalization methods that were used for different sequences of mp-MRI in previous studies. In our research, we applied Min-Max normalization for ADC and T2W images and Z-score normalization for DWI images.

### Min-Max Normalization

Min-Max normalization is a commonly used technique that scales intensity values to a fixed range, typically [0, 1], preserving the relationships between the original values while reducing the influence of intensity outliers.

Let $x$ be the original voxel intensity, $X_{\min}$ and $X_{\max}$ be the minimum and maximum intensity values within the image volume, respectively. The normalized intensity $x'$ is computed as:

$$x' = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

This method is especially suitable for MRI modalities such as ADC and T2W, where absolute intensity values are meaningful and bounded [103].

### Z-Score Normalization

Z-score normalization, as standard score normalization, centers the data by subtracting the mean and scales it by the standard deviation, resulting in a distribution with a mean of 0 and a standard deviation of 1. This technique is effective for harmonizing intensities across subjects and scanners, especially in high-b-value DWI, where signal variability is pronounced.

Let $X$ be a voxel intensity, $\mu$ be the mean, and $\sigma$ be the standard deviation of the voxel intensities within the image volume. The normalized $x'$ is:

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

This approach was applied to the DWI images to account for their typically broad integral range and variable contrast properties [102].

## 2.10 Evaluation Metrics for Segmentation and Classification

### 2.10.1 Segmentation Metrics

For segmentation performance, we employed the DSC and IoU, which quantify the overlap between predicted and ground truth masks.

The DSC is defined as:

$$\text{DSC} = \frac{2|P \cap G|}{|P| + |G|} \tag{2.1}$$

where $P$ is the set of predicted pixels and $G$ is the set of ground truth pixels. DSC ranges from 0 to 1, with 1 indicating perfect overlap.

The IoU, also known as the Jaccard Index, is:

$$\text{IoU} = \frac{|P \cap G|}{|P \cup G|} \tag{2.2}$$

IoU also ranges from 0 to 1 and provides a complementary measure of segmentation accuracy.

### 2.10.2 Classification Metrics

For classification, we used metrics derived from the confusion matrix: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

- **Accuracy**: Overall proportion of correctly classified samples.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{2.3}$$

- **Precision**: Proportion of correctly predicted positive cases among all predicted positives.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2.4}$$

- **Recall (Sensitivity)**: Proportion of correctly predicted positive cases among

all actual positives.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{2.5}$$

- **F1-score**: Harmonic mean of precision and recall, balancing false positives and false negatives.

$$F1 = 2 \times \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{2.6}$$

- **Area Under the ROC Curve (AUC)**: Measures the ability of the classifier to discriminate between classes across all thresholds, with 1 indicating perfect discrimination.

- **Matthews Correlation Coefficient (MCC)**: Accounts for all four confusion matrix categories and is especially useful for imbalanced datasets.

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{2.7}$$

These metrics together provide a comprehensive evaluation of both segmentation quality and classification performance, capturing overlap accuracy, predictive reliability, and robustness to class imbalance.

## 2.11 Software and Hardware

### 2.11.1 3D Slicer

3D Slicer is a widely adopted open-source software platform designed to analyze and visualize medical images. It offers robust automated and interactive segmentation tools, image registration, and volume rendering. Additionally, 3D Slicer is extensively used in research related to image-guided therapy, supporting both clinical and experimental workflows [104].

In this study, the ground truth segmentation masks for the MRI dataset, including T2W, ADC, and DWI modalities, were manually generated using 3D Slicer, an

open-source software platform for medical image informatics and image processing. 3D Slicer offers a comprehensive suite of tools for multi-modal image visualization, manual and semi-automatic segmentation, and volumetric analysis, making it particularly suitable for research applications in radiology. The software's support for DICOM formats and multi-planar image navigation allowed precise delineation of the region of interest across all imaging modalities. All segmentation tasks were performed using the "Segment Editor" module in 3D Slicer, enabling accurate and consistent creation of ground truth masks that served as the reference standard for subsequent analysis and model evaluation [104].

Furthermore, as part of this study, we developed a custom 3D Slicer extension for zonal segmentation, leveraging the platform's integrated Extension Wizard. This plugin was specifically designed to assist in the delineation of anatomical zones within the prostate, streamlining and automating aspects of the segmentation workflow within the 3D Slicer environment, and improving consistency across cases.

### 2.11.2 Python

The majority of the implementation work in this thesis was conducted using the Python programming language, chosen for its extensive ecosystem and strong support for scientific computing and ML. Development primarily relied on a variety of widely used libraries and frameworks, including NumPy [105] for numerical computations, Matplotlib [106] for data visualization, TensorFlow (with its high-level API Keras) [107] for building and training DL models, and SimpleITK [108] for medical image processing and manipulation. All Python development was performed within the Spyder integrated development environment (IDE), which provided a user-friendly interface for writing, debugging, and testing the code efficiently. In addition to TensorFlow, PyTorch [109] was also employed for implementing certain DL architectures. Specifically, networks such as nnU-Net and You Only Look Once You Only Look Once (YOLO) for segmentation tasks were developed using the PyTorch framework due to its flexibility and dynamic computation graph, which are particularly beneficial for research and rapid prototyping.

### 2.11.3 Hardware

All experiments were conducted on a high-performance workstation equipped with two NVIDIA GeForce RTX 2080 GPUs, each with 11 GB of dedicated VRAM. The system included 16 GB of DDR4 RAM operating at 2933 MT/s and was powered by an AMD Ryzen Threadripper 2950X 16-core processor with 32 logical CPUs (threads). This setup provided sufficient computational resources for training and evaluating DL models efficiently.

# Chapter 3

# State of the Art

The literature review in this thesis is structured to provide a comprehensive overview of prior research relevant to the four major components of our work. The first section focuses on prostate zonal segmentation, reviewing traditional and DL–based approaches for delineating the peripheral, transition, and central zones of the prostate on mpMRI. This is followed by an examination of studies on prostate lesion segmentation, where automated methods have been proposed to localize and quantify suspicious regions, highlighting their role in supporting biopsy targeting and treatment planning. The third section explores research on PI-RADS prediction, summarizing advances in AI techniques aimed at standardizing lesion evaluation and improving consistency across radiologists. Finally, the review discusses prior work on prostate cancer detection, particularly studies that integrate imaging features with clinical data to distinguish between malignant and benign findings. Together, these four parts not only summarize the state of the art but also identify the limitations and gaps that motivate the contributions of this thesis.

## 3.1 Prostate Zonal Segmentation

Study [110] evaluated various DL models for prostate segmentation using 3T T2W MRI scans from 204 patients in the PROSTATEx dataset. Manual annotations

of different prostate zones by four operators were used to train the U-Net, ENet, and ERF-Net architectures. Among these, ENet delivered the highest accuracy, achieving Dice similarity coefficients of 91% for the whole gland, 87% for the transition zone, and 71% for the peripheral zone. While U-net and ERF-Net performed slightly less accurately, ENet also demonstrated the shortest training and inference times, highlighting the effectiveness of DL, particularly ENet, for segmenting the prostate in T2W images.

A BASC-Net model was developed in [111] to automatically segment prostate zones in MRI, aiming to enhance prostate cancer diagnosis by accurately delineating the peripheral zone and central gland. Its architecture integrates a Semantic Clustering Attention (SCA) module for extracting features and generating attention maps, along with a Boundary-Aware Contrastive (BAC) loss function that strengthens feature similarity within the same zone while differentiating between zones. When evaluated on the NCI-ISBI 2013 Challenge and Prostate158 datasets, BASC-Net outperformed nine other existing methods. It achieved Dice similarity coefficients of 79.9% for the peripheral zone and 88.6% for the central gland on the NCI-ISBI dataset, and 80.5% for the peripheral zone and 89.2% for the central gland on Prostate158.

An ensemble-based approach was introduced to address both anatomical prostate zone segmentation and the detection of pathologically suspicious regions. This method incorporates pre-training into the standard nnU-Net framework and introduces a modified loss function designed to handle variability in expert annotations. The anatomical segmentation model achieved strong performance, with Dice similarity coefficients of 0.915 for the entire prostate gland, 0.865 for the transition zone, and 0.736 for the peripheral zone, highlighting the effectiveness of this enhanced framework [112].

A 3D U-Net model was designed to perform automatic segmentation of prostate anatomy using MRI data from 223 patients, comprising an internal cohort of 93 cases and two external datasets: ETDpub (n = 141) and ETDpri (n = 59). The

model's accuracy was assessed using Dice similarity coefficients Dice Similarity Coefficients (DSC), 95th percentile Hausdorff distance (95HD), and average boundary distance (ABD), and its results were compared to those of a junior radiologist. For the central gland, the DSCs were 0.909, 0.889, and 0.869 across the respective datasets, while for the peripheral zone, scores were 0.844, 0.755, and 0.764. Notably, the model outperformed the radiologist in peripheral zone segmentation, achieving a DSC of 0.769 compared to 0.706, and also demonstrated superior performance in volume estimation. The study identified central gland volume and MRI vendor as key factors affecting segmentation quality [113].

A new model known as convolution coupled Transformer U-Net (CCT-Unet) was proposed in [114] for prostate segmentation, aiming to leverage the complementary strengths of convolution neural networks and Transformer-based architectures. While Transformers are effective in capturing global context, they often lack sensitivity to local variations, particularly in small prostate MRI datasets. To address this, CCT-Unet incorporates a convolutional embedding block to retain edge information and a convolutional Transformer block to enhance local feature extraction. Experimental results on the ProstateX and Huashan datasets demonstrated that CCT-Unet outperformed existing models, achieving Dice similarity coefficients of 80.39% for the peripheral zone and 87.49% for the transition zone.

Using a diverse dataset of 243 T2W prostate MRI scans collected from seven countries and acquired on ten scanners from three different vendors, a U-Net-based segmentation model was developed with deep supervision and a cyclical learning rate strategy, as described by the authors in [115]. Manual annotations of the central gland-transition zone (central zone-transition zone), peripheral zone, and seminal vesicles (SV) were provided by two experienced radiologists to serve as ground truth. The model was evaluated on 120 test cases, achieving Dice similarity coefficients of $0.88 \pm 0.01$ for the entire prostate gland, $0.85 \pm 0.02$ for the central zone-transition zone, and $0.72 \pm 0.02$ for both the peripheral zone and SV. DSC values above 0.7 were considered indicative of accurate segmentation.

An automated ML approach was proposed for prostate segmentation using MRI, featuring a novel two-stage architecture known as Green Learning (GL). In

this method, the first stage segments the entire prostate gland, while the second focuses on identifying the transition zone and peripheral zone. The study, detailed in [116], included a cohort of 119 men undergoing prostate MRI and biopsy, with manual annotations provided by experienced radiologists. The model was trained on 95 MRI scans, comprising a total of 19,992 T2W images. It achieved mean Dice scores of 0.85 for the whole prostate, 0.81 for the transition zone, and 0.62 for the peripheral zone. In terms of volume estimation, Pearson correlation coefficients were 0.92, 0.93, and 0.62, respectively, all statistically significant ($p < 0.01$). In addition to strong performance, the study also introduced a web-based interface allowing clinicians to refine segmentation results efficiently.

A segmentation framework named PPZ-SegNet, based on a 2D-3D convolution neural networks ensemble, was introduced to automatically delineate the prostate gland and peripheral zone in T2W MRI scans. As reported by the authors in [117], the model was developed and optimized using Bayesian hyperparameter tuning and trained on 150 cases. It leveraged data from four public datasets—including Train 1, Test 1 (from the same patient cohort), and the independent Test 2, Test 3, and Test 4. While anatomical annotations were manually performed by a radiologist for most datasets, Test 4 employed existing ground truth labels. The final model was evaluated on 283 unseen T2W MRI cases from The Cancer Imaging Archives without additional tuning. Evaluation metrics included the DSC and Hausdorff distance, with the model achieving average Dice scores of 0.86, 0.79, 0.81, and 0.62 across the respective test sets.

Dense U-Net, a novel architecture that integrates principles from DenseNet and U-Net, was proposed to enhance automatic segmentation of the prostate and its anatomical zones. Trained on axial T2W MRI scans from 141 patients and evaluated on a separate test set of 47 patients, this network showed robust performance, even when trained with somewhat imprecise labels. According to findings reported by the authors in [118], Dense-2 U-Net outperformed the standard U-Net model, achieving an average Dice score of 92.1% for the entire prostate compared to 90.7% with U-Net, and 78.1% for the peripheral zone, highlighting its improved

accuracy in zonal segmentation. An advanced segmentation technique was developed to delineate the whole prostate gland (WG), central gland, and peripheral zone using both ADC and T2W MRI sequences. As outlined in the work by [119], the framework utilized two separate models, each composed of two U-Net architectures, trained on data from 225 patients. When evaluated on the test set, the method achieved excellent Dice similarity coefficients: 95.33% for WG, 93.75% for the central gland, and 86.78% for the peripheral zone using T2W images. With ADC images, the Dice scores were slightly lower but still strong, reaching 92.09% for WG, 89.89% for the central gland, and 86.1% for the peripheral zone.

## 3.2 Prostate Lesion Segmentation

A study by the authors in [94] focused on segmenting prostate lesions using a total of 140 MRI scans, split evenly between the ProstateX dataset and Prostate158. These cases included both lesion-present and lesion-free examples. They proposed a two-step method: first, identifying the prostate region, then detecting lesions within it. Their model, a hybrid architecture called a dual-pyramid network, blends convolution neural networks with a tokenized multilayer perceptron Multilayer Perceptron (MLP) to better capture features at multiple scales. With the help of attention mechanisms, their approach improved segmentation performance, reaching a Dice score of $56.31 \pm 22.77$ and an Intersection over Union Intersection over Union (IoU) of $44.82 \pm 20.45$ on T2W images.

The researchers introduced a smart Adaptive Window Adjustment (AWA) module that can fine-tune window settings depending on the image type and the task at hand. According to [120], they also took a deep dive into comparing loss functions for prostate segmentation and found that the Boundary Difference over Union (DoU) Loss outperformed others, especially in 3D scenarios. Their approach doesn't stop there; they also designed a cascaded segmentation method that leverages the anatomical layout to improve lesion detection. By combining T2W and ADC images, their multi-stage model was able to segment both the prostate and tumor regions. In tests using Prostate158, ProstateX, and PI-CAI

datasets, they reported mean Dice scores of 48.56 and 49.03, respectively, using 5-fold cross-validation, pointing to the reliability of their framework.

A well-structured model was presented to improve lesion segmentation in prostate MRI, combining three powerful components: SE-Net, Cascade Pyramid Convolution, and an Attention Gate. These elements work together to boost feature extraction. To further refine the segmentation results, the model also includes multi-scale convolutions and a convolution-based edge detection block. It was tested on 390 MRI scans from Guizhou Provincial People's Hospital, and it performed impressively, achieving a DSC of $86.3 \pm 0.027$, showing strong generalization capabilities even during external validation [121].

In [92], the authors introduced a two-stage framework called NaMa (Neighbor-aware Multi-modal Adaptive learning network) aimed at segmenting prostate tumors using anisotropic multi-modal MRI data. At the heart of their approach are two key components: the Modality-Informative Adaptive Learning (MIAL) module, which smartly fuses information from different MRI modalities, and the Activation Mapping Guidance (AMG) module, which helps refine inter-slice feature consistency. To further strengthen feature learning, they used a random masking technique during training. Their method reached a Dice score of 63.76% on their internal dataset and 59.01% on the PI-CAI dataset, showing promising results in cross-dataset performance.

A cascaded two-step SegResNet model was introduced for prostate lesion segmentation in bpMRI. The first stage locates the region of interest, and the second uses T2W, DWI, and ADC to segment the lesion. Trained on the PI-CAI 2022 dataset, the approach, as reported in [122], outperformed the standard SegResNet, with Dice scores of 0.5334 on PI-CAI 2022, 0.4221 on Prostate-158, and 0.3588 on PSU-mpMRI, showing good external generalization.

Lisa C. Adams et al. introduced a U-ResNets model for segmenting prostate cancer lesions using the Prostate158 dataset. The model's performance was measured using the Dice score, reaching 45% in lesion segmentation [123].

A U-Net model was developed to segment prostate lesions, reaching a Dice score of 77% for accuracy. The study, as described in [124], used data from 149

patients (mean age 69.2, range 47–84), split into 80% for training and 20% for internal validation.

In [125], the authors developed an advanced segmentation network combining a cascading pyramid convolution module (CPCM) and a double-input channel attention module (DCAM) to automatically segment prostate cancer lesions from mp-MRI scans. Tested on the I2CVB dataset, their model achieved a Dice score of 79% for lesion delineation.

The study in [126] used the I2CVB dataset to segment prostate cancer lesions. They applied a deep convolutional encoder-decoder model that segmented the prostate, anatomical structures, and malignant lesions, achieving an IoU score of 67% for cancer segmentation.

A new 3D cascaded scoring convolution neural network was developed to automatically segment both the prostate and lesions from MRI images. This network works by first identifying the region of interest, then detecting the object within that region of interest, and finally defining the target area. The method was evaluated in a retrospective study with 77 patients who had MRI scans. The results showed the network achieved a Dice score of 69% for lesion segmentation [127].

A study trained and evaluated a mask region-based convolution neural network for segmenting the prostate gland and lesions. It involved two groups: 78 patients from public datasets (cohort 1) and 42 patients from the authors' institution (cohort 2). The model reached a Dice score of 62% for lesion segmentation [128].

In [129], an FCN was trained on 152 MRI scans to segment prostate lesions. The network achieved a Dice score of 41%.

An experiment used a semi-supervised method combined with a support vector machine Support Vector Machine (SVM) classifier on multispectral MRI data, testing different feature vectors for segmentation. The study, as reported in [130], used 16 MRIs including T2 maps and ADC images and achieved a Dice score of 58% for lesion segmentation, showing improved localization accuracy.

A new method for segmenting prostate cancer in multispectral MR images was introduced using fuzzy Markov random fields (fuzzy MRFs). The model was tested

on 11 MRIs and achieved a Dice score of 62% for lesion segmentation [131].

A non-local Mask R-convolution neural network was improved using different training strategies, and its performance was assessed in a retrospective study of 262 prostate bp-MRI scans. As detailed in [132], the model was compared with baseline Mask R-convolution neural network, 3D U-Net, and manual radiologist annotations using metrics like detection rate, DSC, sensitivity, and Hausdorff Distance. The data was split into three cohorts: 64 patients with histopathology (Cohort 1), 158 with bp-MRI (Cohort 2), and 40 for semi-supervised learning (Cohort 3). The enhanced model achieved the highest DSC of 60%.

In [133], the researchers presented SAM-UNETR, a new method for segmenting clinically significant prostate cancer in MRI scans. The model integrates a transformer-based encoder adapted from the Segment Anything Model (SAM), trained on 11 million images, with a UNETR-style residual-convolution decoder. The study used two datasets, Prostate158 and the PI-CAI Challenge set, each bringing distinct strengths. SAM-UNETR reached a Dice score of 48% in predicting prostate cancer regions.

This study applied advanced DL techniques to create a framework for mp-MRI image segmentation and classification. It followed a two-stage process: first, a U-Net was used to segment the region of interest, and then an LSTM network classified those regions as cancerous or not. The models were trained using the I2CVB dataset and evaluated through comparisons with prior experimental setups. The approach reached a Dice score of 67% for lesion segmentation, reflecting strong performance [134].

A transparent DL model was designed to interpret convolution neural network predictions in prostate tumor segmentation. The model relies on a U-Net architecture, trained on multi-parametric MRI scans from 122 patients. In [135], the authors report that the convolution neural network effectively segmented both the prostate gland and tumor lesions, achieving an average DSC of 62% for the gland and 31% for the tumors.

In [136], researchers assessed a cascaded DL algorithm for identifying and segmenting prostate lesions on biparametric magnetic resonance imaging biparametric

magnetic resonance imaging (bp-MRI). The model, trained and evaluated on 658 patients with 1,029 visible lesions, followed a two-step detection and segmentation strategy. Compared to histopathology and radiologist assessments, it achieved a lesion-level DSC of 0.29. While the sensitivity was moderate, the findings emphasize the potential role of AI in supporting prostate cancer diagnosis.

To enhance prostate cancer segmentation on mpMRI, the authors in [137] introduced Matt-Unet, a model that extends Attention-Unet by adding multiscale modules and a composite loss function. They incorporated both background context and prostate gland masks into the process. Using a dataset of 232 patients who had mpMRI followed by biopsy, Matt-Unet outperformed the standard U-Net, reaching a DSC of 0.7215 and an accuracy of 0.8899, compared to the U-Net's 0.6592 DSC and 0.8601 accuracy.

In [138], a novel AI framework was proposed for prostate cancer lesion segmentation and PI-RADS grading using mp-MRI. The framework includes a Lesion-guided Selective Multi-modal Integration (LeSMI) module, which applies Cross Modality Attention (CMA) and Localized Lesion Attention (LLA) to adaptively integrate key information across and within imaging modalities. Evaluated on Prostate158 and PI-CAI Challenge datasets, the model achieved DSCs of 0.5130 and 0.4381, respectively, and showed improved PI-RADS grading compared to existing approaches.

Thipkasorn et al. [122] proposed a cascaded DL approach using two SegResNet models for prostate cancer lesion segmentation in bp-MRI. The first model localizes the prostate, and the second segments lesions. Trained on PI-CAI 2022 and validated on Prostate-158 and PSU-mpMRI, it achieved DSCs of 0.5334, 0.4221, and 0.3588, respectively, outperforming a single SegResNet and showing good generalizability.

In [139], the authors used ADC MRI maps to segment dominant index lesions in prostate cancer with DL. Six models were tested on three datasets. On the internal dataset (156 lesions, 125 patients), MRRN-DS achieved DSC 0.60, outperforming FPSnet (0.51) and FPSnet-SL (0.54). On ProstateX (299 lesions, 204 patients), MRRN-DS scored 0.54 vs. ResUnet (0.44). On a third dataset (10

lesions), MRRN-DS got 0.45, while FPSnet and FPSnet-SL dropped to 0.06 and 0.24. MRRN-DS showed consistent performance and strong generalizability.

A 3D U-Net model was enhanced by integrating PSAD and patient age using feature-wise transformations (FWTs), allowing clinical biomarkers to influence latent representations and potentially improve segmentation accuracy of prostate cancer lesions on bi-parametric MRI. When evaluated on 214 annotated cases from the PI-CAI challenge dataset, the model reached a mean DSC of 0.36 [140].

A dual-branch, attention-driven multi-scale learning network was developed to automatically segment prostate tumors using mp-MRI data, encompassing T2W, DWI, and DCE sequences. Evaluated on 116 male pelvic magnetic resonance imaging cases, the model achieved a Dice score of 84. 39% for the segmentation of the lesion, demonstrating strong and consistent performance [141].

In [142], a 3D convolutional network incorporating visual and multi-scale attention mechanisms (3D DAG-Net) was employed for segmenting prostate lesions in MRI scans. The study utilized a dataset of 786 DWI MRI images. The model reached a Dice Similarity Coefficient of 92.28% for the prostate and 88.73% for lesion segmentation, emphasizing the effectiveness of dual attention strategies in enhancing segmentation accuracy.

A SegNet-based approach applied to the PROSTATEx dataset achieved Dice scores of 90.45% for the transition zone, 70.04% for the peripheral zone, and 52.73% for segmenting prostate cancer lesions. These findings illustrate the ability of various networks to effectively segment prostate cancer lesions from multi-sequence MRI, underscoring the potential of AI-driven tools in advancing prostate cancer management [143].

AutoProstate [144] used a DL pipeline for prostate lesion segmentation on MRI, trained on PROSTATEx (204 exams) and PICTURE (249 exams). It employed a clinically significant prostate cancer segmenter with test-time augmentation, dropout, and ensembling. The model achieved a mean DSC of 0.46, a median of 0.58, and a range of 0.00–0.90. Its sensitivity matched expert radiologists, though with more false positives, showing promise for clinical support in prostate cancer assessment.

The study [145] utilized a cascaded Res-UNet for fully automated segmentation of clinically significant prostate cancer on MRI. Trained and tested on 347 patients (235 with csPCa) using T2W and ADC images, the model achieved a mean DSC of $0.64 \pm 0.24$ with bp-MRI and $0.66 \pm 0.23$ using ADC alone. It segmented the whole prostate, zones, and clinically significant prostate cancer lesions.

Shiri et al. [146] applied AI for prostate lesion segmentation on T2W MRI using data from 400 patients across 8 centers. Their two-stage DL model combined a modified 3D U-Net and a Dual Attention 2D U-Net, trained with the Federated Averaging (FedAvg) algorithm. On a 30% test set, the federated model achieved a Dice of $0.84 \pm 0.05$ and Jaccard of $0.72 \pm 0.08$, outperforming center-based models (Dice $0.77 \pm 0.06$, Jaccard $0.63 \pm 0.09$).

Authors in [147] proposed a three-stage cascaded model combining 3D U-Net and 3D ResNet architectures for prostate segmentation, lesion detection, and PI-RADS classification. The model was trained and evaluated on 1,247 scans (1,043 in-house and 204 from ProstateX), achieving a lesion segmentation DSC of 0.385, detection sensitivity of 71.65%, and PI-RADS classification accuracy of 30.8%.

## 3.3 PI-RADS Prediction

Sciarra et al. [148] extracted 609 PyRadiomics texture features from mp-MRI of 90 prostate lesions to perform multiclass classification across PI-RADS 1–5. After two feature selection steps, they applied Naïve Bayes (120 feature subset) and Random Forest (52 feature subset) classifiers, achieving top AUCs of 0.7442 and 0.7394, respectively. This demonstrates the feasibility of using radiomics across all PI-RADS scores using standard clinical MRI protocols.

The study [149] built a DL model to sort prostate lesions into PI-RADS categories 2 to 5 using mp-MRI scans (T2W, DWI, ADC) from 687 patients. On a slice level, the AI matched the radiologists' scores 58% of the time, and the same percentage held for lesion-based comparison, with a moderate agreement (kappa = 0.40). The best match was for PI-RADS 5 (80%), while the weakest was for PI-RADS 2 (6%). The model often upgrades lesions, especially those labeled as

PI-RADS 3. When a $\pm 1$ category difference was allowed, the accuracy jumped to 86%, showing that the model can be a helpful support tool but not a complete replacement for radiologists.

Winkel et al. [150] developed a DL algorithm to automatically classify prostate lesions (PI-RADS 1–5) using bi-parametric MRI (T2W and DWI/ADC) in 121 patients. Compared with five radiologists, the AI achieved an AUROC of 0.828 for detecting clinically significant cancer (Gleason $\geq 7$), performing better than less experienced readers but below an expert (0.914). At PI-RADS $\geq 4$, it reached 76.7% sensitivity and 85.9% specificity. While matching average clinical performance, it did not reduce variability between readers.

In [151], 252 lesions from 188 patients were analyzed using bi-parametric MRI (T2W and ADC). The study aimed to refine PI-RADS classification by using mean ADC (mADC) and radiomic features. Lesions rated PI-RADS 3–5 were sometimes downgraded, and PI-RADS 1–2 were upgraded based on quantitative thresholds. The "Down(ADC)" approach increased specificity (39.2% $\rightarrow$ 56.7%) with a small drop in sensitivity, while radiomic-based methods performed similarly, improving PI-RADS assessment without full ML models.

Radiomic features from T2W and ADC MRI, combined with clinical data, were used to improve the detection of clinically significant prostate cancer (Gleason $\geq 7$) in PI-RADS 4–5 lesions. ML models trained on 111 lesions from 99 patients (71% significant) achieved 79% accuracy, with approximately 79% sensitivity and 80% specificity. For peripheral zone lesions, accuracy rose to 84% and sensitivity to 86%, showing that combining radiomics with clinical data boosts performance [152].

The study [153] explored using MRI radiomics combined with ML to enhance PI-RADS v2 in detecting clinically significant prostate cancer. MRI scans from multiple patients, including T2W, DWI, and ADC sequences, were analyzed. The model classified lesions as clinically significant (PI-RADS 4–5) or non-significant/benign (PI-RADS 1–3). The combined approach achieved an AUC of 0.85, with 83% sensitivity and 78% specificity, showing improved diagnostic accuracy compared to PI-RADS v2 alone.

In a retrospective study of 203 patients (141 for training, 62 for validation), radiomic features were extracted from T2W, DWI, ADC, and DCE MRI using PyRadiomics [154]. A ML–based Rad-score was created and combined with PI-RADS v2.1. This combined model outperformed PI-RADS alone in detecting prostate cancer, with AUCs of 0.989 vs. 0.905 for training and 0.931 vs. 0.845 for validation. It also improved sensitivity (92.3% to 79.4%) and specificity (98.4% to 96.4%), particularly for PI-RADS 3 and peripheral zone lesions. Adding prostate-specific antigen did not enhance performance. Radiomics helped overcome the limitations of PI-RADS in terms of specificity and lesion stratification.

In a multicenter study, Zhu et al. [155] evaluated 1,186 lesions from 927 patients to predict clinically significant prostate cancer using radiomics. A logistic regression model classified lesions as PI-RADS 3–5. The radiomics model achieved AUCs of 0.85 (training), 0.87 (internal test), and 0.83–0.85 (external validation), with accuracy between 75.5% and 81.4% and sensitivities up to 89.2%. LASSO was used for feature selection, and radiomics features were combined with radiologist-assigned PI-RADS scores. The approach improved the detection of clinically significant cancer, particularly for PI-RADS 3 lesions, highlighting its potential for clinical application.

In the study [156], 615 patients were categorized into clinically significant (grade group $\geq 2$) and non-significant/benign lesions using a deep radiomics approach combining nnU-Net segmentation with an XGBoost classifier. The model reached a patient-level AUROC of 0.91, with 90% sensitivity and 73% specificity, similar to PI-RADS $\geq 3$ (AUROC 0.94). However, lesion-level sensitivity was lower (68% vs. 84% for PI-RADS). These findings indicate that deep radiomics could support PI-RADS but is not yet ready to replace it in clinical practice.

In a study of 102 patients with PI-RADS 3 and upgraded PI-RADS 4 lesions, a radiomics approach using mp-MRI was applied. Features extracted from T2W and ADC images were analyzed with a random forest classifier to identify clinically significant prostate cancer (Gleason score $\geq 7$) versus non-significant cases. The model reached an AUC of 0.82, outperforming PI-RADS alone in lesion stratification. Preprocessing and feature selection were used to improve accuracy, targeting

better classification specifically for PI-RADS 3 and 4 lesions [157].

The study [158] created a semi-automated ML system to score PI-RADS v2.1 using mp-MRI from 59 patients (PI-RADS 2:18, 3:10, 4:16, 5:15). Two approaches were tested: multiclass classification across PI-RADS 2–5 and binary classification separating low-risk (2+3) from high-risk (4+5) lesions. The system combined prostate segmentation, 3D co-registration, and lesion region of interest extraction with classifiers such as LDA, linear Support Vector Machine, and Gaussian Support Vector Machine. The multiclass model achieved 88.0% $\pm$ 0.98% accuracy and an AUC of 0.94, while the binary classifier performed better, with 93.2% $\pm$ 2.1% accuracy and an AUC of 0.99.

The study [159] evaluated a commercial deep learning AI for automated PI-RADS v2.1 scoring and lesion detection using mp-MRI from 91 patients across three centers. For clinically significant prostate cancer ($PI-RADS \geq 4$ ), the AI achieved 81% sensitivity and 78% specificity at the lesion level, comparable to radiologists' performance (90% sensitivity, 70% specificity). The approach utilized a convolution neural network trained on large datasets for lesion segmentation and classification.

mp-MRI from 346 patients was used to create a DL ensemble model combining ResNet and DenseNet for classifying (clinically significant prostate cancer, $PI-RADS \geq 4$) versus non-significant cases (PI-RADS $\leq$ 3). Individually, ResNet reached 87% accuracy and DenseNet 85%, while the ensemble improved performance to 89% accuracy, achieving 91% sensitivity and 87% specificity [160].

In a study of 453 patients with prostate-specific antigen 4–10 ng/mL, combining PI-RADS v2.1 scores and PSAD improved the detection of clinically significant prostate cancer. Both factors were independent predictors, and a logistic regression model achieved an AUC of 0.935, with better sensitivity and specificity than either alone. Low-risk patients (PI-RADS $\leq 2$ or PI-RADS 3 with PSAD $\leq 0.33$ ng/mL/mL ) had minimal clinically significant prostate cancer detection, suggesting biopsies could be avoided [161].

The study [162] evaluated 287 PI-RADS 3 lesions using logistic regression, Support Vector Machine, XGBoost, and random forest classifiers to distinguish

clinically significant prostate cancer from non-significant or benign lesions. The random forest model performed best, achieving an AUC of 0.832 (internal) and 0.688 (external), with 87% sensitivity and 50% specificity. Integrating clinical and imaging features helped minimize unnecessary biopsies.

A radiomics-based ML model was developed using 107 features from T2W MRI to classify clinically significant prostate cancer versus non-significant lesions in equivocal PI-RADS 3 cases. A random forest classifier achieved an AUC of 0.76, outperforming prostate-specific antigen density (AUC 0.61) and prostate volume (AUC 0.62), enhancing diagnostic accuracy for clinically significant prostate cancer stratification in PI-RADS 3 lesions [163].

## 3.4 Prostate Cancer Detection

Taguelmimt et al. [164] developed a bimodal DL model combining bp-MRI with clinical data (prostate-specific antigen, prostate volume, age) for prostate cancer detection. They used a convolution neural network architecture with uncertainty quantification to evaluate prediction confidence. The dataset included bpMRI and clinical data, though the exact size was not reported. Their model achieved an AUC of 0.82±0.03 and sensitivity of 0.73±0.04, outperforming unimodal approaches. This study highlights the benefit of integrating clinical information and uncertainty measures to improve MRI-based prostate cancer detection reliability.

In [165], a self-adapting DL model using 3D nnU-Net was developed to classify clinically significant prostate cancer versus non-cancer regions on bp-MRI (T2W, ADC, high-b DWI). The network incorporates probabilistic prostate masks and anatomical priors to focus on relevant regions. It was trained on 1,500 multi-center MRI scans and tested on 1,036 in-house scans. The model achieved an AUROC of 0.888–0.889 in validation/testing and 0.886 in independent data.

The study [166] aims to evaluate different convolution neural network models for prostate cancer classification using a dataset of 400 prostate MRI images. The ResNet50 DL model was employed for feature extraction and classification. The performance of ResNet50 was compared with other convolution neural network

architectures such as VGG16 and VGG19. The results showed that ResNet50 outperformed the other models in terms of sensitivity, specificity, and overall accuracy. Specifically, ResNet50 achieved a classification accuracy of 99.64%, demonstrating its effectiveness for prostate cancer detection.

A computer-aided diagnosis model was developed for prostate cancer classification using multi-parametric MRI (mpMRI). The approach combined a ResNet50 deep neural network for feature extraction with a random forest classifier. Using a dataset of 400 prostate MRI images, the model achieved 99.64% accuracy, outperforming other convolution neural network architectures such as VGG16 and VGG19 in sensitivity, specificity, and overall classification [167].

Litjens et al. [168] developed a cascaded classification framework using a Support Vector Machine to distinguish prostate cancer from benign confounders on multi-parametric MRI. The study included 31 prostatectomy patients with voxel-wise annotations from 3T mpMRI registered to histopathology. Features were derived from T2W, diffusion, and dynamic contrast-enhanced sequences. The cascaded Support Vector Machine achieved an accuracy of 0.76, outperforming the conventional single-shot classifier.

A deep transfer learning (DTL)-based model was developed to classify prostate cancer lesions using 3T multi-parametric MRI. The study utilized 169 lesions from 110 patients, with whole-mount histopathology as the reference standard. The DTL model achieved an area under the curve Area Under the Curve (AUC) of 0.726, outperforming a DL model without transfer learning (AUC 0.687) and the PI-RADS v2 score (AUC 0.711) in distinguishing clinically significant from indolent prostate cancer lesions [169].

Wang et al. [170] developed a multi-input convolutional neural network (MISN) to classify clinically significant and clinically insignificant (CiS) prostate cancer using mp-MRI. The study utilized a dataset comprising 650 cases from a single center (400 training, 100 validation, 150 internal test) and 100 cases from three external centers for testing. The MISN model achieved an accuracy of 82.5% in distinguishing clinically significant from CiS prostate cancer. To optimize performance, a pruning strategy was employed to select the most informative sequences

for classification in specific prostate zones.

Authors in [171] evaluated ML models for classifying prostate MRI lesions as cancerous or non-cancerous using single- and multi-institutional datasets. The study used 112 lesions from the ProstateX-2 dataset and 63 lesions from a local institutional dataset, totaling 175 mpMRI lesions. Models trained on single-institution data achieved AUCs ranging from 0.82 to 0.98 on their respective datasets. Performance dropped (AUC decreased 0.23–0.50) when tested on data from a different institution. Models trained on combined datasets achieved AUCs from 0.83 to 0.99, demonstrating improved generalizability.

The study [172] aimed to improve prostate cancer diagnosis using ML on MRI images. A dataset of 738 healthy and 3,514 cancerous prostate images was analyzed. DL models VGG16, VGG19, ResNet50, and ResNet50V2 were used for feature extraction, followed by Random Forest classification. Classification results showed accuracies of 98.56% (VGG16), 99.64% (VGG19), 99.52% (ResNet50), and 99.41% (ResNet50V2). Overall, VGG19 achieved the highest performance, demonstrating the effectiveness of transfer learning for prostate cancer detection.

The PROSTATEx Challenge dataset includes mpMRI scans from 346 patients with clinically significant and CiS prostate cancer lesions. For each patient, nine MRI sequences were used: T2W in three planes (T2T, T2C, T2S), DWI with b-values of 0, 400, 800, and 1400, ADC maps, and DCE-derived Ktrans images. The dataset was divided into training (204 patients, 330 lesions) and test (142 patients, 208 lesions) sets, covering the peripheral zone, transition zone, and anterior stroma (AS) zones. DL-based feature extraction combined with ML classification achieved an overall AUC of 0.87, with zone-specific AUCs of transition zone: 0.96, peripheral zone: 0.80, and AS: 0.85 [173].

In order to improve prostate cancer detection in MRI images, [174] proposed an ensemble DL approach combining 3D convolution neural network, ResNet, and Inception models using a soft voting strategy. The method was evaluated on the SPIE-AAPM-NCI PROSTATEx dataset. It achieved an accuracy of 91.3%, a sensitivity of 90.2%, a specificity of 92.1%, a precision of 89.8%, and an F1-score of 90.0%, demonstrating strong diagnostic performance.

The study [175] proposed a 3D Efficient CapsNet model that integrates 3D Convolution neural network modules, primary capsule layers, and fully connected capsule layers, combined with a dynamic weighted margin loss to address class imbalance. The dataset included 976 patients with T2W prostate MRI images from The Cancer Imaging Archive. The model was evaluated across different stratifications, including low + medium versus high risk, low versus high, medium versus high, and low versus medium risk groups. The results showed an AUC of 0.79 and an F1-score of 0.75 for low + medium versus high risk, an AUC of 0.83 and an F1-score of 0.64 for low versus high, an AUC of 0.75 and an F1-score of 0.69 for medium versus high, and an AUC of 0.59 and an F1-score of 0.57 for low versus medium.

The authors in [176] developed a computer-aided diagnostic (CAD) system for prostate cancer detection. It used 300 mp-MRI images and applied denoising filters, segmentation via Particle Swarm Optimization (PSO), and feature extraction. For classification, support vector machine, multilayer perceptron, and LDA were employed. The system achieved an accuracy of 92.5%, and the PSO-based segmentation improved the accuracy by 7.8%.

A hybrid DL model was developed to classify prostate cancer as benign or malignant using 620 MRI images (440 for training, 180 for testing) with 256 $\times$ 256 pixels. On the ISUP Grade-wise Dataset, MobileNet achieved 88.91% accuracy and 89.22% precision, while on the Transverse Plane Dataset, MobileNet and ResNet152 reached 100% accuracy and precision. For segmentation, the PANDA dataset with U-Net achieved 98.46% accuracy and an AUC of 0.9778, demonstrating the model's effectiveness [177].

## 3.5 Existing Gaps and Contributions

Despite the impressive performance of existing deep learning methods for prostate zonal segmentation, several limitations remain. Most models achieve high accuracy for the whole gland and central zone, but segmentation of the peripheral zone is still challenging, especially in smaller or heterogeneous datasets. Differences in MRI

acquisition protocols, scanner vendors, and annotation quality can affect model generalization, and some advanced architectures require extensive computational resources or large amounts of data, limiting their applicability in routine clinical practice.

Furthermore, in many studies, the discussion of limitations is embedded within the description of each method, making it difficult to clearly identify the remaining gaps. Factors such as robustness, reproducibility across centers, and efficiency are not consistently addressed, which leaves room for further improvement.

In our study, we addressed some of these gaps by demonstrating that effective segmentation can be achieved even with smaller datasets. By employing innovative model architectures and training strategies, our framework achieves competitive performance for both central gland and peripheral zone segmentation. This not only reinforces the feasibility of obtaining high-quality results with limited data but also provides a foundation for ongoing research into practical and reliable prostate imaging solutions.

Automated segmentation of prostate lesions has seen substantial progress with deep learning techniques, yet accurately delineating lesions remains more challenging than zonal segmentation. Many recent approaches, including dual-pyramid networks, cascaded U-Nets, attention-based models, and transformer-enhanced frameworks, have demonstrated promising Dice scores ranging from approximately 0.36 to 0.92 depending on the dataset and lesion characteristics. Multi-stage strategies that first localize the prostate and then segment lesions, as well as methods incorporating multi-modal MRI (T2W, ADC, DWI) and attention mechanisms, have further improved detection of clinically significant tumors. These studies highlight the potential of AI to support radiologists in identifying subtle lesions and standardizing assessments across datasets.

Despite these advancements, several limitations persist. Segmentation performance tends to decline for low-conspicuity, lower-risk PI-RADS 3 lesions, underscoring the sensitivity of models to lesion characteristics and dataset composition. Many models rely on relatively modest datasets, limiting generalizability, and

multi-input networks may require further optimization to fully exploit complementary MRI sequences. Moreover, some architectures focus on single-modality input, such as ADC, which can restrict the ability to capture spatiotemporal or multi-parametric information.

In our study, we addressed these gaps by carefully selecting imaging sequences and leveraging temporal modeling with Bi-ConvLSTM, demonstrating that even ADC-based inputs can provide strong discriminative power when combined with appropriate network design. Our findings indicate that thoughtful integration of multi-input networks and spatiotemporal modeling can improve automated lesion segmentation, while highlighting that careful handling of multi-modal information is necessary to prevent performance degradation. Overall, this work emphasizes the importance of dataset composition, sequence selection, and model architecture in advancing reliable prostate lesion segmentation, supporting future research and potential clinical translation.

Accurate prediction of PI-RADS categories is crucial for guiding prostate cancer diagnosis, yet it remains a challenging task due to the subtle and overlapping imaging characteristics of intermediate-risk lesions. Numerous studies have explored both radiomics and deep learning approaches to automate PI-RADS classification, leveraging features from T2W, ADC, and DWI sequences, sometimes combined with clinical variables. These methods achieve promising results for high-risk lesions (PI-RADS 4–5), often reaching AUCs above 0.85 and sensitivity above 80%, demonstrating the potential of AI to support radiologists in stratifying clinically significant cancers. Ensemble models, multi-modal feature integration, and advanced classifiers such as Random Forest, XGBoost, and deep neural networks further enhance performance, particularly in challenging cases like PI-RADS 3 lesions.

Despite these advances, several limitations persist. Intermediate categories (PI-RADS 3) remain difficult to classify accurately, due to ambiguous imaging features that overlap with both low- and high-risk lesions. Many studies also rely on relatively modest datasets, which may reduce model generalizability across patient populations and imaging centers. Additionally, some approaches do not

fully exploit the complementary information across multiple MRI sequences or integrate clinical features, limiting their clinical robustness.

In our work, we addressed these gaps by combining radiomic features from multiple MRI sequences with ensemble learning techniques, improving classification performance for high-risk lesions while highlighting the particular challenges of PI-RADS 3. Although our dataset size was limited, the study demonstrates that careful feature selection, multi-sequence integration, and ensemble modeling can enhance PI-RADS prediction. These findings emphasize the need for continued refinement of feature extraction, model training, and clinical data incorporation to achieve robust, clinically applicable performance, particularly for intermediate-risk lesions.

Automated prostate cancer detection using MRI has progressed rapidly, with recent studies demonstrating strong performance using deep learning, transfer learning, and hybrid machine learning approaches. Models based on convolutional neural networks, ensemble strategies, and multi-input architectures have achieved high accuracy and AUC values, particularly when combining mp-MRI sequences with clinical information such as PSA, prostate volume, and patient age. These findings confirm the potential of AI-based systems to support prostate cancer diagnosis and reduce inter-reader variability, especially for distinguishing clinically significant from insignificant disease.

However, several limitations remain evident across existing approaches. Many studies rely on slice-based analysis rather than full volumetric information, which may overlook spatial heterogeneity and multifocal tumor patterns. Image downsampling and aggressive preprocessing, while computationally efficient, can remove fine morphological details critical for detecting subtle lesions. Generalizability also remains a concern, as a large proportion of models are trained on single-center datasets and evaluated without robust external validation. Furthermore, in many cases, ground truth labels are derived from radiologist assessments or biopsy results rather than whole-mount histopathology, introducing a degree of diagnostic uncertainty.

In our work, these limitations were explicitly considered. The slice-based design simplifies the workflow but does not fully capture the three-dimensional nature of prostate cancer, and the use of reduced image resolution may limit sensitivity to subtle findings. Additionally, training on single-center data and reliance on PI-RADS–based labels may restrict generalization. These observations point to clear directions for future research, including the use of multi-slice or full-volume inputs, higher-resolution imaging, automated lesion localization, and multi-center datasets. As larger and more diverse data become available, integrating 3D architectures, attention mechanisms, or transformer-based models could further enhance robustness and clinical applicability.

# Chapter 4

# Materials and Methods

The Materials and Methods chapter of this thesis provides a detailed description of the datasets, computational models, and implementation strategies employed in each of the four stages of this project. For each stage, prostate zonal segmentation, prostate lesion segmentation, PI-RADS prediction, and real prostate cancer detection, the corresponding dataset is described, including acquisition parameters, preprocessing steps, and any clinical or imaging annotations. This is followed by a discussion of the neural network architectures and ML models utilized for each task, highlighting specific design choices, training procedures, and hyperparameter configurations. Finally, implementation details are presented, covering aspects such as data splitting, training protocols, evaluation metrics, and software and hardware environments. By organizing the chapter in this manner, the reader gains a clear understanding of how the project was executed, how each stage builds upon the previous one, and how methodological rigor was maintained throughout the development of the AI-based prostate cancer diagnostic pipeline.

## 4.1 Prostate Zonal Segmentation

Prostate zonal segmentation is the process of delineating the anatomical zones of the prostate gland, most commonly the peripheral zone, transition zone, central zone, and anterior fibromuscular stroma, on magnetic resonance imaging (MRI).

This task is clinically significant because prostate cancer occurrence, morphology, and prognosis vary across zones, and many diagnostic systems, such as PI-RADS, explicitly depend on accurate identification of zonal anatomy. For example, cancer in the peripheral zone is typically easier to detect, while those in the transition zone present greater challenges due to heterogeneous background tissue. Accurate zonal segmentation, therefore, provides essential information for cancer localization, risk stratification, and treatment planning [11] [178].

With the advent of AI, particularly DL, automated prostate zonal segmentation has gained significant attention. Traditional manual segmentation by radiologists is time-consuming and prone to inter-observer variability, which limits reproducibility in both clinical and research settings. AI-based approaches offer the potential to provide fast, consistent, and accurate zonal delineation, which is critical for computer-aided diagnosis, quantitative imaging biomarker extraction, and large-scale studies. Furthermore, robust automated zonal segmentation is a foundational step for many downstream applications, such as prostate volume estimation, therapy response monitoring, and integration into multi-parametric diagnostic workflows [179] [168].

### 4.1.1 Data Acquisition

For zonal segmentation, we used the Prostate158 dataset, which comprises 158 annotated multiparametric 3T prostate MRIs, including T2W and DW sequences, as well as ADC maps. Magnetic resonance imaging was acquired with a Siemens VIDA and Skyra 3T clinical scanner, following standardized protocols and guidelines, including B1 calibration. The T2W sequences had a slice thickness of 3 mm, no inter-slice gap, and an in-plane resolution of 0.47 $\times$ 0.47 mm. The DWIs were acquired similarly with a slice thickness of 3 mm and an in-plane resolution of 1.4 $\times$ 1.4 mm. ADC maps were generated with b values ranging from 50 to 1000 s/mm$^3$, with some values even reaching 1400 s/mm$^3$, using pre-installed scanner software (version VE11A). After acquisition and anonymization, the images were stored on a local PACS and then transferred to an internal server. For segmentation, two experienced board-certified radiologists performed detailed pixel-by-pixel

annotations of the central gland, peripheral zone, and prostate cancer lesions on the axial T2W images using open-source ITK-Snap software (version 3.8.0) [123]. Of the 158 MRIs, 139 T2W sequences included zonal masks, which were used in our analysis.

### 4.1.2 Neural Networks for Zonal Segmentation

In the study, we implemented three distinct approaches for zonal prostate segmentation. The first approach involved using 5-fold cross-validation to train three separate models: Att-R-Net, Vanilla-Net, and V-Net, resulting in a total of 15 unique models. We then applied average ensemble learning across these models to improve the accuracy of prostate zonal segmentation.

The second approach was using the MetaNet [180] architecture, which had not been applied to segmentation tasks before. This represents the first attempt to utilize MetaNet for zonal prostate segmentation. We incorporated the models from our ensemble learning approach and integrated them through MetaNet. The best performing combinations of MetaNet models were identified based on their performance in the validation set and then tested in the test set. MetaNet's ability to integrate models and enhance performance highlights its potential as a powerful tool in medical imaging.

The third approach used YOLOv8 for detecting and segmenting prostate zones. YOLOv8, one of the most advanced iterations of the YOLO family, is known for its exceptional speed and accuracy in object detection. Its ability to process images in real time makes it especially suited for medical imaging applications. By employing YOLOv8, we aimed to boost the precision and efficiency of prostate zonal segmentation, which could support faster analysis and potentially improve clinical decision-making.

The following sections provide a more detailed explanation of these methodologies and the architectural design of the networks we developed.

### Ensemble of Deep Neural Networks for Prostate Zonal Segmentation

We implemented an ensemble approach for prostate zonal segmentation by combining the predictions of three DL architectures: Attention-Res-U-Net, Vanilla-Net, and V-Net. The ensemble uses a simple averaging strategy, where the final segmentation is obtained by averaging the outputs of the individual networks.

In the following sections, we describe the architectures and key features of the networks included in the ensemble.

**Attention-Res-U-Net (Att-R-Net)**

Att-R-Net is an advanced neural network architecture specifically designed for medical image segmentation tasks. This model combines the advantages of Residual Blocks and Attention Gates to improve segmentation accuracy and robustness. Residual Blocks help address the challenge of vanishing gradients in deep networks, while Attention Gates enhance the model's ability to focus on relevant areas of the image. By incorporating these elements, Att-R-Net not only maintains high resolution in its feature maps across different layers but also selectively highlights important anatomical structures, minimizing the influence of irrelevant information. This approach leads to more accurate and reliable segmentation results. In the following, we provide further insight into how Residual Blocks and Attention Gates work: Residual Blocks: Researchers have suggested that increasing the depth of neural networks can improve performance, but deeper networks often face significant issues like vanishing gradients [181]. To solve this, He et al. [182] introduced a deep residual learning framework that uses identity mapping, making it easier to train deeper models. Alongside this, Ronneberger et al. developed the UNet architecture, which combines features from multiple layers to enhance segmentation performance. UNet has become a cornerstone in biomedical image segmentation due to its ability to merge low-level and high-level features. Building on this, Zhang et al. [181] created ResUNet, a deeper version of UNet that integrates the benefits of residual learning with UNet's architecture [45]. Residual blocks aid in training deeper networks by simplifying the process, and the model's skip connections allow efficient information flow without disrupting the network's

structure, leading to better performance in semantic segmentation tasks and reducing the need for excessive parameters [181]. Attention Gates: Attention gates are widely used in various fields, such as Natural Language Processing, image analysis, and knowledge graphs, and can be classified into two types: soft attention and hard attention [183]. Hard attention methods, like cyclic region classifiers and pruning, are typically non-differentiable and often rely on reinforcement learning to adjust parameters, which can make training difficult [184]. In contrast, soft attention methods are probabilistic and use standard backpropagation, which is easier to train. For example, additive soft attention is employed in tasks like sentence rephrasing, and recently in image classification [183]. In the deep layers of the network, the feature representation becomes more comprehensive, but using cascaded convolutions and nonlinear activation functions can lead to a loss of spatial information in the high-level output maps [185]. This can make it harder to reduce false detections, particularly for small lesions that vary greatly in size and shape. Figure 4.1 illustrates the architecture of our proposed Att-R-Net.

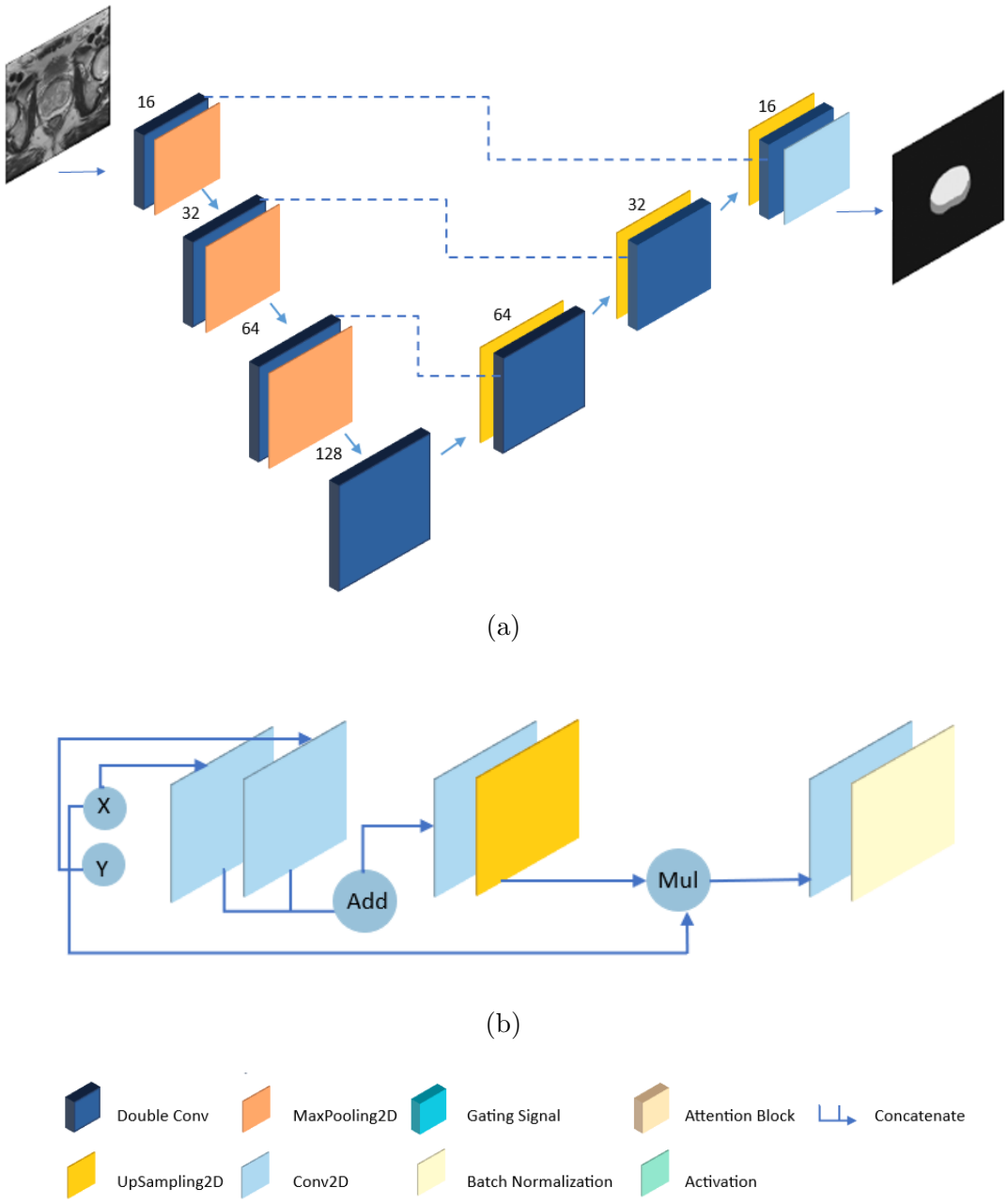



Figure 4.1: *The architecture and details of the Att-R-Net neural network. (a) Overall network, and (b) the Attention Block architecture. The Double Conv contains two convolution layers with ReLU activation, and the Gating Signal contains one convolution layer with ReLU activation.*

**Vanilla-Net**

Vanilla-Net is distinguished by its minimalist approach to neural network architecture. Unlike traditional models, it avoids deep layers, shortcuts, and complex operations such as self-attention. This design focuses on simplicity and robustness, with each layer optimized for compactness. To maintain its original structure, nonlinear activation functions are removed after training. This minimalist design effectively reduces the complexity typically associated with conventional networks, making Vanilla-Net ideal for environments where resources are limited. Despite its simplicity, the model delivers strong performance, performing on par with well-established deep neural networks and vision transformers. This demonstrates the potential of minimalism in DL. Vanilla-Net's innovative approach could revolutionize the field and challenge conventional models, setting a new benchmark for elegant and efficient network design [186]. The architecture of Vanilla-Net is shown in Figure 4.2.

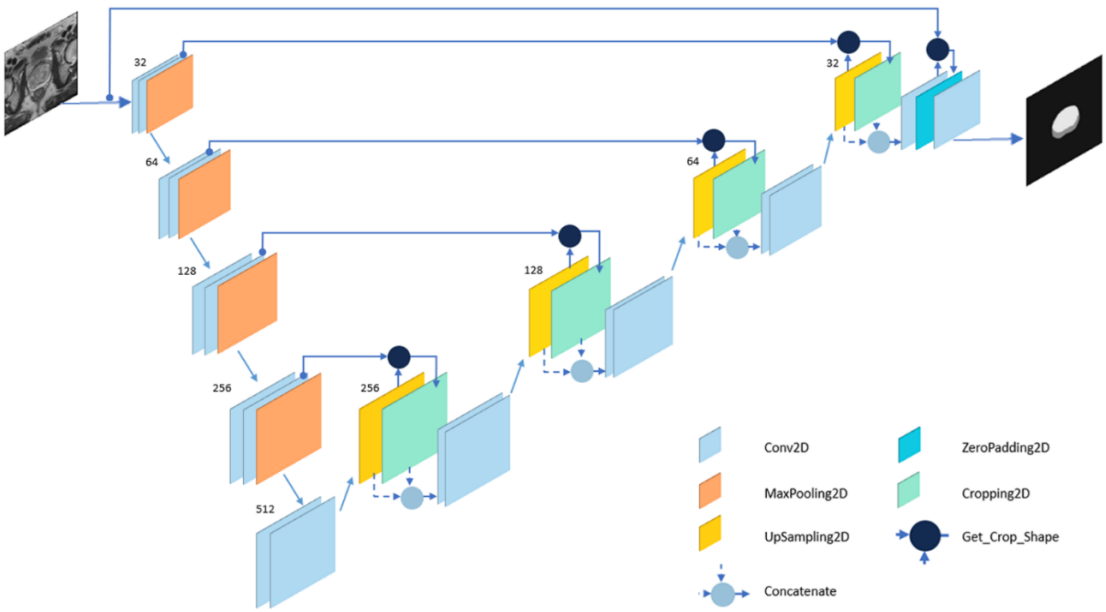


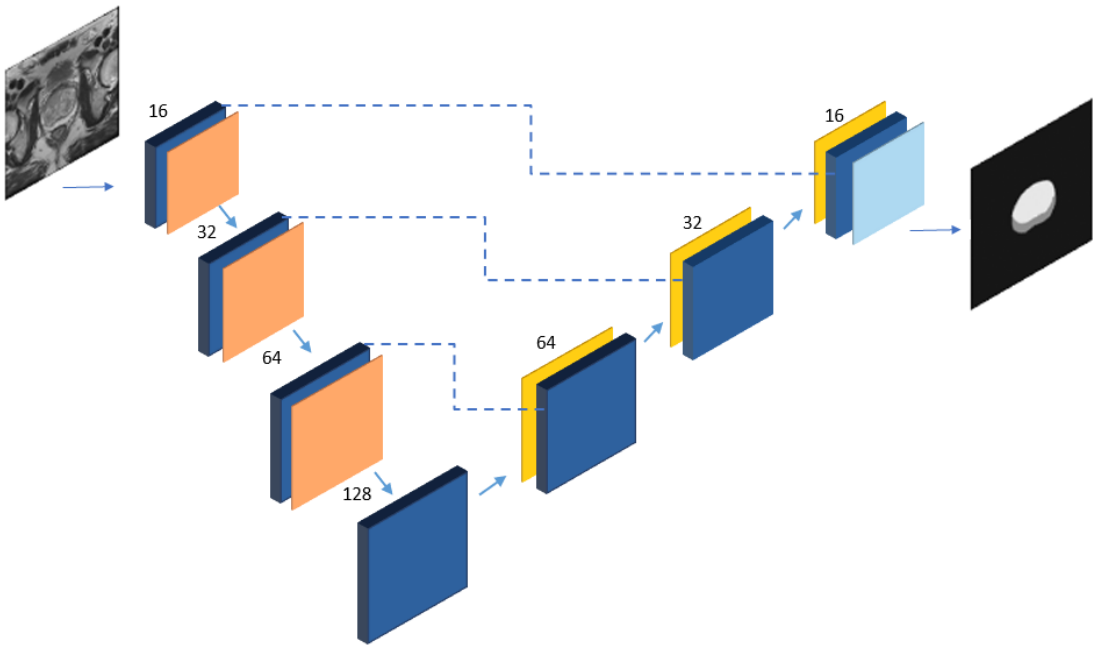

Figure 4.2: *The architecture and details of the Vanilla-Net neural network*

The get-crop-shape function plays an essential role in the U-Net architecture

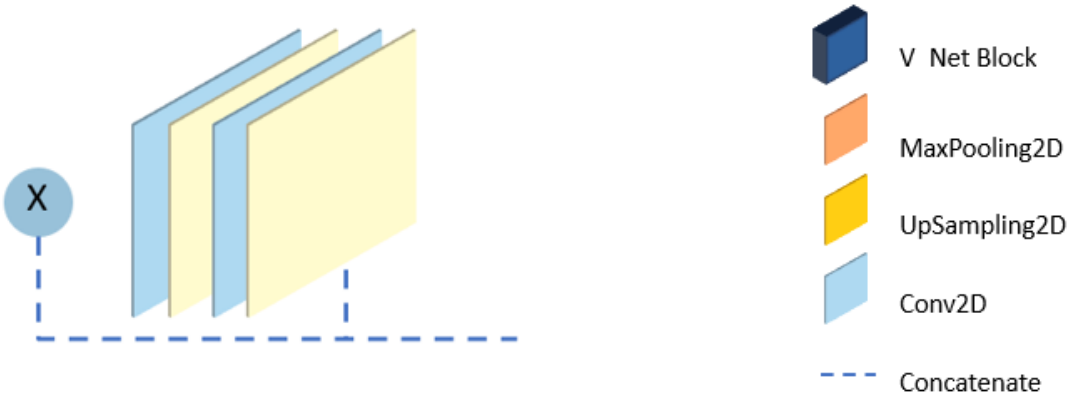
by ensuring proper alignment between the encoder and decoder feature maps during the up-sampling process. It computes the necessary cropping dimensions by comparing the spatial sizes of the feature map from the encoder with the reference map from the decoder. By correcting differences in width and height, the function ensures the feature maps are aligned accurately during concatenation. This alignment helps preserve spatial information, thereby improving the model's performance in segmentation tasks [93].

**V-Net**

V-Net, initially designed for 3D volumetric data segmentation, has demonstrated considerable potential for adapting its architecture to 2D image segmentation tasks. While it is primarily associated with medical imaging, the underlying principles of V-Net can be effectively applied to segment a wide range of 2D images, particularly those requiring precise boundary delineation. The V-Net architecture is inspired by the well-known U-Net model, incorporating an encoder-decoder structure with skip connections. These skip connections are essential for preserving high-resolution features, which are crucial for achieving accurate segmentation results [93]. In its original version, V-Net uses 3D convolutions tailored for volumetric medical image segmentation [187]. However, by adapting it to 2D convolutional layers, V-Net can now be applied to conventional image datasets, efficiently capturing spatial hierarchies in 2D images. The model's use of a Dice coefficient-based loss function is especially beneficial for addressing class imbalances, such as in tumor segmentation or small organ detection [188]. Furthermore, V-Net's architecture is optimized for fast training and inference, making it well-suited for real-time applications. This efficiency, coupled with its strong performance, enables V-Net to tackle a variety of 2D segmentation tasks. Its versatility extends beyond medical imaging, making it suitable for applications like remote sensing and autonomous driving, where it can be used to identify land cover types in satellite images or segment objects in driving scenarios. The architecture of the proposed V-Net neural network is illustrated in Figure 4.3.




(a)




(b)

Figure 4.3: *The architecture and details of the V-Net neural network, a) The overall network, and b) the architecture of the V-Net Block.*

### Meta-Net

MetaNet, based on the Theory of Independent Judges (TIJ), is applied to the field of substance use and misuse. In the TIJ framework, each artificial neural network Artificial Neural Network (ANN) acts as an expert judge for a specific problem,

with its reliability determined by its performance during testing and validation. For classification tasks involving one of N categories, each judge-artificial neural network has different levels of reliability for different aspects of the problem, with each aspect corresponding to one of the N input classes. This reliability is captured in the Confusion Matrix (CM), which shows how well the network performed during testing. According to TIJ, by combining the outputs of M judge-artificial neural networks and their corresponding M Confusion Matrices, a MetaNet can be created. The MetaNet takes the combined outputs from all M judges and produces the N classes as its output. A weight matrix is formed through an algorithm that processes the judges' CMs, and signal propagation follows a cooperative and competitive feed-forward process (see Figure 4.4) [189].

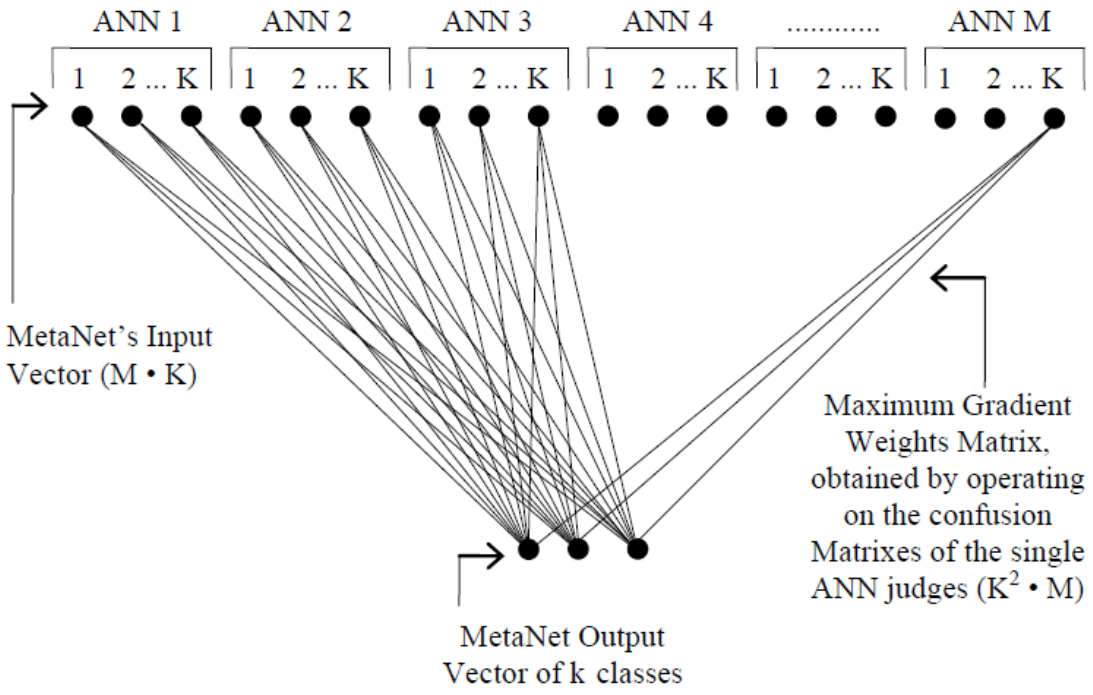



Figure 4.4: *The structure of Meta-Net* [189].

The purpose of MetaNet's Weights Matrix is to establish the local credibility of each judge about the classification task. For an artificial neural network to be

included in the pool of MetaNet judges, it must demonstrate a performance history (curriculum) relevant to the specific problem. In this context, the Confusion Matrix (CM), which compares correct classifications with predicted classifications generated by each artificial neural network during the testing phase, serves as an important curriculum. The CM of each judge-ANN is interpreted as shown in Figure 4.5.

| | | *Output* $_j$ | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | ANN 1 | A | B | C | ..... | Z | *Null* | *Tot. Row* |
| | A | $X_{1,1}$ | $X_{1,2}$ | $X_{1,3}$ | ..... | $X_{1,Z}$ | $X_{1,Z+1}$ | $\dot{R}_1$ |
| | B | $X_{2,1}$ | $X_{2,2}$ | $X_{2,3}$ | ..... | $X_{2,Z}$ | $X_{2,Z+1}$ | $\dot{R}_2$ |
| *Target* $_i$ | C | $X_{3,1}$ | $X_{3,2}$ | $X_{3,3}$ | ..... | $X_{3,Z}$ | $X_{3,Z+1}$ | $\dot{R}_3$ |
| | ..... | ..... | ..... | ..... | ..... | ..... | ..... | ..... |
| | Z | $X_{Z,1}$ | $X_{Z,2}$ | $X_{Z,3}$ | ..... | $X_{Z,Z}$ | $X_{Z,Z+1}$ | $\dot{R}_Z$ |
| | *Tot. Column* | $\dot{C}_1$ | $\dot{C}_2$ | $\dot{C}_3$ | ..... | $\dot{C}_Z$ | $\dot{C}_{Z+1}$ | |

Figure 4.5: *Confusion Matrix of each judge-ANN*

From which:

$$\text{Successes } (S) = \frac{X_{ii}}{R_i}$$

$$\text{Failed Blows } (B) = \frac{\sum_{i \neq j}^{Z+1} (X_{ij} - X_{ii})}{R_i}$$

$$\text{False Attributions } (F) = \frac{\sum_{i \neq j}^{Z+1} (X_{ij} - X_{ii})}{C_i}$$

$$\text{Correct Eliminations } (E) = \frac{X_{jj}}{C_j}$$

Beginning with the CM of each judge-artificial neural network, several methods can be used to calculate the Weight Matrix for MetaNet [146]. In this study, the weight matrix was calculated using the following formula:

$$w_{ij} = -\log_n \left( \frac{B_j F_i}{S_j E_i} \right)$$

In this formula, Sj represents the recall for class i (referred to as Successes in [190]), while Bj is calculated as 1 minus S for class i (called Failed Blows). Ei stands for the precision for class j (known as Correct Eliminations), and Fi is 1 minus E for class j (known as False Attributions). In earlier research, MetaNet was used only for classification tasks, where each judge-artificial neural network helped decide a single final label from a group of categories. However, in our study, we expand MetaNet's use to segmentation tasks, where the goal is to classify each pixel separately instead of labeling the whole image. This changes the problem from overall classification to making many small decisions at the pixel level, which requires a more detailed way of measuring each model's reliability. For this work, we used a version of MetaNet called MetaBayes.

### YOLO-v8 Architecture

YOLO was first introduced to computer vision in 2015 through a paper by Joseph Redmon and colleagues called "You Only Look Once: Unified, Real-Time Object Detection." [191]. This approach changed object detection by treating it as a simple regression task, where the system directly predicts bounding boxes and class labels from image pixels. Its unified framework allowed it to detect multiple objects at once with both high speed and accuracy. Since its release in 2016, the YOLO family of models has evolved rapidly. While Joseph Redmon stopped working on YOLO after version 3 [192], other researchers have continued to improve the model, leading up to the most recent version, YOLOv10, released in 2024 [193]. In this study, we used YOLOv8 to segment the zones of the prostate. YOLOv8

was introduced in January 2023 by Ultralytics, the same team that developed YOLOv5. Although the official research paper is still pending, early tests show that YOLOv8 outperforms its previous versions and sets a new benchmark in the YOLO series [194]. YOLOv8 has reached state-of-the-art performance by upgrading its architecture, supporting both anchor-based and anchor-free methods, and using extensive data augmentation techniques. It is capable of handling tasks like object detection, instance segmentation, and image classification, making it highly flexible for different applications. The model includes several improvements, such as a redesigned structure, new handling of anchor boxes, and an updated loss function, all of which contribute to better accuracy and faster performance in segmentation tasks [195]. Figure 4.6 illustrates the architecture of the YOLOv8 model.

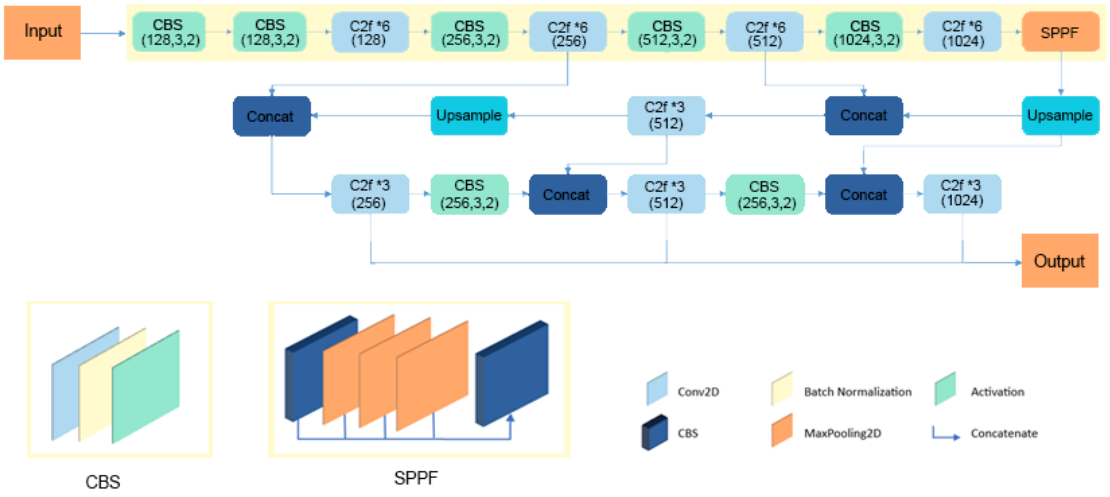



Figure 4.6: *The diagram of the YOLOv8 network structure (figure by the authors) illustrates that the CBS component consists of convolution, batch normalization, and SiLu activation functions. Additionally, the SPPF is built from three tiers of Maxpooling integrated with two CBS units, as in [1]*

### 4.1.3 Implementation Details for Zonal Segmentation

The study presents three distinct methods for segmenting prostate zones (central gland and peripheral zone) using T2W images. A total of 3,553 T2W slices from 139 patients were used, sourced from the publicly available Prostate158 dataset.

For model development and evaluation, the dataset was split into three subsets: 2,306 slices from 90 patients for training, 637 slices from 25 patients for validation, and 610 slices from 24 patients for testing. To standardize image intensity across the dataset, we applied a custom normalization technique based on intensity windowing. Each image was loaded using SimpleITK, and its mean and standard deviation were calculated. Pixel values were then rescaled to fall between one standard deviation below and two standard deviations above the mean, with the resulting range mapped to a 0–255 scale. This preprocessing step helped ensure consistent contrast levels across images, improving the effectiveness of downstream segmentation. The first segmentation method used an average ensemble of three models: Att-R-Net, Vanilla-Net, and V-Net. These models were trained end-to-end using the Adam optimizer, with the following hyperparameters: an initial learning rate of $1 \times 10^{-4}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 1 \times 10^{-8}$. Training was performed with a batch size of 8 for up to 300 epochs. Early stopping was applied after 50 epochs without improvement in validation loss. Each input image was resized to 144×144 pixels, and pixel intensities were normalized to zero mean and unit variance. Data augmentation techniques were used during training to prevent overfitting. Since the loss function plays a key role in how a neural network learns, we carefully selected one suitable for segmentation tasks. For training, we used the fixed focal loss, which is especially useful for handling class imbalance by down-weighting well-classified examples and focusing learning on harder, misclassified pixels. The formula for the fixed focal loss is provided below.

$$L_{\text{focal}}(y_{\text{true}}, y_{\text{pred}}) = -\alpha\,(1 - p_{\text{true}})^{\gamma}\,\log(p_{\text{true}} + \epsilon) - (1 - \alpha)\,(p_{\text{false}})^{\gamma}\,\log(1 - p_{\text{false}} + \epsilon)$$

Where $\alpha$ is the balancing factor and is set to 0.25, $\gamma$ is the focusing parameter and is set to 2, and $\epsilon$ is a small constant added for numerical stability to avoid taking the logarithm of zero.

In this formula, $y_{\text{true}}$ is the original mask and $y_{\text{pred}}$ is the predicted mask obtained from our models. Furthermore, $p_{\text{true}}$ and $p_{\text{false}}$ are defined based on the true labels $y_{\text{true}}$:

$$\text{If } y_{\text{true}} = 1, \quad p_{\text{true}} = y_{\text{pred}}, \quad p_{\text{false}} = 1 - y_{\text{pred}}.$$

$$\text{If } y_{\text{true}} = 0, \quad p_{\text{true}} = 1 - y_{\text{pred}}, \quad p_{\text{false}} = y_{\text{pred}}.$$

For the YOLO-based segmentation approach, we utilized the official GitHub repository provided by Ultralytics [196]. To enable training, the ground truth segmentation masks were converted into the "YOLO format," which involves encoding object classes along with normalized bounding box coordinates and associated segmentation data. We employed the YOLOv8-nano segmentation model, a lightweight and efficient version tailored for high-speed inference with minimal computational demand. Training was conducted using a batch size of 16, an image resolution of 256×256 pixels, and a learning rate of 0.01. This configuration was selected to balance training speed and performance while ensuring compatibility with hardware limitations. The YOLO-v8 model's compact design made it particularly well-suited for rapid deployment in clinical imaging environments where both speed and accuracy are critical.

## 4.2 Prostate Lesion Segmentation

Accurate segmentation of prostate lesions on MRI is a crucial step in improving the detection, diagnosis, and treatment planning of prostate cancer. Segmentation refers to delineating the boundaries of both the prostate gland and suspicious lesion regions, or "intraprostatic lesions" (ILs), from MRI scans. This process enables the quantification of lesion size, shape and location — key factors for assessing risk (e.g. via PI-RADS or Gleason scores), guiding biopsies, and planning interventions such as focal therapies or radiotherapy boost. Traditionally, lesion segmentation has been done manually by radiologists, but this is time-consuming and subject to inter-observer variability. DL-based methods have been increasingly used to automate segmentation from MRI, using architectures such as U-Net, Mask R-convolution neural network, cascaded convolutional networks, and

attention mechanisms, often with multimodal inputs like T2W images plus ADC or DWI. For example, Dai et al. proposed a non-local Mask R-convolution neural network approach that uses histopathological ground truth and showed improved segmentation accuracy of ILs compared to baseline methods (Mask R-convolution neural network, 3D U-Net) [132]. Another work by Ren et al. combined dense blocks, attention mechanisms, and spatial pyramidal pooling to segment both the prostate gland and lesion regions with high Dice and IoU scores [197].

Despite these advances, lesion segmentation remains challenging. Lesion boundaries can be ambiguous, particularly in the transition zone or for small lesions, and MRI image quality varies across scanners, sequences, and patients. Also, the ground truth for lesions is often limited (e.g., only certain slices or only the dominant lesion) and can vary between experts, contributing to evaluation variability. Studies show that even among radiologists, agreement (e.g., via Dice similarity) for manual segmentations of lesions is moderate at best, which sets a realistic ceiling for what automated methods can achieve [198]. Moreover, many published models are trained on relatively small or institution-specific datasets, which can limit generalizability. Because of this, robustness (cross-site performance), clarity of how datasets are selected, and external validation are important factors to consider when developing segmentation algorithms [9].

### 4.2.1 Data Acquisition

In this study, a specialized dataset was created in-house by systematically collecting and processing raw multi-parametric MRI scans. The final dataset includes a total of 311 MRI cases, with 58 labeled as PI-RADS 3 and 253 as PI-RADS 4 or 5, by the PI-RADS criteria. The original imaging data were acquired at the Department of Diagnostic Imaging and Stereotactic Radiosurgery at Centro Diagnostico Italiano S.p.A. (CDI), located on Via S. Saint Bon 20, Milan 20147, Italy. This curated dataset provides a solid foundation for the development and evaluation of ML models aimed at enhancing the detection and localization of clinically significant prostate lesions.

### 4.2.2 Proposed Neural Networks for Lesion Segmentation

This study investigated three DL models for segmenting prostate lesions: U-Net, Dense U-Net, and Attention U-Net. Each architecture incorporates distinct design principles to better extract spatial and contextual information from mp-MRI scans, aiming to enhance segmentation performance. The models were trained using three key MRI sequences. To assess the impact of different input strategies, two training approaches were employed: (1) training each model separately with individual modalities (T2W, ADC, or DWI), and (2) applying a multi-branch or multi-input framework based on U-Net, where all three sequences were integrated and processed jointly, as illustrated in Figure 4.7.

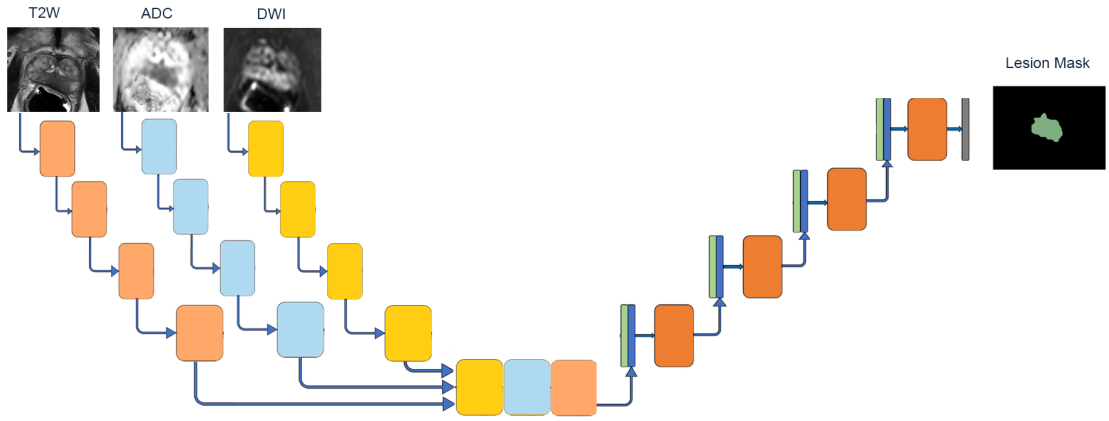



Figure 4.7: *The overall architecture of the multi-encoder U-Net-Shaped networks for prostate lesion segmentation*

As the original dataset included separate lesion masks for each MRI modality, a unified ground truth was needed for the multi-input training approach. To create this, the individual binary masks from the T2W, ADC, and DWI sequences were combined using a pixel-wise logical OR operation. This ensured that any voxel identified as part of a lesion in at least one modality was included in the final segmentation mask. By integrating information across modalities, this method captured the full extent of lesion regions, leveraging the complementary strengths of each sequence. The resulting combined mask was converted to the float32 format to meet the input requirements of the neural networks. Figure 4.8 illustrates

examples of these unified segmentation masks across cases with different PI-RADS scores.

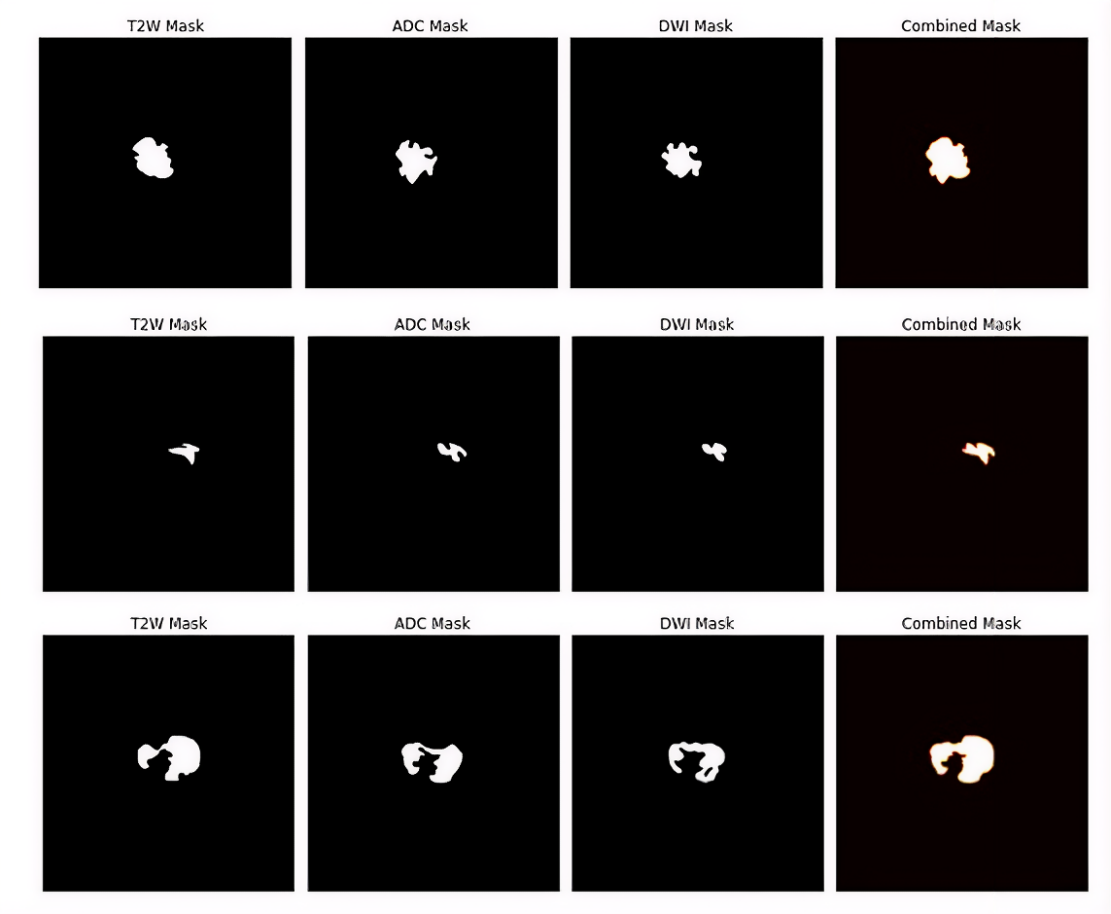



Figure 4.8: *Representative ground truth lesion masks for T2W, ADC, and DWI modalities, and their corresponding merged mask*

To evaluate how well the models generalize to different lesion types, we designed training experiments using two distinct subsets of the dataset: one comprising all cases with PI-RADS scores of 3, 4, and 5, and another limited to cases with PI-RADS 4 and 5 only. The latter subset is of greater clinical importance, as lesions in these categories are more likely to correspond to significant prostate cancer. This comparative setup enabled a more targeted assessment of the models' effectiveness in identifying high-risk lesions. The following sections offer an in-depth overview of the model architectures, training methodologies, and the rationale behind our

data selection and segmentation mask fusion strategies.

### U-net

The U-Net model consists primarily of two components: an encoder and a decoder. The central concept behind U-Net is to extract meaningful features from input images by progressively reducing their spatial dimensions through the encoder, thereby capturing increasingly abstract representations. This down-sampling process enables the network to learn deep, hierarchical features. To preserve spatial context and fine-grained details, U-Net incorporates skip connections that directly link corresponding layers in the encoder and decoder. These connections allow the decoder to integrate both low-level and high-level features, improving the accuracy of the reconstructed segmentation map. During the encoding phase, the network applies successive convolutional layers to extract features at multiple levels, with each layer learning increasingly complex representations. In the decoding phase, the network employs up-sampling layers to gradually reconstruct the spatial resolution, ultimately producing a dense segmentation mask. The quality of feature extraction and segmentation is influenced by the training of convolutional filters, the design of the down-sampling and up-sampling operations, and the effectiveness of the skip connections. Furthermore, the backbone architecture, responsible for defining the arrangement of layers, plays a crucial role in structuring both the encoding and decoding pathways of the network [199].

### Dense-Unet

Dense-UNet is an advanced extension of the U-Net architecture that incorporates dense concatenation to improve feature propagation and network efficiency. The architecture consists of a dense downsampling path and a dense upsampling path, which are symmetrically arranged and interconnected through multiple skip connection channels. In the downsampling path, convolutional operations are employed to capture multi-scale semantic and contextual features. To address the depth limitations of the traditional U-Net, standard pooling and convolution layers are replaced with dense blocks and transition blocks. This design allows the

network to achieve greater depth without compromising on computational efficiency or feature extraction quality. Each dense block is composed of layers that are directly connected to all preceding layers in a feed-forward manner, enabling extensive feature reuse. This means that the input of each layer includes the outputs of all previous layers, a strategy referred to as dense concatenation. For this dense connectivity to function properly, the feature maps at each layer must have consistent spatial dimensions. As a result, Dense-UNet facilitates deeper architectures with enhanced feature learning capacity, making it well-suited for complex image segmentation tasks [200].

### Attention U-Net

Attention U-Net is a specialized neural architecture developed for medical image segmentation, incorporating attention mechanisms that allow the model to selectively emphasize relevant regions within the input data. By focusing computational resources on informative features, these attention gates enhance both the accuracy and consistency of segmentation results [201]. Attention Mechanism: Attention gates have been extensively adopted across various domains, including natural language processing, computer vision, and knowledge representation. Broadly, they can be categorized into hard and soft attention [183]. Hard attention approaches, such as cyclic classifiers and pruning, are non-differentiable and typically require reinforcement learning, which adds complexity to the training process [40]. In contrast, soft attention is differentiable and compatible with standard backpropagation techniques, eliminating the need for methods like Monte Carlo sampling. Additive soft attention, in particular, is frequently used in applications like sentence generation and image classification [183]. In deep convolutional networks, while the encoder captures increasingly abstract features, repeated convolutions and non-linear activations may compromise spatial resolution [184]. This loss of detail can hinder the detection of small or irregularly shaped lesions. The integration of attention gates addresses this issue by preserving critical spatial information, thereby improving the model's sensitivity to subtle and heterogeneous lesion patterns.

**LSTM U-Net**

In this study, we developed a U-Net–based architecture for prostate lesion segmentation in which a Bidirectional ConvLSTM (Bi-ConvLSTM) was incorporated at the bottleneck [202]. This design choice was motivated by recent studies demonstrating the effectiveness of combining recurrent modules with attention mechanisms to capture contextual information in medical images. The Bi-ConvLSTM allows the network to aggregate multi-scale spatial features from the encoder while preserving fine structural details, improving segmentation performance on small and irregular lesions. By modeling spatial dependencies within the feature maps, the network can better capture relationships between different regions of the prostate. This approach follows the strategies proposed in the Residual Attention U-Net framework for skin lesion segmentation [203] and the Multi-scale Attention U-Net [204], where attention mechanisms and recurrent blocks are employed at the bottleneck to enhance feature representation without significantly increasing computational cost. Overall, the integration of convolutional, recurrent, and attention-based modules provides a more expressive and robust feature representation, contributing to improved segmentation accuracy on challenging prostate MRI datasets.

### 4.2.3 Implementation Details for Prostate Lesion Segmentation

The main phases of lesion segmentation, depicted in Figure 4.9, are as follows: 1. Data Collection: We gathered 311 multi-parametric MRI (mpMRI) scans from the CDI data repository, covering the period from 2023 to 2025. These datasets encompass various imaging modalities crucial for the diagnosis and evaluation of prostate cancer.

2. Anonymization and Format Conversion: To comply with privacy regulations, all original DICOM files underwent anonymization. Using custom Python scripts, we extracted three essential MRI sequences for prostate cancer analysis: T2W, ADC, and DWI. These sequences were then converted into NIfTI format

via Python tools such as dicom2nifti. In particular, DWI scans with a b-value of 1600 were selected due to their improved lesion contrast and ability to distinguish malignant from benign tissues [205] [206]. Corresponding ADC maps were also obtained for these DWI images.

3. Image Registration and Alignment: To bring T2W, ADC, and DWI images into spatial correspondence, we employed the SimpleITK library. Since these sequences differ in resolution, orientation, and may be affected by patient movement, preprocessing and spatial transformation were necessary [28].

• Preprocessing: Images were standardized in format and intensity ranges were normalized to enable meaningful comparison [207].

• Registration: T2W images served as the fixed reference, with ADC and DWI scans aligned to this space using rigid and affine transformations optimized via mutual information metrics.

• Multi-resolution Approach: To improve accuracy, registration was performed progressively at different scales, refining the alignment step-by-step [208].

• Resampling: Finally, all images were resampled to match the reference resolution and dimensions, ensuring spatial consistency across modalities. This automated pipeline effectively minimized misalignments, laying the groundwork for precise segmentation.

4. Lesion Annotation: Initial lesion segmentation was carried out using 3D Slicer [209] by a doctoral candidate trained under a certified radiologist. Lesions were delineated across all three MRI sequences to capture comprehensive lesion characteristics.

5. Expert Validation: Two experienced radiologists independently reviewed and refined these annotations, resulting in high-quality segmentations that served as the reference standard for subsequent model evaluation.

6. Image Resizing: To maintain uniform input size for neural network training, all images were resized to 256×256 pixels.

7. Center Cropping: Focusing on the prostate region, typically centralized within pelvic scans, images were cropped to 128×128 pixels. This step reduces

irrelevant background information and enhances the model's ability to detect relevant features [82].

8. Intensity Normalization: Given the variability of MRI intensity values across scanners and sequences, tailored normalization methods were applied:

• T2W and ADC images were normalized using Min-Max scaling to map intensities between 0 and 1.

• DWI images underwent Z-score normalization, standardizing intensities by subtracting the mean and dividing by the standard deviation.

These normalization strategies have been demonstrated to enhance training stability and improve model accuracy in multi-modal imaging tasks [210].

9. Model Training: Various U-Net-based neural networks were trained for prostate lesion segmentation. Multiple preprocessing and training setups were explored to assess their impact on segmentation performance. Subsequent sections delve deeper into select steps of this pipeline, offering detailed technical insights and justification for key methodological decisions.

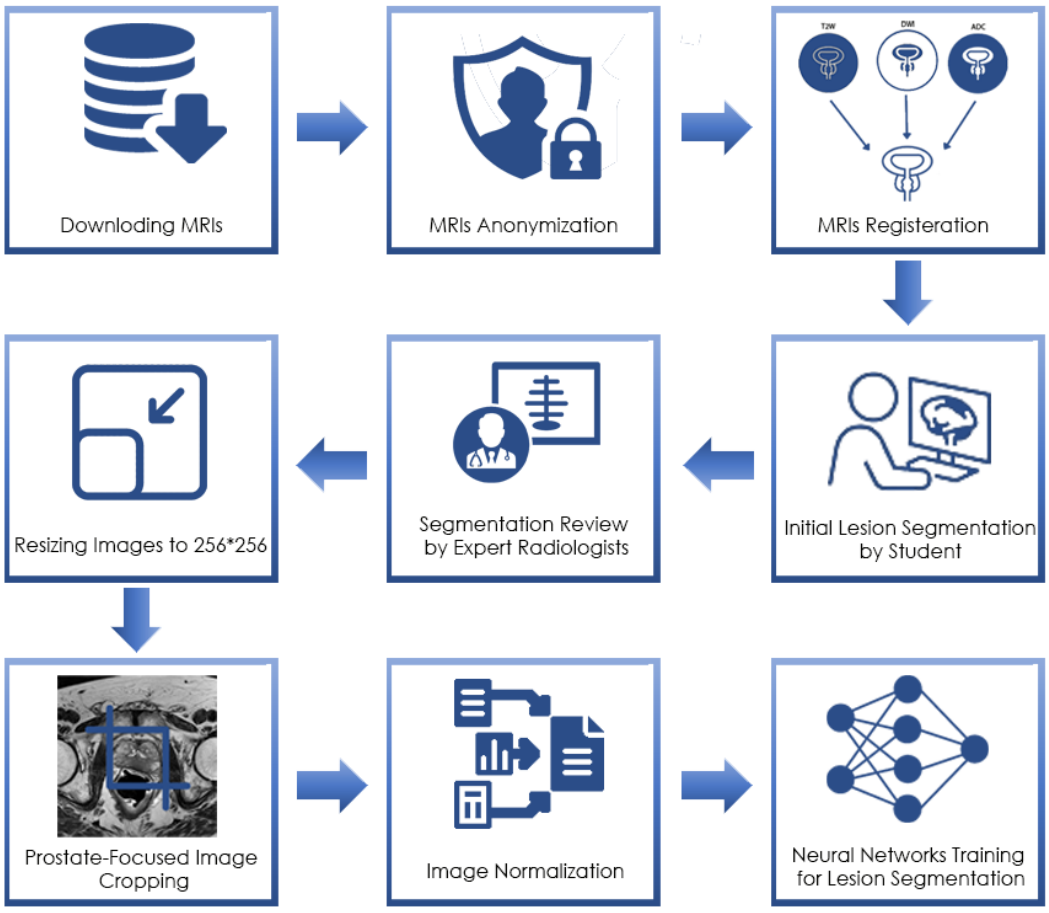



Figure 4.9: *The main steps of the research for preparing and pre-processing our in-house dataset for prostate lesion segmentation*

## 4.3 PI-RADS Classification

The PI-RADS was developed to standardize and streamline the interpretation of prostate MRI. Currently in its latest revision, PI-RADS v2.1, the system assigns scores ranging from 1 to 5 to indicate the probability that a lesion represents clinically significant prostate cancer, based on assessments from T2W imaging (T2WI), DWI, and DCE sequences [11]. This structured framework enhances clarity in radiology reports and assists urologists in making informed decisions regarding biopsy or active surveillance [211].

Although PI-RADS has significantly improved consistency and communication among radiologists, its application remains largely reliant on expert visual interpretation, which can be subjective and prone to variability. Consequently, there is increasing interest in CAD systems and AI-driven approaches that leverage quantitative features of mp-MRI to support or automate PI-RADS scoring [212].

For PI-RADS classification, we applied three complementary strategies, each designed to extract meaningful features from mp-MRI data. The first strategy involved hand-crafted radiomic feature extraction from manually segmented lesions using the PyRadiomics library, capturing intensity, shape, and texture characteristics across multiple MRI sequences. The second strategy built upon this by incorporating fully automated lesion and prostate zone segmentations, creating a workflow that more closely resembles clinical practice without relying on manual input. In the third strategy, we developed a custom convolution neural network to automatically learn and extract high-level features from ADC images and their associated lesion masks. Features derived from all three approaches were subsequently used to train ML models for multi-class PI-RADS classification, enabling a comparative evaluation of manual, automated, and deep-learning-based feature extraction methods.

### 4.3.1 Extracting PyRadiomic Features (Approach 1)

In this study, radiomic features were derived from mp-MRI scans using PyRadiomics, an open-source Python library specifically developed to enable standardized and reproducible radiomics analyses [65]. Radiomics is a computational technique that converts medical images into high-dimensional, quantitative data by characterizing patterns, textures, shapes, and intensity distributions within defined Regions of Interest [60]. Through this process, it is possible to extract hundreds or even thousands of features that capture subtle variations in tissue heterogeneity, many of which remain imperceptible to the human eye [213].

Feature extraction was performed from clinically relevant MRI sequences, including T2W, ADC, and DWI. Two separate configurations were implemented. In the first configuration (Feature Set 1), features were extracted from the following

image types: Original, Laplacian of Gaussian (LoG; $\sigma$ = 2.0, 3.0, 4.0, and 5.0), Wavelet, Square, SquareRoot, and Logarithm. From each of these filtered images, a comprehensive set of feature categories was calculated, including first-order intensity features, shape-based descriptors, and texture-based metrics derived from the GLCM, GLRLM, GLSZM, Gray Level Dependence Matrix (GLDM), and Neighboring Gray Tone Difference Matrix (NGTDM). The extraction settings were standardized across all images as follows: binWidth = 25, resampledPixelSpacing = [1, 1, 1], interpolator = sitkBSpline, normalize = True, removeOutliers = 3, label = 1, correctMask = True, force2D = False, and preCrop = True.

In the second configuration (Feature Set 2), the feature set was expanded by incorporating additional image filters, namely Exponential, Gradient, Local Binary Patterns in 2D (LBP2D), and Local Binary Patterns in 3D (LBP3D), while also retaining the filters used in the first configuration (Original, LoG with an extended range from $\sigma$ = 1.0–5.0, Wavelet, Square, SquareRoot, and Logarithm). This enhanced setup enabled the extraction of a broader spectrum of intensity, shape, and texture descriptors. The same feature classes (first-order, shape, GLCM, GLRLM, GLSZM, GLDM, and NGTDM) were computed, but applied to a more diverse set of filtered images. Furthermore, pre-processing included normalizeScale = 100 to ensure consistency between scanners and acquisition protocols.

The extracted radiomic features were used to classify prostate lesions into PI-RADS categories 3, 4, and 5, thereby providing quantitative support for the radiological scoring system. Beyond image-derived features, the anatomical location of each lesion was incorporated as an additional input variable. Specifically, lesions were categorized according to their location in the peripheral zone, central gland, or both zones simultaneously, and were numerically encoded as 1 (peripheral zone), 2 (central gland), or 3 (peripheral zone+central gland). This anatomical context enriched the feature set with zone-specific information, which is clinically relevant as lesion appearance and diagnostic interpretation in PI-RADS are strongly influenced by zonal location [214].

By comparing results across both configurations, the study aimed to determine

whether the inclusion of advanced filters and descriptors improved PI-RADS classification performance. In particular, the second configuration was expected to capture more complex and discriminative image patterns, especially through the use of LBP and Gradient filters, which have been shown to improve classification accuracy in prior radiomics studies [215].

Finally, all extracted features were employed to train a set of ML models for multi-class classification of PI-RADS scores (3, 4, and 5). Model performance was evaluated using an independent test set consisting of unseen cases. For each sample, both the lesion masks and the corresponding anatomical images were available, ensuring accurate region-specific feature extraction and validation of the models' predictive capabilities.

### 4.3.2 Pyradiomics Feature Extraction from Automated Zone and Lesion Segmentations (Approach 2)

In the first approach, radiomic features were extracted from manually segmented lesions, with the lesion's anatomical zone incorporated as an additional variable to improve PI-RADS classification. This strategy relied on the availability of reference lesion masks and zone labels within the test dataset, which were utilized for both feature extraction and model evaluation.

The second approach was developed to more closely reflect a real-world clinical setting by employing a fully automated pipeline based on previously established segmentation models. First, a zonal segmentation model, published in prior work [201], was applied to T2W images to automatically delineate prostate zones. Subsequently, a lesion segmentation model, trained on ADC images from the same dataset, was used to identify lesions automatically. To objectively assign a zonal label to each lesion, a custom Python script was implemented to calculate the spatial overlap between the automatically segmented lesion masks and zone masks. This ensured reproducible and accurate zone assignment without manual input.

Radiomic feature extraction in this approach followed the same procedures as in the manual workflow but was now based entirely on automated lesion and zone segmentations. Finally, the PI-RADS classification model was tested within this

end-to-end automated pipeline, enabling evaluation of its performance under conditions that simulate routine clinical practice, where both segmentation and zone determination are carried out automatically. Figure 4.10 illustrates the sequential steps of Approach 2 used for PI-RADS classification.

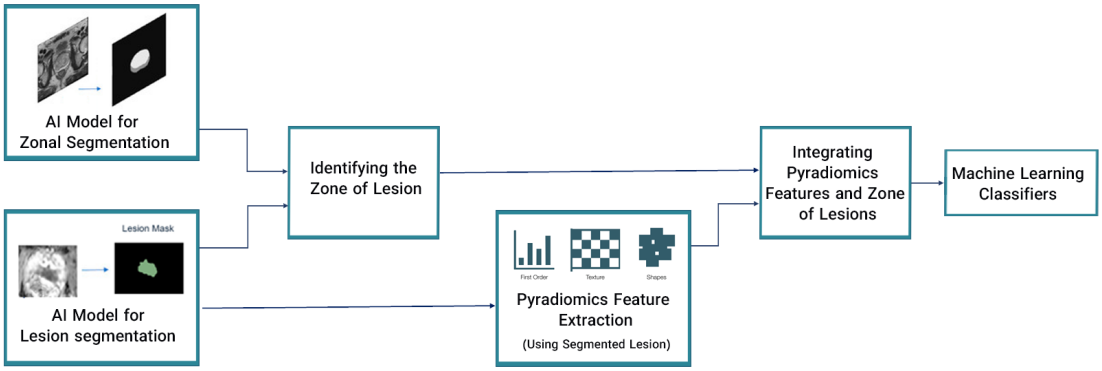



Figure 4.10: *Workflow of the Fully Automated Pipeline (Approach 2) for PI-RADS Classification*

### 4.3.3 Feature Extraction Using Proposed Convolution Neural Network (Approach 3)

In an alternative strategy, we employed a convolutional neural network for feature extraction, followed by the application of ML classifiers for PI-RADS categorization. The Convolutional neural network was trained using ADC images alongside their corresponding lesion masks, which were simultaneously provided to the network. This paired-input design enabled the model to capture both visual information and spatial context, thereby enhancing its ability to focus on the lesion-specific region [75] [92].

The convolution neural network architecture comprised three convolutional layers, each regularized with dropout to reduce overfitting. Before training, both images and masks were preprocessed by resizing to 256 $\times$ 256 pixels, applying center cropping to isolate the prostate region, and subsequently resizing to 128 $\times$ 128 pixels. To further improve generalizability, data augmentation was applied, including random rotations, horizontal and vertical flips, and small translations to

simulate variability in patient positioning and imaging conditions. The network was trained for 200 epochs with a learning rate of 1e-4 and a batch size of 32. Once trained, features were extracted from the final dense layer and subsequently used as inputs for the ML classifiers to perform PI-RADS classification.

### 4.3.4 Implementation Details for PI-RADS Classification

After extracting features in different approaches, all classifiers were trained using a 5-fold cross-validation strategy [216] to provide a reliable estimate of performance. Before training, the features were standardized with StandardScaler [217], which normalizes data by removing the mean and scaling to unit variance, an essential step for distance-based algorithms such as SVM and k-nearest neighbor k-Nearest Neighbor (KNN). To address dimensionality, we employed the Least Absolute Shrinkage and Selection Operator (LASSO) feature selection [218], which eliminates less informative features through penalization. The classifiers were trained both with and without feature selection, with LASSO consistently leading to improved results.

## 4.4 Prostate Cancer Detection

In prostate cancer diagnostics, mp-MRI has become a standard tool for identifying suspicious lesions and assigning PI-RADS scores, which guide clinical decision-making. Patients with higher PI-RADS scores (typically $\geq 3$) are often referred for targeted biopsy to confirm or exclude malignancy [11][178]. While biopsy remains the gold standard for diagnosis, it is invasive and carries potential risks such as infection, bleeding, and discomfort. Consequently, there is significant interest in developing non-invasive predictive models that can distinguish between cancer and non-cancer cases using a combination of imaging biomarkers and clinical data.

Clinical factors such as patient age, prostate-specific antigen levels, prostate volume, and prostate-specific antigen density are known to be important indicators of cancer risk [58] [219]. When integrated with quantitative imaging features extracted from MRI, these variables can improve diagnostic accuracy and reduce

unnecessary biopsies. ML and AI-based approaches are increasingly being used to combine radiological and clinical data for binary classification tasks, separating patients into cancer and non-cancer categories. Such multimodal models hold promise for improving individualized risk assessment, enhancing biopsy decision-making, and ultimately supporting precision medicine in prostate cancer management [220][221].

### 4.4.1 Data Acquisition

Data was collected from Trita Hospital, located in the 22nd district of Tehran, Iran. The images were produced using an 18-channel device with a 10-coil receiver. Biopsy results were extracted from medical reports, and additional clinical information was gathered from imaging reports, including patient age, prostate-specific antigen levels, and prostate dimensions and volume. Because prostate cancer is often difficult to diagnose due to similarities with other conditions, the expertise of a urology specialist was used to identify the relevant diagnostic factors. Prostate-specific antigen levels can indicate whether the prostate is healthy, but when combined with prostate size and dimensions, they help distinguish cancer from conditions such as prostatitis. Age is also an important factor in diagnosis. For this reason, these variables were considered together, with none regarded as more important than the others. Based on the reports, patients were assigned PI-RADS-V2 scores ranging from 1 to 5. In total, 345 patients were included in the study. These patients visited the imaging center between October 2020 and December 2021, and informed consent was obtained from all participants or their legal guardians. The imaging data included multiple MRI modalities: ADC maps, DWI, and T2W scans in axial, coronal, and sagittal views. Figure 4.11 illustrates the different MRI modalities used in this study, including ADC maps, DWI, and T2W scans in axial, coronal, and sagittal views.

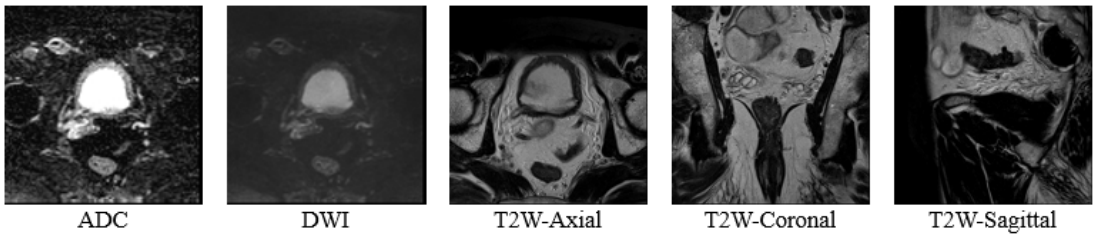



Figure 4.11: *Examples of the MRI modalities utilized for cancer classification: ADC map, DWI, and T2W scans (axial, coronal, and sagittal views)*

### 4.4.2 Proposed Neural Network for Prostate Cancer Detection

In this study, a classification approach was developed using a specialized dataset of prostate magnetic resonance imaging. For each patient, the data set contained a single slice of magnetic resonance imaging that, based on the radiologist's assessment, clearly showed the most suspicious lesion. This approach aligns with practices commonly seen in datasets designed for Gleason grade classification or prostate cancer detection, where only the slice with the most prominent lesion is selected for analysis.

A custom convolutional neural network was trained using these images. The proposed convolution neural network is designed for binary classification of prostate MRI images into cancer and non-cancer categories. The network takes grayscale images of size 64×64 pixels as input. It consists of a sequence of four convolutional blocks, each composed of a 2D convolutional layer followed by a max-pooling layer and a dropout layer. The convolutional layers progressively increase the number of filters (16, 32, 64, 128) to allow the network to learn hierarchical feature representations, from simple edges and textures in the early layers to more complex structures in deeper layers. ReLU activation is applied after each convolution to introduce non-linearity.

Max-pooling layers reduce the spatial dimensions of feature maps, providing translation invariance and reducing computational cost. Dropout layers with a rate of 0.1 are used to prevent overfitting by randomly deactivating a small fraction of

neurons during training. After the convolutional blocks, feature maps are flattened into a one-dimensional vector and normalized using batch normalization to improve training stability and convergence. A fully connected (dense) layer with 128 units serves as the main feature embedding layer, which can also be used for extracting high-level image features for further analysis or fusion with clinical data. The final layer uses a softmax activation function to output class probabilities for cancer and non-cancer predictions.

Overall, this architecture balances model complexity and regularization, allowing efficient learning of relevant imaging features while mitigating overfitting, making it suitable for integrating imaging features with clinical variables for robust prostate cancer classification.

After training the network, deep features were extracted from the final layer of the convolution neural network to represent the imaging information in a compact form. To improve diagnostic performance, these imaging features were combined with additional clinical data, including patient age, prostate-specific antigen levels, prostate volume, and prostate gland size. The integration of imaging and clinical characteristics was designed to capture both structural information from MRI and relevant clinical risk factors, thus providing a more comprehensive basis for classification.

The combined feature set was then used as input for several traditional ML classifiers, including both individual models and ensemble approaches, to categorize patients into two groups: cancer and non-cancer. Two ensemble strategies were employed to improve robustness and reduce the risk of overfitting. The first, referred to as Ensemble 1, combined XGBoost and support vector machine using Logistic Regression as a meta-learner in a stacked configuration. The second, referred to as Ensemble 2, implemented a simple voting scheme between XGBoost and support vector machine classifiers. This multimodal strategy allowed the framework to leverage complementary sources of information, enhancing the reliability of prostate cancer detection compared to using only imaging or clinical data.

### 4.4.3 Implementation Details for Prostate Cancer Detection

For the implementation of the proposed convolution neural network, all MRI images were resized to 64×64 pixels to standardize input dimensions, as the original images were of varying and relatively small sizes. Min-max normalization was applied to scale the image intensity values between 0 and 1, ensuring consistent input ranges for the network and improving convergence during training. To enhance the generalization capability of the model and reduce overfitting, several data enhancement techniques were used, including random rotations, flips, and shifts.

The network was trained using Adam Optimizer with a learning rate of $1 \times 10^{-5}$, a batch size of 64, and the categorical cross-entropy loss function. Training was performed for a maximum of 300 epochs; however, early stopping and callbacks were used to terminate training once the validation loss stopped improving, preventing overfitting. Additionally, to address the imbalance between cancer and non-cancer samples in the dataset, a weighting technique was applied during training. Clinical variables such as age, prostate-specific antigen levels, prostate volume, and gland size were scaled using the StandardScaler method to standardize their distribution and ensure compatibility when combined with the extracted imaging characteristics for subsequent ML classification.

The network was trained using 288 images (including 80 cancer and 168 non-cancer) and evaluated on a separate test set containing 72 images (24 cancer and 48 non-cancer). The test data were completely unseen during training and were not used in any preprocessing or feature extraction steps, ensuring an unbiased assessment of the model's performance. These preprocessing and training strategies, combined with careful network design, contributed to the robustness and reliability of the cancer versus non-cancer classification framework.

## 4.5 Machine Learning Classifiers Involved in the Pipeline

For the two classification tasks, PI-RADS scoring and prostate cancer detection, several ML models were employed to analyze imaging and clinical data. These models were selected to evaluate different learning strategies and to identify which approaches perform best for each task. All models were trained on the corresponding datasets and validated using appropriate cross-validation techniques to ensure reliable performance. By applying multiple ML algorithms, the study aimed to compare their effectiveness in predicting PI-RADS categories and detecting prostate cancer, providing a robust framework for automated decision support.

### Multilayer Perceptron (MLP)

A multilayer perceptron is a class of feedforward artificial neural network. It consists of an input layer, one or more hidden layers, and an output layer. Each neuron (node) in a layer is connected to neurons in the next layer. The nodes compute weighted sums of inputs, add a bias, then pass the result through an activation function (e.g., sigmoid, ReLU). The network is trained by backpropagation: errors at the output are propagated back through the network to update weights via gradient descent [222]. multilayer perceptrons are powerful for learning complex, non-linear decision boundaries [223].

### Random Forest (RF)

Random Forest is an ensemble learning method that constructs multiple decision trees during training and aggregates their predictions to improve accuracy and reduce overfitting. It is particularly robust against noise and can handle high-dimensional datasets effectively [224].

### Decision Tree

A decision tree is a supervised learning model used for classification (and regression) that recursively partitions the feature space using simple decision rules. At each internal node of the tree, a feature is selected and a threshold or categorical split is made so as to best separate the classes (according to some measure like Gini impurity, information gain, etc.). The tree grows until stopping criteria are met (such as minimum node size or maximum depth), and then pruning can be applied to avoid overfitting. One of its major advantages is interpretability: the paths from root to leaf can be read as human-interpretable rules [225].

### Support Vector Machine (SVM)

SVMs are supervised learning models that classify data by finding the optimal hyperplane that maximizes the margin between classes. They are powerful in high-dimensional feature spaces and can model complex relationships using kernel functions [226].

### Extreme Gradient Boosting (XGBoost)

XGBoost is a scalable implementation of gradient boosting designed for efficiency and accuracy. It employs system optimizations such as parallelization, regularization, and sparsity-aware learning, making it highly effective for large-scale ML tasks [227].

### AdaBoost

AdaBoost (Adaptive Boosting) combines multiple weak classifiers iteratively, giving higher weights to misclassified instances in each round. This method builds a strong classifier that often performs well with minimal parameter tuning [228].

### Gradient Boosting

Gradient Boosting is an ensemble method that builds models sequentially, where each new model minimizes the errors (residuals) of the previous ones by using

gradient descent optimization. It is flexible in handling various loss functions and achieves state-of-the-art performance in many prediction tasks [229].

### LightGBM

LightGBM is a gradient boosting framework developed by Microsoft that utilizes novel techniques, including Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB), to enhance both speed and memory efficiency. It is highly scalable and effective for large datasets [230].

### Nearest Neighbor

The nearest neighbor or k-Nearest Neighbor algorithm is a non-parametric, instance-based learning method. Given a new (unlabeled) example, it compares this example to all (or many) stored examples (the training set) using a distance metric (e.g., Euclidean, Manhattan). The simplest case is 1-NN: the label of the single closest training example is used for classification. In k-NN, a majority vote (or weighted vote) among the k closest training examples determines the class label. Because it makes no assumptions about the distribution of the data (non-parametric), it can adapt to complex data but may suffer in high dimensions or with large datasets (because of computation cost and "curse of dimensionality") [231][232].

# Chapter 5

# Results

In this chapter, the results of the four main stages of the project are presented in a structured and comprehensive manner. Each section corresponds to one of the stages: prostate zonal segmentation, prostate lesion segmentation, PI-RADS prediction, and real prostate cancer detection. For each stage, all relevant quantitative and qualitative outcomes are reported, including performance metrics, visualizations, and comparisons where appropriate. The results are organized into four distinct parts to provide clarity and allow the reader to evaluate the effectiveness of the methods applied at each step of the project. This structured presentation highlights both the individual performance of each component and the overall integration of the multi-stage AI-based prostate cancer diagnostic pipeline.

## 5.1 Prostate Zonal Segmentation

Tables 5.1 and 5.2 summarize the results of all 15 models and the average of each model and the average ensemble model for zonal segmentation on the validation and test sets, respectively.

Table 5.1: The results of 5-fold cross-validation of Att-R-Net, Vanilla-Net, V-Net, and average ensemble model for zonal segmentation on validation data.

| **Models** | **CG** | | **PZ** | |
|---|---|---|---|---|
| | **IoU** | **DSC** | **IoU** | **DSC** |
| Att-R-Net fold 1 | 78.4% | 87.9% | 54.5% | 70.5% |
| Att-R-Net fold 2 | 73.2% | 84.5% | 54.0% | 70.1% |
| Att-R-Net fold 3 | 73.3% | 84.6% | 58.3% | 73.6% |
| Att-R-Net fold 4 | 75.3% | 85.9% | 58.2% | 73.6% |
| Att-R-Net fold 5 | 70.8% | 82.9% | 58.1% | 73.5% |
| Vanilla-Net fold 1 | 78.5% | 87.6% | 55.7% | 71.5% |
| Vanilla-Net fold 2 | 78.3% | 87.8% | 57.9% | 73.3% |
| Vanilla-Net fold 3 | 77.4% | 87.3% | 54.3% | 70.3% |
| Vanilla-Net fold 4 | 79.3% | 88.2% | 56.4% | 72.1% |
| Vanilla-Net fold 5 | 78.5% | 87.9% | 58.4% | 73.8% |
| V-Net fold 1 | 76.7% | 86.8% | 56.1% | 71.9% |
| V-Net fold 2 | 76.2% | 86.3% | 57.3% | 72.9% |
| V-Net fold 3 | 76.9% | 86.9% | 56.7% | 72.3% |
| V-Net fold 4 | 77.7% | 87.4% | 57.3% | 72.8% |
| V-Net fold 5 | 76.8% | 86.9% | 58.6% | 73.9% |
| Average of Att-R-Net folds | 74.2% | 85.16% | 56.62% | 72.26% |
| Average of Vanilla-Net folds | 78.4% | 87.76% | 56.54% | 72.2% |
| Average of V-Net folds | 76.86% | 86.86% | 57.2% | 72.76% |
| Ensemble | 80.4% | 89.1% | 60.6% | 75.5% |

Table 5.2: The results of 5-fold cross-validation of Att-R-Net, Vanilla-Net, V-Net, and average ensemble model for zonal segmentation on test data.

| **Models** | **CG** | | **PZ** | |
|---|---|---|---|---|
| | **IoU** | **DSC** | **IoU** | **DSC** |
| Att-R-Net fold 1 | 78.9% | 88.2% | 49.4% | 66.1% |
| Att-R-Net fold 2 | 70.6% | 82.7% | 47.3% | 64.1% |
| Att-R-Net fold 3 | 74.5% | 85.3% | 49.8% | 66.4% |
| Att-R-Net fold 4 | 76.5% | 86.7% | 52.8% | 69.1% |
| Att-R-Net fold 5 | 71.7% | 83.5% | 49.8% | 66.5% |
| Vanilla-Net fold 1 | 77.9% | 87.6% | 52.8% | 69.1% |
| Vanilla-Net fold 2 | 78.6% | 88.0% | 53.6% | 69.8% |
| Vanilla-Net fold 3 | 78.2% | 87.8% | 50.8% | 67.4% |
| Vanilla-Net fold 4 | 77.5% | 87.3% | 50.1% | 66.7% |
| Vanilla-Net fold 5 | 77.6% | 87.4% | 52.7% | 69.0% |
| V-Net fold 1 | 75.0% | 85.7% | 48.2% | 65.1% |
| V-Net fold 2 | 77.0% | 87.0% | 50.5% | 67.1% |
| V-Net fold 3 | 75.7% | 86.1% | 48.2% | 65.0% |
| V-Net fold 4 | 76.7% | 86.8% | 52.0% | 68.4% |
| V-Net fold 5 | 77.1% | 87.0% | 50.2% | 66.9% |
| Average of Att-R-Net folds | 74.44% | 85.28% | 49.82% | 66.44% |
| Average of Vanilla-Net folds | 77.96% | 87.62% | 52% | 68.4% |
| Average of V-Net folds | 76.3% | 86.52% | 49.82% | 66.5% |
| Ensemble | 79.3% | 88.4% | 54.5% | 70.5% |

Tables 5.3 and 5.4 present the results of the Meta-Net, which utilizes various combinations of Att-R-Net, Vanilla-Net, and V-Net for zonal segmentation on both the validation and test sets. In these tables, Att-R-Net encompasses all models from folds 1 to 5 of the network, and the same applies to Vanilla-Net and V-Net. Additionally, the notation " Att-R-Net + Vanilla-Net " indicates the combination

of all attention models with Vanilla-Net models, and similar combinations are represented throughout the tables.

Table 5.3: The results of Meta-Net using different combinations of U-net-based networks for zonal segmentation on the validation dataset.

| **Models** | **CG** | | **PZ** | |
|---|---|---|---|---|
| | **IoU** | **DSC** | **IoU** | **DSC** |
| Att-R-Net | 75% | 85% | 56% | 72% |
| Vanilla-Net | 79% | 88% | 58% | 73% |
| V-Net | 78% | 88% | 57% | 72% |
| Att-R-Net + Vanilla-Net | 80% | 89% | 58% | 74% |
| Att-R-Net + V-Net | 79% | 88% | 56% | 72% |
| Vanilla-Net + V-Net | 80% | 89% | 58% | 73% |
| Att-R-Net + Vanilla-Net + V-Net | 80% | 89% | 58% | 73% |

Table 5.4: The results of Meta-Net using different combinations of U-net-based networks for zonal segmentation on the test dataset.

| **Models** | **CG** | | **PZ** | |
|---|---|---|---|---|
| | **IoU** | **DSC** | **IoU** | **DSC** |
| Att-R-Net | 74% | 85% | 49% | 65% |
| Vanilla-Net | 78% | 88% | 54% | 70% |
| V-Net | 78% | 87% | 52% | 69% |
| Att-R-Net + Vanilla-Net | 79% | 88% | 51% | 68% |
| Att-R-Net + V-Net | 78% | 88% | 51% | 68% |
| Vanilla-Net + V-Net | 78% | 88% | 54% | 71% |
| Att-R-Net + Vanilla-Net + V-Net | 79% | 88% | 53% | 69% |

The outcomes of YOLO-V8 are presented in Tables 5.5 and 5.6 for the validation dataset and test set.

Figure 5.1 shows a comparison between the segmented masks and the coarsely annotated labels across several samples of the test set.

Table 5.5: Validation set results (IoU and DSC) for YOLO-V8.

| **Model** | **CG** | | **PZ** | |
|---|---|---|---|---|
| | **IoU** | **DSC** | **IoU** | **DSC** |
| YOLO-V8 | 81% | 91% | 60% | 74% |

Table 5.6: Test set results (IoU and DSC) for YOLO-V8.

| **Model** | **CG** | | **PZ** | |
|---|---|---|---|---|
| | **IoU** | **DSC** | **IoU** | **DSC** |
| YOLO-V8 | 80% | 89% | 58% | 73% |

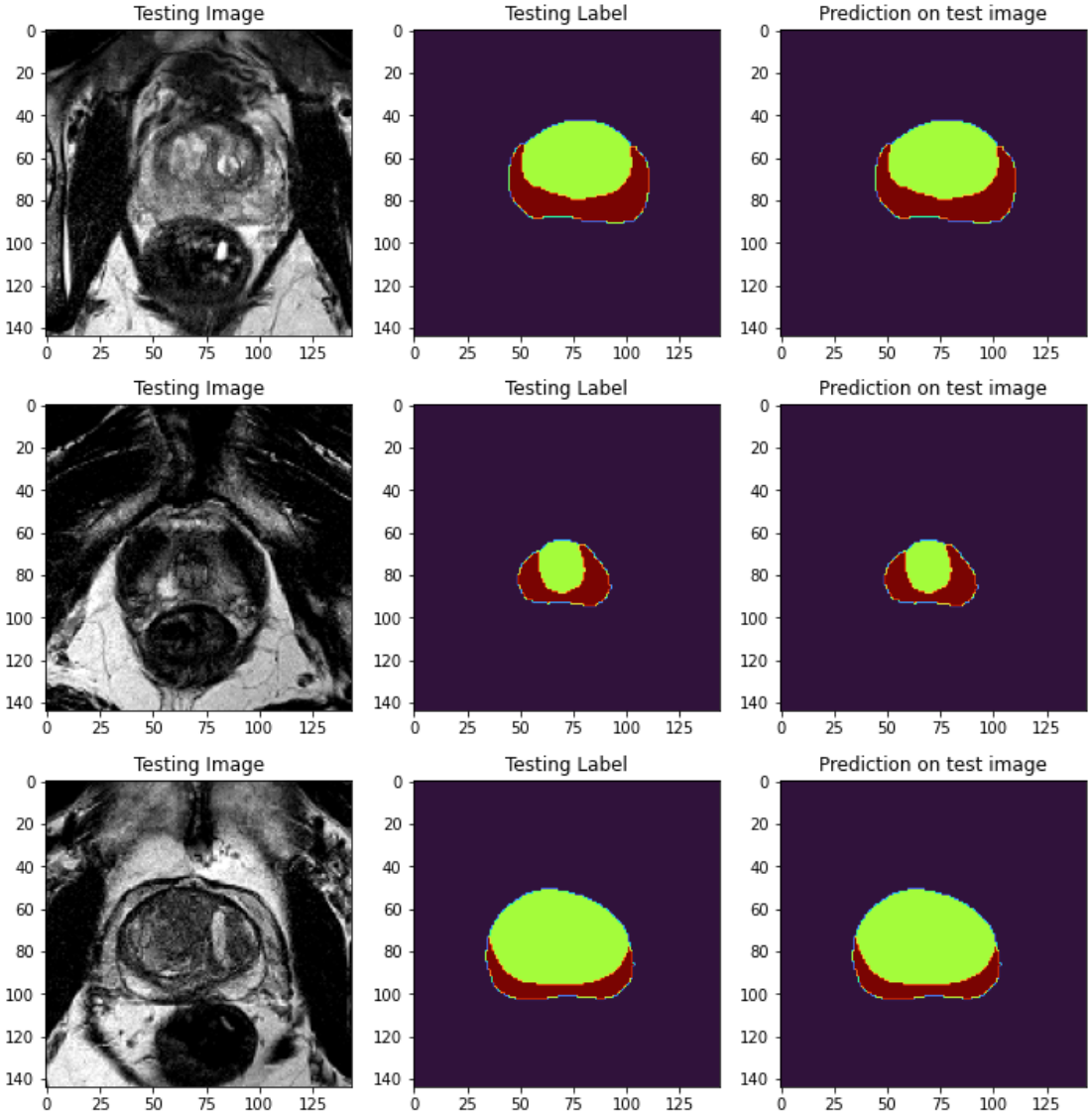





Figure 5.1: *Segmentation results of the prostate zones using the ensemble model of three examples of the test set. Columns (left to right): original image, original mask, predicted mask of CG and PZ.*

Figure 5.2 illustrates some examples of zonal segmentation using Meta-Net (Vanilla-Net + V-Net) on the test set.

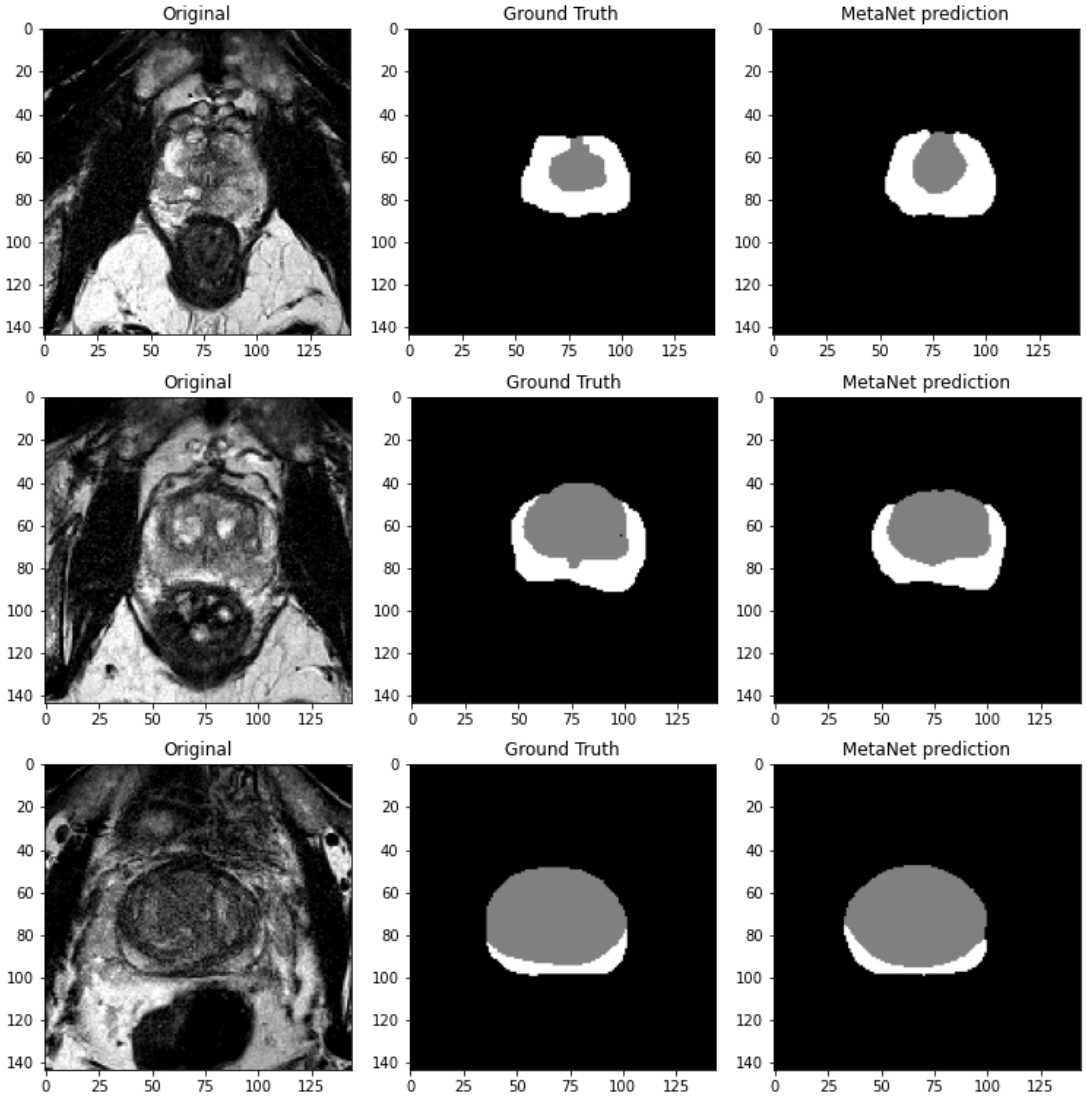



Figure 5.2: *Segmentation results of the Meta-Net: (left) original image, (middle) ground truth, and (right) predicted segmentation mask.*

Figure 5.3 illustrates YOLO's zonal detection and segmentation performance on some examples from the test.

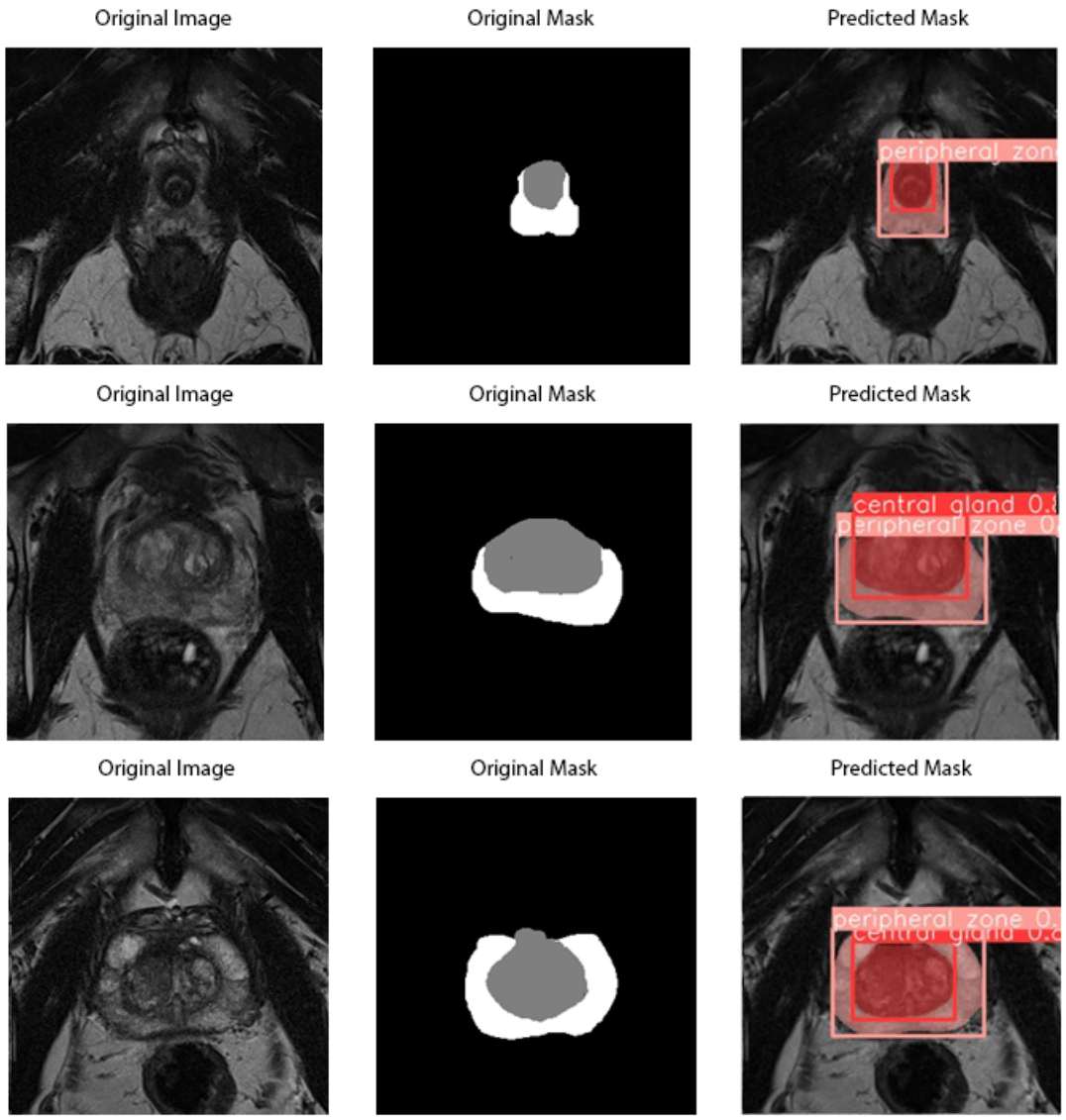



Figure 5.3: *Detection and segmentation results of the YOLO-V8: (left) original image, (middle) ground truth, and (right) predicted segmentation mask*

## 5.2 Prostate Lesion Segmentation

To comprehensively assess the performance of our proposed segmentation networks, two distinct training configurations were implemented, each followed by complementary evaluation strategies. The first configuration focused exclusively

on high-risk prostate cancer cases, using MRI scans labeled with PI-RADS 4 and 5. From this subset, 127 images were used for training and 24 for testing, with 5-fold cross-validation applied to ensure robust performance estimates. In the second configuration, the dataset was expanded to include PI-RADS 3 cases alongside PI-RADS 4 and 5, thereby increasing the diversity of lesion characteristics. This larger dataset comprised 270 training and 33 testing images and was similarly evaluated using 5-fold cross-validation. Two evaluation strategies were applied to the test set to gain detailed insights into segmentation performance. In the first strategy, the DSC was calculated on a per-patient basis by averaging DSC scores across slices within each scan, followed by reporting the mean and median values across all patients. In the second strategy, predictions were aggregated across all slices in the test set, and overall mean and median DSC values were computed at the dataset level. Patient-wise evaluation generally yielded lower DSC scores, likely due to intra-patient variability, including differences in image quality, contrast, and lesion size, which challenged consistent model performance. In contrast, the slice-level aggregation benefited from an averaging effect, where strong results on higher-quality scans mitigated weaker outcomes on lower-quality scans, resulting in improved overall segmentation accuracy.

The lesion segmentation results using the dataset containing PI-RADS 4 and 5 cases are summarized in Table 5.7.

Table 5.7: The results of lesion segmentation using the dataset containing PI-RADS 4 and 5.

| **Sequence** | **U-Net** | **Dense-U-Net** | **Attention-U-Net** | **LSTM U-Net** |
|---|---|---|---|---|
| ADC | 68% | 69% | 64% | 67% |
| DWI | 65% | 66% | 58% | - |
| T2W | 48% | 47% | 47% | - |
| ADC + DWI + T2W | 41% | 58% | 53% | - |

Across the three models, Dense-U-Net generally achieved the highest Dice scores, particularly for the ADC (69%) and DWI (66%) sequences, whereas Attention-U-Net consistently showed the lowest performance. Segmentation based on the

ADC sequence yielded the best results overall, indicating that lesions are more clearly distinguishable in this modality. In contrast, the T2W sequence produced the lowest accuracy for all models (47–48%), suggesting that lesion boundaries are less distinct in these images. When combining all three sequences (ADC + DWI + T2W), only Dense-U-Net showed a notable improvement (58%), while U-Net and Attention-U-Net did not benefit substantially, highlighting that effective multi-sequence integration depends strongly on the network architecture. Overall, these results indicate that Dense-U-Net is the most robust model for lesion segmentation and that ADC provides the most informative input for delineating high-risk lesions.

Table 5.8 reports the lesion segmentation performance on the dataset, including PI-RADS 3, 4, and 5 cases.

Table 5.8: The results of lesion segmentation using the dataset containing PI-RADS 3, 4, and 5.

| **Sequence** | **U-Net** | **Dense-U-Net** | **Attention-U-Net** | **LSTM U-Net** |
|---|---|---|---|---|
| ADC | 66% | 68% | 63% | 65% |
| DWI | 62% | 63% | 53% | - |
| T2W | 40% | 48% | 43% | - |
| ADC + DWI + T2W | 53% | 54% | 48% | - |

Among the models, Dense-U-Net consistently delivered the highest Dice scores, achieving 68% for ADC and 63% for DWI, while Attention-U-Net showed the lowest results across most sequences. The ADC sequence again proved to be the most informative for lesion delineation, whereas segmentation based on T2W images was the least accurate (40–48%), reflecting the difficulty in identifying lesions in this modality. For the combined sequences (ADC + DWI + T2W), performance improvements were modest, with Dense-U-Net reaching 54%, suggesting that integrating multi-sequence data offers only a limited advantage in this dataset. Overall, Dense-U-Net demonstrated the greatest robustness, and ADC remained the most valuable sequence for accurate lesion segmentation, even when lower-risk PI-RADS 3 cases were included. Figure 5.4 illustrates examples of lesion segmentation on

ADC images for cases with PI-RADS 4 and 5, highlighting the model's ability to delineate high-risk lesions. In contrast, Figure 5.5 presents segmentation results on ADC images for the broader dataset, including PI-RADS 3, 4, and 5, demonstrating how the models perform when lower-risk lesions are also considered

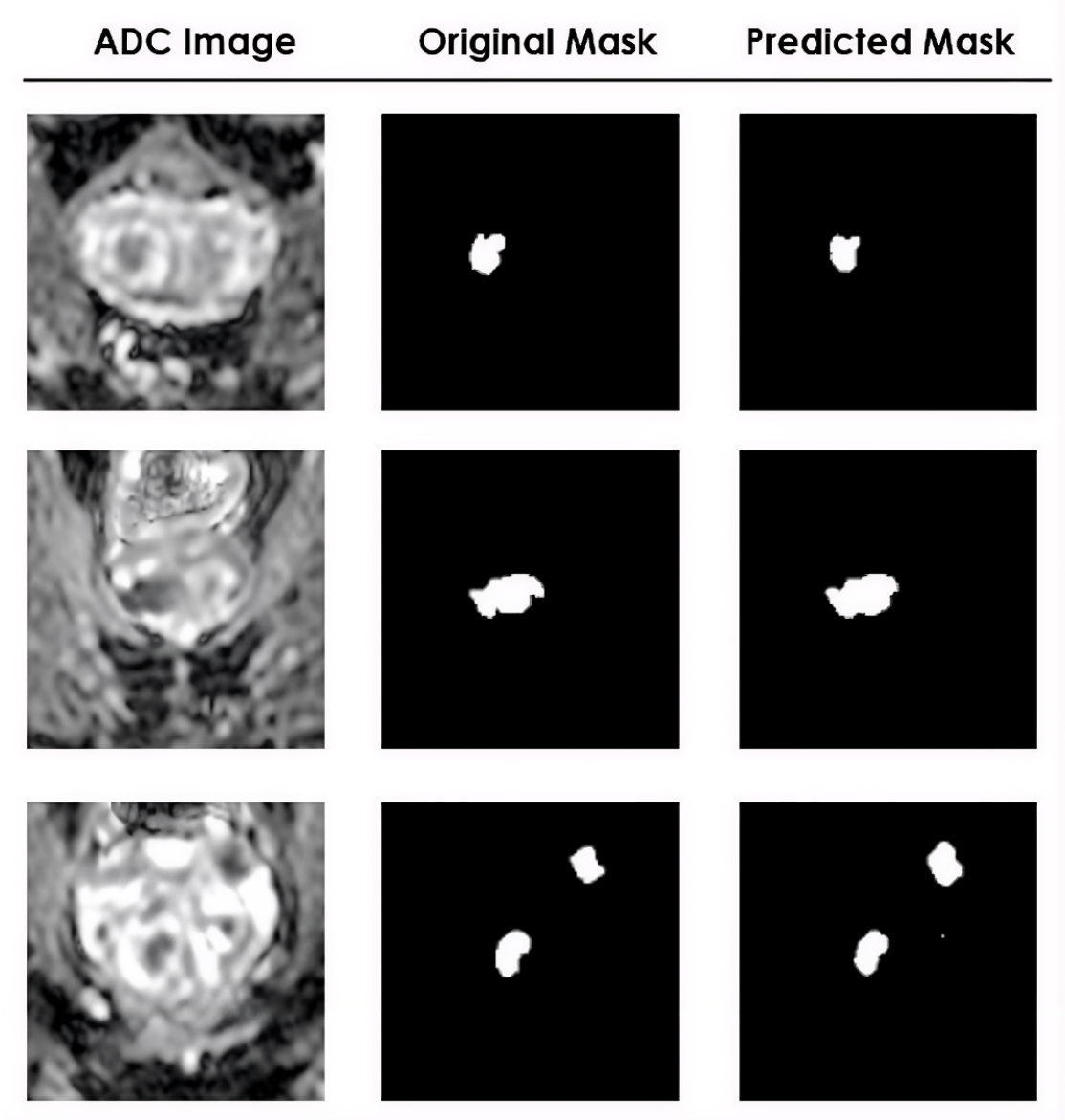



Figure 5.4: *Examples of lesion segmentation on ADC images for PI-RADS 4 and 5 cases, showing model performance on high-risk and mixed-risk lesions*

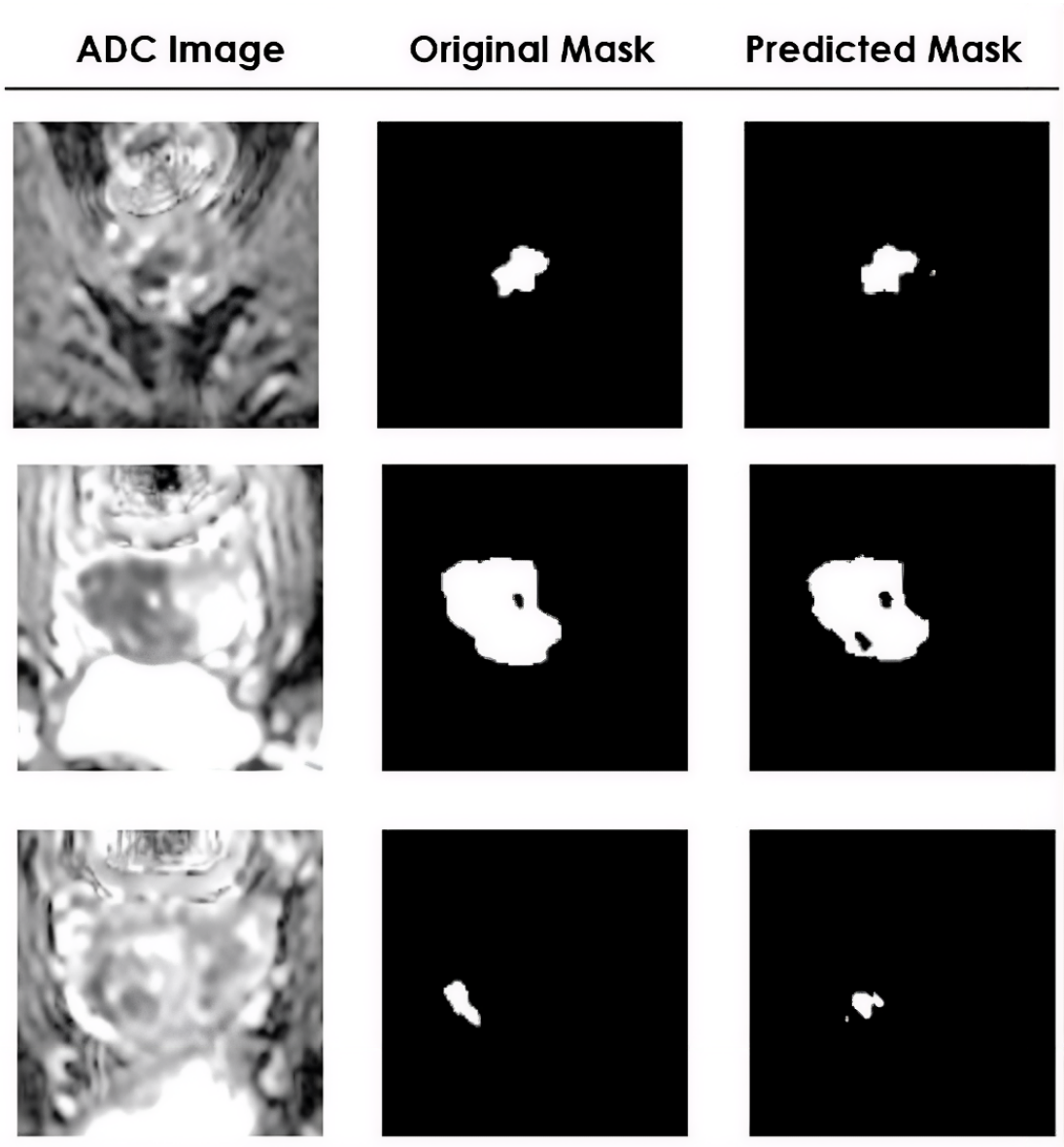



Figure 5.5: *Examples of lesion segmentation on ADC images for PI-RADS 3, 4, and 5 cases, showing model performance on high-risk and mixed-risk lesions*

## 5.3 PI-RADS Classification Results

In the following sections, we present the results obtained for the three proposed approaches. For Approach 1, radiomic features were extracted from all three MRI

sequences (T2W, ADC, and DWI), after which the features from these sequences were combined. This procedure generated four distinct result sets. The extracted features were subsequently used to train ML models for PI-RADS classification into three categories: PI-RADS 3, PI-RADS 4, and PI-RADS 5. As outlined in the description of Approach 1, two separate categories of features were considered, and the results corresponding to each category are reported independently. Model performance was assessed using precision, recall, and AUC for each PI-RADS class. In addition, the average accuracy and mean AUC across all classes are also presented. For clarity and consistency, the results of each approach are presented in dedicated sections.

### 5.3.1 Results of Approach 1 Using Pyradiomics Feature Set 1

Tables 5.9–5.12 summarize the results obtained with Feature Set 1 across different input configurations: Table 1 corresponds to ADC, Table 2 to DWI, Table 3 to T2W, and Table 4 to the combined sequences. For each configuration, performance metrics are reported in terms of precision, recall, and AUC for the individual PI-RADS classes, along with the overall average accuracy and mean AUC across all classes.

Table 5.9: Performance of PI-RADS Classification Models of Approach 1, Using Pyradiomics Feature Set 1 Across ADC Images

| **Model** | **Acc.** | **AUC** | **PI-RADS 3** | | | **PI-RADS 4** | | | **PI-RADS 5** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | **Prec.** | **Recall** | **AUC** | **Prec.** | **Recall** | **AUC** | **Prec.** | **Recall** | **AUC** |
| RF | 0.65 | 0.84 | 0.46 | 0.58 | 0.79 | 0.60 | 0.63 | 0.76 | 1.00 | 0.73 | 0.97 |
| XGBoost | 0.67 | 0.83 | 0.50 | 0.66 | 0.79 | 0.63 | 0.63 | 0.75 | 1.00 | 0.73 | 0.94 |
| AdaBoost | 0.67 | 0.84 | 0.57 | 0.66 | 0.75 | 0.63 | 0.77 | 0.81 | 1.00 | 0.53 | 0.97 |
| Gradient Boosting | 0.67 | 0.84 | 0.50 | 0.58 | 0.79 | 0.62 | 0.68 | 0.76 | 1.00 | 0.73 | 0.96 |
| LightGBM | 0.65 | 0.83 | 0.47 | 0.66 | 0.78 | 0.65 | 0.68 | 0.77 | 1.00 | 0.60 | 0.94 |
| SVM | 0.61 | 0.77 | 0.46 | 0.58 | 0.72 | 0.59 | 0.72 | 0.65 | 1.00 | 0.46 | 0.94 |
| Ensemble 1 | 0.77 | 0.83 | 0.66 | 0.66 | 0.77 | 0.72 | 0.81 | 0.77 | 1.00 | 0.80 | 0.96 |
| Ensemble 2 | 0.65 | 0.82 | 0.50 | 0.66 | 0.78 | 0.63 | 0.63 | 0.75 | 0.90 | 0.66 | 0.93 |
| Ensemble 3 | 0.69 | 0.81 | 0.50 | 0.66 | 0.75 | 0.66 | 0.63 | 0.75 | 1.00 | 0.80 | 0.94 |
| Ensemble 4 | 0.67 | 0.83 | 0.50 | 0.66 | 0.78 | 0.63 | 0.63 | 0.75 | 1.00 | 0.73 | 0.96 |

Table 5.10: Performance of PI-RADS Classification Models of Approach 1, Using Pyradiomics Feature Set 1 Across DWI Images

| **Model** | **Acc.** | **AUC** | **PI-RADS 3** | | | **PI-RADS 4** | | | **PI-RADS 5** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | **Prec.** | **Recall** | **AUC** | **Prec.** | **Recall** | **AUC** | **Prec.** | **Recall** | **AUC** |
| RF | 0.63 | 0.78 | 0.37 | 0.25 | 0.70 | 0.58 | 0.77 | 0.67 | 0.91 | 0.73 | 0.97 |
| XGBoost | 0.59 | 0.72 | 0.37 | 0.25 | 0.63 | 0.54 | 0.77 | 0.63 | 0.90 | 0.60 | 0.89 |
| AdaBoost | 0.59 | 0.81 | 0.38 | 0.41 | 0.71 | 0.56 | 0.63 | 0.73 | 0.90 | 0.66 | 0.97 |
| Gradient Boosting | 0.57 | 0.74 | 0.30 | 0.25 | 0.61 | 0.53 | 0.68 | 0.63 | 0.90 | 0.66 | 0.97 |
| LightGBM | 0.49 | 0.70 | 0.20 | 0.16 | 0.59 | 0.46 | 0.63 | 0.59 | 0.88 | 0.53 | 0.93 |
| SVM | 0.63 | 0.72 | 0.00 | 0.00 | 0.61 | 0.56 | 0.95 | 0.61 | 0.90 | 0.66 | 0.95 |
| Ensemble 1 | 0.59 | 0.77 | 0.00 | 0.00 | 0.66 | 0.54 | 0.90 | 0.70 | 0.90 | 0.60 | 0.93 |
| Ensemble 2 | 0.63 | 0.75 | 0.66 | 0.16 | 0.67 | 0.56 | 0.95 | 0.64 | 0.88 | 0.53 | 0.93 |
| Ensemble 3 | 0.61 | 0.74 | 0.44 | 0.33 | 0.65 | 0.56 | 0.77 | 0.66 | 0.90 | 0.60 | 0.91 |
| Ensemble 4 | 0.63 | 0.78 | 0.45 | 0.41 | 0.71 | 0.59 | 0.72 | 0.67 | 0.90 | 0.66 | 0.96 |

Table 5.11: Performance of PI-RADS Classification Models of Approach 1, Using Pyradiomics Feature Set 1 Across T2W Images

| **Model** | **Acc.** | **AUC** | **PI-RADS 3** | | | **PI-RADS 4** | | | **PI-RADS 5** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | **Prec.** | **Recall** | **AUC** | **Prec.** | **Recall** | **AUC** | **Prec.** | **Recall** | **AUC** |
| RF | 0.61 | 0.73 | 0.00 | 0.00 | 0.61 | 0.54 | 0.77 | 0.61 | 0.86 | 0.86 | 0.96 |
| XGBoost | 0.57 | 0.68 | 0.16 | 0.20 | 0.50 | 0.51 | 0.63 | 0.60 | 0.85 | 0.80 | 0.93 |
| AdaBoost | 0.61 | 0.80 | 0.37 | 0.25 | 0.71 | 0.57 | 0.72 | 0.73 | 0.84 | 0.73 | 0.93 |
| Gradient Boosting | 0.59 | 0.67 | 0.16 | 0.08 | 0.47 | 0.53 | 0.68 | 0.59 | 0.86 | 0.86 | 0.96 |
| LightGBM | 0.61 | 0.67 | 0.33 | 0.16 | 0.48 | 0.55 | 0.72 | 0.57 | 0.85 | 0.80 | 0.96 |
| SVM | 0.63 | 0.73 | 0.00 | 0.00 | 0.63 | 0.55 | 0.95 | 0.60 | 0.90 | 0.66 | 0.96 |
| Ensemble 1 | 0.63 | 0.72 | 0.00 | 0.00 | 0.58 | 0.55 | 0.90 | 0.65 | 0.84 | 0.73 | 0.93 |
| Ensemble 2 | 0.46 | 0.70 | 0.23 | 0.33 | 0.53 | 0.47 | 0.50 | 0.63 | 0.88 | 0.53 | 0.95 |
| Ensemble 3 | 0.57 | 0.69 | 0.00 | 0.00 | 0.53 | 0.51 | 0.77 | 0.58 | 0.84 | 0.73 | 0.96 |
| Ensemble 4 | 0.63 | 0.72 | 0.33 | 0.16 | 0.61 | 0.57 | 0.72 | 0.61 | 0.86 | 0.86 | 0.96 |

Table 5.12: Performance of PI-RADS Classification Models of Approach 1, Using Pyradiomics Feature Set 1 Across Combined Features of ADC, DWI, and T2W Images

| **Model** | **Acc.** | **AUC** | **PI-RADS 3** | | | **PI-RADS 4** | | | **PI-RADS 5** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | **Prec.** | **Recall** | **AUC** | **Prec.** | **Recall** | **AUC** | **Prec.** | **Recall** | **AUC** |
| RF | 0.653 | 0.83 | 0.471 | 0.667 | 0.765 | 0.632 | 0.545 | 0.743 | 0.923 | 0.8 | 0.982 |
| XGBoost | 0.653 | 0.812 | 0.471 | 0.667 | 0.743 | 0.632 | 0.545 | 0.714 | 0.923 | 0.8 | 0.978 |
| AdaBoost | 0.592 | 0.803 | 0.357 | 0.417 | 0.743 | 0.56 | 0.636 | 0.677 | 1 | 0.667 | 0.988 |
| Gradient Boosting | 0.653 | 0.816 | 0.471 | 0.667 | 0.764 | 0.632 | 0.545 | 0.702 | 0.923 | 0.8 | 0.982 |
| LightGBM | 0.592 | 0.782 | 0.438 | 0.583 | 0.721 | 0.545 | 0.545 | 0.672 | 0.909 | 0.667 | 0.953 |
| SVM | 0.531 | 0.746 | 0.143 | 0.083 | 0.7 | 0.485 | 0.727 | 0.581 | 1 | 0.6 | 0.957 |
| Ensemble 1 | 0.653 | 0.841 | 0.533 | 0.667 | 0.815 | 0.609 | 0.636 | 0.732 | 0.909 | 0.667 | 0.975 |
| Ensemble 2 | 0.51 | 0.731 | 0.286 | 0.333 | 0.678 | 0.48 | 0.545 | 0.581 | 0.9 | 0.6 | 0.933 |
| Ensemble 3 | 0.694 | 0.813 | 0.533 | 0.667 | 0.795 | 0.667 | 0.636 | 0.702 | 0.923 | 0.8 | 0.941 |
| Ensemble 4 | 0.633 | 0.83 | 0.444 | 0.667 | 0.784 | 0.611 | 0.5 | 0.731 | 0.92 | 0.8 | 0.976 |

The results revealed several noteworthy trends. Models trained on ADC features consistently outperformed those based solely on DWI or T2W, with the best performance achieved by Ensemble 1 on ADC, yielding an accuracy of 0.77 and an AUC of 0.83. In contrast, models using DWI or T2W features demonstrated lower accuracies, generally ranging from 0.49 to 0.63, although certain configurations attained competitive AUC values exceeding 0.80 (e.g., AdaBoost on DWI with 0.81). When features from all three sequences were combined, the models exhibited more balanced performance. Several ensemble and gradient boosting approaches achieved accuracies above 0.65 and AUCs greater than 0.81. Notably, Ensemble 3 with combined features reached an accuracy of 0.69 and produced strong class-specific AUCs, particularly for PI-RADS 5 (commonly  0.94). Across all experiments, classification of PI-RADS 5 was the most accurate and stable, with AUC values frequently above 0.94, while PI-RADS 3 proved more challenging, showing lower precision and recall, especially when trained on DWI and T2W inputs. Collectively, these findings underscore the discriminative strength of ADC-derived features and highlight the added value of integrating multi-sequence information for more robust PI-RADS classification.

### 5.3.2 Results of Approach 1 Using Pyradiomics Feature Set 2

Tables 5.13–5.16 present the results derived from Feature Set 2 under various input configurations: ADC (Table 5.13), DWI (Table 5.14), T2W (Table 5.15), and the combination of all sequences (Table 5.16). For each configuration, the reported metrics include precision, recall, and AUC for each PI-RADS class, as well as the overall average accuracy and mean AUC across all classes.

Table 5.13: Performance of PI-RADS Classification Models of Approach 1, Using PyRadiomics Feature Set 2 Across ADC Images

| **Model** | **Acc.** | **AUC** | **PI-RADS 3** | | | **PI-RADS 4** | | | **PI-RADS 5** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | **Prec.** | **Recall** | **AUC** | **Prec.** | **Recall** | **AUC** | **Prec.** | **Recall** | **AUC** |
| RF | 0.65 | 0.84 | 0.42 | 0.50 | 0.77 | 0.62 | 0.68 | 0.774 | 1.00 | 0.73 | 0.97 |
| XGBoost | 0.71 | 0.86 | 0.53 | 0.66 | 0.80 | 0.70 | 0.63 | 0.79 | 0.92 | 0.86 | 0.98 |
| AdaBoost | 0.69 | 0.82 | 0.55 | 0.83 | 0.79 | 0.66 | 0.63 | 0.69 | 1.00 | 0.66 | 0.98 |
| Gradient Boosting | 0.71 | 0.83 | 0.53 | 0.58 | 0.76 | 0.68 | 0.77 | 0.75 | 1.00 | 0.73 | 0.97 |
| LightGBM | 0.71 | 0.84 | 0.58 | 0.58 | 0.79 | 0.66 | 0.72 | 0.74 | 0.92 | 0.80 | 0.97 |
| SVM | 0.67 | 0.83 | 0.50 | 0.41 | 0.75 | 0.63 | 0.86 | 0.77 | 1.00 | 0.60 | 0.96 |
| Ensemble 1 | 0.73 | 0.87 | 0.57 | 0.66 | 0.83 | 0.69 | 0.72 | 0.81 | 1.00 | 0.80 | 0.98 |
| Ensemble 2 | 0.71 | 0.85 | 0.46 | 0.58 | 0.78 | 0.71 | 0.68 | 0.77 | 1.00 | 0.86 | 0.99 |
| Ensemble 3 | 0.75 | 0.86 | 0.57 | 0.66 | 0.79 | 0.72 | 0.72 | 0.80 | 1.00 | 0.86 | 0.98 |
| Ensemble 4 | 0.71 | 0.85 | 0.53 | 0.66 | 0.80 | 0.70 | 0.63 | 0.77 | 0.92 | 0.86 | 0.97 |

Table 5.14: Performance of PI-RADS Classification Models of Approach 1, Using PyRadiomics Feature Set 2 Across DWI Images

| **Model** | **Acc.** | **AUC** | **PI-RADS 3** | | | **PI-RADS 4** | | | **PI-RADS 5** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | **Prec.** | **Recall** | **AUC** | **Prec.** | **Recall** | **AUC** | **Prec.** | **Recall** | **AUC** |
| RF | 0.63 | 0.77 | 0.37 | 0.25 | 0.69 | 0.58 | 0.77 | 0.67 | 0.91 | 0.73 | 0.95 |
| XGBoost | 0.65 | 0.73 | 0.50 | 0.41 | 0.64 | 0.60 | 0.77 | 0.62 | 0.90 | 0.66 | 0.92 |
| AdaBoost | 0.49 | 0.72 | 0.26 | 0.41 | 0.62 | 0.44 | 0.36 | 0.56 | 0.91 | 0.73 | 0.97 |
| Gradient Boosting | 0.59 | 0.73 | 0.33 | 0.25 | 0.61 | 0.55 | 0.72 | 0.63 | 0.90 | 0.66 | 0.95 |
| LightGBM | 0.63 | 0.72 | 0.45 | 0.41 | 0.62 | 0.59 | 0.72 | 0.63 | 0.90 | 0.66 | 0.92 |
| SVM | 0.61 | 0.74 | 0.00 | 0.00 | 0.62 | 0.55 | 0.90 | 0.65 | 0.90 | 0.66 | 0.95 |
| Ensemble 1 | 0.59 | 0.75 | 0.00 | 0.00 | 0.60 | 0.54 | 0.86 | 0.70 | 0.90 | 0.66 | 0.94 |
| Ensemble 2 | 0.44 | 0.64 | 0.26 | 0.33 | 0.54 | 0.45 | 0.50 | 0.52 | 0.70 | 0.46 | 0.86 |
| Ensemble 3 | 0.57 | 0.68 | 0.33 | 0.25 | 0.55 | 0.55 | 0.72 | 0.58 | 0.81 | 0.60 | 0.92 |
| Ensemble 4 | 0.57 | 0.74 | 0.30 | 0.25 | 0.64 | 0.53 | 0.68 | 0.63 | 0.90 | 0.66 | 0.95 |

Table 5.15: Performance of PI-RADS Classification Models of Approach 1, Using PyRadiomics Feature Set 2 Across T2W Images

| Model | Acc. | AUC | PI-RADS 3 | | | PI-RADS 4 | | | PI-RADS 5 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Prec. | Recall | AUC | Prec. | Recall | AUC | Prec. | Recall | AUC |
| RF | 0.63 | 0.71 | 0.40 | 0.16 | 0.55 | 0.56 | 0.77 | 0.62 | 0.85 | 0.80 | 0.96 |
| XGBoost | 0.63 | 0.68 | 0.42 | 0.25 | 0.50 | 0.57 | 0.72 | 0.58 | 0.85 | 0.80 | 0.95 |
| AdaBoost | 0.63 | 0.77 | 0.40 | 0.16 | 0.67 | 0.56 | 0.77 | 0.66 | 0.85 | 0.80 | 0.97 |
| Gradient Boosting | 0.61 | 0.67 | 0.28 | 0.16 | 0.47 | 0.55 | 0.72 | 0.58 | 0.92 | 0.80 | 0.98 |
| LightGBM | 0.57 | 0.68 | 0.25 | 0.16 | 0.51 | 0.51 | 0.63 | 0.56 | 0.85 | 0.80 | 0.96 |
| SVM | 0.61 | 0.70 | 0.00 | 0.00 | 0.61 | 0.53 | 0.95 | 0.59 | 0.90 | 0.60 | 0.89 |
| Ensemble 1 | 0.65 | 0.71 | 0.00 | 0.00 | 0.62 | 0.57 | 0.90 | 0.63 | 0.92 | 0.80 | 0.88 |
| Ensemble 2 | 0.53 | 0.76 | 0.00 | 0.00 | 0.66 | 0.50 | 0.72 | 0.67 | 0.90 | 0.66 | 0.95 |
| Ensemble 3 | 0.63 | 0.70 | 0.33 | 0.16 | 0.54 | 0.56 | 0.77 | 0.59 | 0.92 | 0.80 | 0.95 |
| Ensemble 4 | 0.55 | 0.71 | 0.22 | 0.16 | 0.55 | 0.50 | 0.59 | 0.60 | 0.85 | 0.80 | 0.97 |

Table 5.16: Performance of PI-RADS Classification Models of Approach 1, Using PyRadiomics Feature Set 2 Across Combined Features of ADC, DWI, and T2W Images

| Model | Acc. | AUC | PI-RADS 3 | | | PI-RADS 4 | | | PI-RADS 5 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Prec. | Recall | AUC | Prec. | Recall | AUC | Prec. | Recall | AUC |
| RF | 0.69 | 0.83 | 0.50 | 0.66 | 0.76 | 0.68 | 0.59 | 0.75 | 0.92 | 0.86 | 0.98 |
| XGBoost | 0.67 | 0.80 | 0.47 | 0.66 | 0.73 | 0.66 | 0.54 | 0.72 | 0.92 | 0.86 | 0.95 |
| AdaBoost | 0.67 | 0.81 | 0.50 | 0.75 | 0.77 | 0.65 | 0.59 | 0.68 | 1.00 | 0.73 | 0.98 |
| Gradient Boosting | 0.67 | 0.78 | 0.46 | 0.58 | 0.71 | 0.65 | 0.59 | 0.70 | 0.92 | 0.86 | 0.94 |
| LightGBM | 0.67 | 0.79 | 0.50 | 0.50 | 0.71 | 0.62 | 0.68 | 0.70 | 0.92 | 0.80 | 0.95 |
| SVM | 0.59 | 0.78 | 0.50 | 0.25 | 0.73 | 0.52 | 0.86 | 0.65 | 1.00 | 0.46 | 0.97 |
| Ensemble 1 | 0.69 | 0.84 | 0.54 | 0.50 | 0.80 | 0.64 | 0.72 | 0.75 | 0.92 | 0.80 | 0.98 |
| Ensemble 2 | 0.71 | 0.83 | 0.53 | 0.58 | 0.79 | 0.68 | 0.68 | 0.73 | 0.92 | 0.86 | 0.97 |
| Ensemble 3 | 0.69 | 0.83 | 0.54 | 0.50 | 0.79 | 0.64 | 0.72 | 0.74 | 0.92 | 0.80 | 0.98 |
| Ensemble 4 | 0.67 | 0.82 | 0.46 | 0.58 | 0.75 | 0.65 | 0.59 | 0.73 | 0.92 | 0.86 | 0.97 |

Evaluation of Pyradiomics Feature Set 2 across the various input configurations demonstrated that models trained on ADC features generally outperformed those based solely on DWI or T2W images. Within the ADC-based models, Ensemble 3 achieved the highest overall accuracy (0.755) with an AUC of 0.860, closely followed by Ensemble 1 and several gradient boosting variants. For these models, classification of PI-RADS 5 remained consistently strong (AUC $\geq$ 0.97), while PI-RADS 3 proved the most challenging, with notably lower precision and recall.

Models based on DWI features exhibited moderate performance, with accuracies typically ranging from 0.57 to 0.65, and relatively few configurations surpassed an AUC of 0.75. Within this modality, AdaBoost and Ensemble 2 showed the weakest results, whereas XGBoost and Random Forest produced more balanced outcomes. Similarly, T2W-based models generally underperformed compared to ADC and combined-sequence configurations, with most accuracies also between 0.57 and 0.65, although some ensemble methods maintained reasonable AUC values for PI-RADS 5. Combining features from ADC, DWI, and T2W improved performance relative to single-sequence DWI or T2W models and approached the levels observed with ADC alone. Notably, Ensemble 2 trained on combined features achieved an accuracy of 0.714 and an AUC of 0.834, highlighting the advantage of multi-sequence integration for more robust PI-RADS classification. Across all configurations and modalities, PI-RADS 5 was consistently predicted with high reliability, whereas PI-RADS 3 remained the most difficult class to classify accurately, reflecting the inherent challenge of differentiating intermediate cases.

### 5.3.3 Results of Approach 2

In the second approach, the complete workflow, from zonal and lesion segmentation to zone assignment, was executed automatically using pre-trained segmentation models in combination with a custom spatial overlap algorithm. Radiomic features were subsequently extracted from these automated segmentations, and PI-RADS classification was performed following the same procedure as in the first approach. Table 5.17 reports the results obtained with ADC images using Feature Set 1, while Table 5.18 presents the corresponding outcomes for Feature Set 2.

Table 5.17: Performance of PI-RADS Classification Models of Approach 2, Using PyRadiomics Feature Set 1 Across ADC Images

| Model | Acc. | AUC | PI-RADS 3 | | | PI-RADS 4 | | | PI-RADS 5 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Prec. | Recall | AUC | Prec. | Recall | AUC | Prec. | Recall | AUC |
| RF | 0.60 | 0.82 | 0.16 | 0.16 | 0.74 | 0.61 | 0.57 | 0.78 | 0.78 | 0.84 | 0.95 |
| XGBoost | 0.69 | 0.83 | 0.42 | 0.50 | 0.74 | 0.66 | 0.57 | 0.78 | 0.85 | 0.92 | 0.98 |
| AdaBoost | 0.63 | 0.79 | 0.25 | 0.16 | 0.71 | 0.56 | 0.64 | 0.71 | 0.84 | 0.84 | 0.94 |
| Gradient Boosting | 0.63 | 0.84 | 0.28 | 0.33 | 0.78 | 0.63 | 0.50 | 0.78 | 0.80 | 0.92 | 0.98 |
| LightGBM | 0.69 | 0.82 | 0.42 | 0.50 | 0.77 | 0.66 | 0.57 | 0.74 | 0.85 | 0.92 | 0.95 |
| SVM | 0.51 | 0.76 | 0.00 | 0.00 | 0.66 | 0.50 | 0.42 | 0.72 | 0.64 | 0.84 | 0.89 |
| Ensemble 1 | 0.66 | 0.83 | 0.20 | 0.16 | 0.79 | 0.61 | 0.57 | 0.74 | 0.86 | 1.00 | 0.97 |
| Ensemble 2 | 0.57 | 0.82 | 0.40 | 0.33 | 0.83 | 0.05 | 0.42 | 0.70 | 0.68 | 0.84 | 0.93 |
| Ensemble 3 | 0.69 | 0.80 | 0.33 | 0.33 | 0.71 | 0.66 | 0.57 | 0.74 | 0.86 | 1.00 | 0.96 |
| Ensemble 4 | 0.66 | 0.84 | 0.33 | 0.33 | 0.73 | 0.61 | 0.57 | 0.81 | 0.85 | 0.92 | 0.97 |

Table 5.18: Performance of PI-RADS Classification Models of Approach 2, Using PyRadiomics Feature Set 2 Across ADC Images

| Model | Acc. | AUC | PI-RADS 3 | | | PI-RADS 4 | | | PI-RADS 5 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Prec. | Recall | AUC | Prec. | Recall | AUC | Prec. | Recall | AUC |
| RF | 0.60 | 0.81 | 0.16 | 0.16 | 0.71 | 0.63 | 0.50 | 0.81 | 0.75 | 0.92 | 0.91 |
| XGBoost | 0.66 | 0.82 | 0.37 | 0.50 | 0.74 | 0.70 | 0.50 | 0.78 | 0.80 | 0.92 | 0.93 |
| AdaBoost | 0.72 | 0.81 | 0.33 | 0.16 | 0.73 | 0.64 | 0.78 | 0.76 | 0.92 | 0.92 | 0.93 |
| Gradient Boosting | 0.63 | 0.80 | 0.28 | 0.33 | 0.72 | 0.77 | 0.50 | 0.74 | 0.70 | 0.92 | 0.92 |
| LightGBM | 0.69 | 0.81 | 0.37 | 0.50 | 0.72 | 0.80 | 0.57 | 0.77 | 0.80 | 0.92 | 0.92 |
| SVM | 0.57 | 0.78 | 0.16 | 0.16 | 0.66 | 0.57 | 0.57 | 0.77 | 0.76 | 0.76 | 0.91 |
| Ensemble 1 | 0.63 | 0.82 | 0.33 | 0.16 | 0.76 | 0.55 | 0.71 | 0.79 | 0.83 | 0.76 | 0.91 |
| Ensemble 2 | 0.48 | 0.64 | 0.00 | 0.00 | 0.58 | 0.46 | 0.50 | 0.58 | 0.64 | 0.69 | 0.76 |
| Ensemble 3 | 0.60 | 0.81 | 0.28 | 0.33 | 0.74 | 0.53 | 0.50 | 0.77 | 0.84 | 0.84 | 0.92 |
| Ensemble 4 | 0.66 | 0.81 | 0.33 | 0.50 | 0.71 | 0.77 | 0.50 | 0.79 | 0.80 | 0.92 | 0.91 |

For Feature Set 1 (Table5.17), the best accuracy (0.697) was achieved by XGBoost, LightGBM, and Ensemble 3, while the highest overall AUC (0.849) was obtained with Gradient Boosting. Notably, XGBoost demonstrated strong class-specific results, particularly for PI-RADS 4 (precision = 0.667, recall = 0.571, AUC = 0.789) and PI-RADS 5 (precision = 0.857, recall = 0.923, AUC = 0.985).

For Feature Set 2 (Table 5.18), AdaBoost achieved the highest accuracy (0.727) alongside a competitive AUC of 0.812. LightGBM and XGBoost also performed

consistently well, especially in the classification of PI-RADS 5 (AUC = 0.927 and 0.935, respectively). In contrast, Ensemble 2 yielded the weakest outcomes, with an accuracy of 0.485 and an AUC of 0.645.

Overall, transitioning from Feature Set 1 to Feature Set 2 resulted in accuracy improvements for certain models (e.g., AdaBoost), but not across all configurations. XGBoost and LightGBM emerged as consistently strong performers across both feature sets, whereas support vector machine and Ensemble 2 tended to underperform in this fully automated workflow.

### 5.3.4 Results of Approach 3

In Approach 3, a different strategy was adopted by discarding Pyradiomics features. Instead, feature extraction was performed using the proposed convolution neural network, trained on paired ADC images and their corresponding masks. The extracted features were then provided as input to ML models for PI-RADS classification. The results of this approach are summarized in Table 5.19.

Table 5.19: Performance of PI-RADS Classification Models of Approach 3, Using Extracted Features by Proposed Convolution Neural Network Across ADC Images

| **Model** | **Acc.** | **AUC** | **PI-RADS 3** | | | **PI-RADS 4** | | | **PI-RADS 5** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | **Prec.** | **Recall** | **AUC** | **Prec.** | **Recall** | **AUC** | **Prec.** | **Recall** | **AUC** |
| RF | 0.65 | 0.84 | 0.37 | 0.25 | 0.78 | 0.60 | 0.77 | 0.77 | 0.92 | 0.80 | 0.96 |
| XGBoost | 0.67 | 0.84 | 0.42 | 0.25 | 0.77 | 0.61 | 0.86 | 0.78 | 1.00 | 0.73 | 0.96 |
| AdaBoost | 0.53 | 0.82 | 1.00 | 0.08 | 0.76 | 0.48 | 1.00 | 0.74 | 1.00 | 0.20 | 0.96 |
| Gradient Boosting | 0.63 | 0.81 | 0.37 | 0.25 | 0.73 | 0.60 | 0.77 | 0.77 | 0.84 | 0.73 | 0.94 |
| LightGBM | 0.69 | 0.82 | 0.42 | 0.25 | 0.73 | 0.63 | 0.86 | 0.79 | 1.00 | 0.80 | 0.94 |
| SVM | 0.63 | 0.72 | 0.50 | 0.33 | 0.58 | 0.58 | 0.63 | 0.67 | 0.76 | 0.86 | 0.92 |
| Ensemble 1 | 0.59 | 0.72 | 0.26 | 0.16 | 0.55 | 0.53 | 0.63 | 0.67 | 0.81 | 0.86 | 0.92 |
| Ensemble 2 | 0.67 | 0.80 | 0.46 | 0.50 | 0.76 | 0.66 | 0.72 | 0.74 | 0.91 | 0.73 | 0.91 |
| Ensemble 3 | 0.69 | 0.79 | 0.57 | 0.33 | 0.70 | 0.62 | 0.90 | 0.74 | 1.00 | 0.66 | 0.94 |
| Ensemble 4 | 0.73 | 0.84 | 0.66 | 0.33 | 0.77 | 0.65 | 0.95 | 0.79 | 1.00 | 0.73 | 0.96 |

As shown in Table 5.19, Ensemble 4 achieved the highest performance, with an accuracy of 0.735 and an AUC of 0.844, demonstrating particularly strong results for PI-RADS 4 and PI-RADS 5. Other models, including LightGBM and

Ensemble 3, also yielded competitive outcomes. However, performance for PI-RADS 3 remained consistently lower across all models, confirming it as the most challenging category to classify.

## 5.4 Prostate Cancer Detection

In this section, we present the results of the classification of patients into cancer and non-cancer categories. For each classifier and imaging modality, we report the average area under the AUC, accuracy, precision, recall, and F1-score for both classes. These metrics provide a comprehensive evaluation of model performance, allowing comparison across different classifiers and input images.

Table 5.20 summarizes the performance of different ML classifiers for distinguishing between cancer and non-cancer cases using ADC images.

Table 5.20: The results of cancer and non-cancer classification using ADC images

| Model | Acc | AUC | Non-Cancer | | | Cancer | | |
|---|---|---|---|---|---|---|---|---|
| | | | Precision | Recall | F1-Score | Precision | Recall | F1-Score |
| MLP | 0.83 | 0.87 | 0.84 | 0.91 | 0.87 | 0.83 | 0.70 | 0.76 |
| KNN | 0.72 | 0.74 | 0.76 | 0.82 | 0.79 | 0.65 | 0.56 | 0.60 |
| XGBoost | 0.80 | 0.74 | 0.73 | 0.91 | 0.81 | 0.75 | 0.44 | 0.56 |
| Decision Tree | 0.72 | 0.70 | 0.78 | 0.78 | 0.78 | 0.63 | 0.63 | 0.63 |
| Random Forest | 0.67 | 0.77 | 0.66 | 0.96 | 0.78 | 0.71 | 0.19 | 0.29 |
| SVM | 0.79 | 0.84 | 0.89 | 0.76 | 0.82 | 0.68 | 0.85 | 0.75 |
| LGBM | 0.76 | 0.81 | 0.78 | 0.87 | 0.82 | 0.73 | 0.59 | 0.65 |
| GradientBoosting | 0.74 | 0.83 | 0.75 | 0.87 | 0.80 | 0.70 | 0.52 | 0.60 |
| Ensemble 1 | 0.70 | 0.83 | 0.73 | 0.84 | 0.78 | 0.65 | 0.48 | 0.55 |
| Ensemble 2 | 0.73 | 0.83 | 0.72 | 0.93 | 0.81 | 0.78 | 0.41 | 0.54 |

The results in the Table 5.20 indicate that the multilayer perceptron and support vector machine models achieved the highest overall performance in distinguishing cancer from non-cancer cases, with the multilayer perceptron slightly outperforming in terms of accuracy (0.83) and AUC (0.87). Multilayer perceptron

also demonstrated balanced class-specific performance, with high precision, recall, and F1-scores for both non-cancer and cancer classes. Support Vector Machine showed a particularly strong recall for the cancer class (0.85), indicating better sensitivity in detecting positive cases. Traditional ensemble methods (Ensemble 1 and Ensemble 2) provided moderate improvements in AUC but had lower overall accuracy compared to individual top-performing models. Random Forest achieved very high precision for non-cancer cases (0.96) but performed poorly in detecting cancer, as reflected by the low recall (0.19), highlighting an imbalance in class performance. Overall, these results suggest that while tree-based methods are effective for certain metrics, neural network-based and support vector machine models provide more balanced and reliable performance across both classes.

Table 5.21 presents the classification performance of various ML models in differentiating cancer from non-cancer cases based on DWI images.

Table 5.21: The results of cancer and non-cancer classification using DWI images

| Model | Acc | AUC | Non-Cancer | | | Cancer | | |
|---|---|---|---|---|---|---|---|---|
| | | | Precision | Recall | F1-Score | Precision | Recall | F1-Score |
| MLP | 0.75 | 0.79 | 0.78 | 0.84 | 0.81 | 0.70 | 0.59 | 0.64 |
| KNN | 0.67 | 0.75 | 0.69 | 0.84 | 0.76 | 0.59 | 0.37 | 0.45 |
| XGBoost | 0.74 | 0.74 | 0.73 | 0.91 | 0.81 | 0.75 | 0.44 | 0.56 |
| Decision Tree | 0.62 | 0.59 | 0.70 | 0.71 | 0.70 | 0.50 | 0.48 | 0.49 |
| Random Forest | 0.68 | 0.72 | 0.67 | 0.96 | 0.79 | 0.75 | 0.22 | 0.34 |
| SVM | 0.75 | 0.82 | 0.83 | 0.76 | 0.79 | 0.65 | 0.74 | 0.69 |
| LGBM | 0.71 | 0.71 | 0.71 | 0.89 | 0.79 | 0.69 | 0.41 | 0.51 |
| GradientBoosting | 0.74 | 0.78 | 0.72 | 0.96 | 0.82 | 0.83 | 0.37 | 0.51 |
| Ensemble 1 | 0.80 | 0.85 | 0.79 | 0.93 | 0.86 | 0.84 | 0.59 | 0.69 |
| Ensemble 2 | 0.71 | 0.84 | 0.71 | 0.91 | 0.79 | 0.71 | 0.37 | 0.49 |

Table 5.21 presents the performance of different classifiers for distinguishing cancer from non-cancer cases using DWI images. Ensemble 1 achieved the best overall performance, with an accuracy of 0.80 and an AUC of 0.85, demonstrating strong discriminative power. Multilayer perceptron and support vector machine also performed well, providing balanced precision, recall, and F1-scores for both

classes. Tree-based models such as Random Forest and Gradient Boosting showed high precision for non-cancer cases but relatively low recall for cancer, indicating that they are better at identifying negative cases than positive ones. Ensemble 2 offered improved AUC compared to several individual models but did not surpass Ensemble 1 in overall accuracy. These results suggest that combining imaging features with ensemble approaches can enhance classification reliability for DWI-based prostate cancer detection.

Table 5.22 shows the performance of different classifiers for distinguishing cancer and non-cancer cases using T2W axial images. The table reports accuracy, AUC, and the precision, recall, and F1-scores for both classes.

Table 5.22: The results of cancer and non-cancer classification using T2W axial images

| Model | Acc | AUC | Non-Cancer | | | Cancer | | |
|---|---|---|---|---|---|---|---|---|
| | | | Precision | Recall | F1-Score | Precision | Recall | F1-Score |
| MLP | 0.64 | 0.70 | 0.70 | 0.72 | 0.71 | 0.52 | 0.50 | 0.51 |
| KNN | 0.59 | 0.63 | 0.63 | 0.81 | 0.71 | 0.42 | 0.23 | 0.29 |
| XGBoost | 0.71 | 0.72 | 0.70 | 0.92 | 0.80 | 0.73 | 0.36 | 0.48 |
| Decision Tree | 0.67 | 0.65 | 0.73 | 0.75 | 0.74 | 0.57 | 0.55 | 0.56 |
| Random Forest | 0.62 | 0.70 | 0.63 | 0.94 | 0.76 | 0.50 | 0.09 | 0.15 |
| SVM | 0.69 | 0.81 | 0.76 | 0.72 | 0.74 | 0.58 | 0.64 | 0.61 |
| LGBM | 0.67 | 0.70 | 0.71 | 0.81 | 0.75 | 0.59 | 0.45 | 0.51 |
| GradientBoosting | 0.69 | 0.70 | 0.70 | 0.89 | 0.78 | 0.67 | 0.36 | 0.47 |
| Ensemble 1 | 0.74 | 0.78 | 0.72 | 0.94 | 0.82 | 0.82 | 0.41 | 0.54 |
| Ensemble 2 | 0.71 | 0.79 | 0.69 | 0.94 | 0.80 | 0.78 | 0.32 | 0.45 |

Regarding Table 5.22, the results show the classification performance of different models using T2W axial images. Ensemble 1 achieved the highest accuracy (0.74) and strong AUC (0.78), indicating reliable differentiation between cancer and non-cancer cases. Multilayer perceptron and support vector machine also demonstrated balanced performance across both classes, with good F1-scores. Tree-based models, such as Random Forest and Gradient Boosting, had high precision for non-cancer cases but lower recall for cancer, suggesting challenges in

detecting positive cases. Overall, ensemble approaches and neural network models provided the most consistent and robust classification results.

Table 5.23 presents the classification performance of various ML models on T2W coronal images. The metrics include accuracy, AUC, and precision, recall, and F1-scores for both cancer and non-cancer classes.

Table 5.23: The results of cancer and non-cancer classification using T2W coronal images

| Model | Acc | AUC | Non-Cancer | | | Cancer | | |
|---|---|---|---|---|---|---|---|---|
| | | | Precision | Recall | F1-Score | Precision | Recall | F1-Score |
| MLP | 0.69 | 0.76 | 0.74 | 0.78 | 0.76 | 0.60 | 0.56 | 0.58 |
| KNN | 0.67 | 0.60 | 0.68 | 0.89 | 0.77 | 0.62 | 0.30 | 0.40 |
| XGBoost | 0.68 | 0.68 | 0.69 | 0.80 | 0.74 | 0.55 | 0.41 | 0.47 |
| Decision Tree | 0.60 | 0.59 | 0.70 | 0.62 | 0.66 | 0.47 | 0.56 | 0.51 |
| Random Forest | 0.64 | 0.69 | 0.64 | 0.96 | 0.77 | 0.60 | 0.11 | 0.19 |
| SVM | 0.78 | 0.83 | 0.84 | 0.80 | 0.82 | 0.69 | 0.74 | 0.71 |
| LGBM | 0.68 | 0.74 | 0.71 | 0.82 | 0.76 | 0.60 | 0.44 | 0.51 |
| GradientBoosting | 0.68 | 0.70 | 0.70 | 0.87 | 0.77 | 0.62 | 0.37 | 0.47 |
| Ensemble 1 | 0.75 | 0.81 | 0.74 | 0.91 | 0.82 | 0.76 | 0.48 | 0.59 |
| Ensemble 2 | 0.71 | 0.82 | 0.71 | 0.91 | 0.79 | 0.71 | 0.37 | 0.49 |

Based on Table 5.23, the classification results using T2W coronal images indicate that Ensemble 1 achieved the highest accuracy (0.75) and AUC (0.81), reflecting strong overall performance in differentiating cancer from non-cancer cases. support vector machine and multilayer perceptron also performed well, showing balanced precision, recall, and F1-scores across both classes. Tree-based models, including Random Forest and Decision Tree, demonstrated high precision for non-cancer cases but lower recall for cancer, suggesting limited sensitivity to positive cases. Ensemble 2 improved AUC compared to several individual models but did not surpass Ensemble 1 in accuracy. Overall, these results highlight that ensemble strategies and neural network models provide more reliable and consistent classification outcomes.

Table 5.24 shows the performance of different ML classifiers for cancer versus

non-cancer classification using T2W sagittal images. The results include overall accuracy, AUC, and class-specific precision, recall, and F1-scores for both cancer and non-cancer groups.

Table 5.24: The results of cancer and non-cancer classification using T2W sagittal images

| Model | Acc | AUC | Non-Cancer | | | Cancer | | |
|---|---|---|---|---|---|---|---|---|
| | | | Precision | Recall | F1-Score | Precision | Recall | F1-Score |
| MLP | 0.70 | 0.76 | 0.70 | 0.91 | 0.79 | 0.71 | 0.37 | 0.49 |
| KNN | 0.75 | 0.75 | 0.73 | 0.93 | 0.82 | 0.80 | 0.44 | 0.57 |
| XGBoost | 0.73 | 0.72 | 0.73 | 0.91 | 0.81 | 0.75 | 0.44 | 0.56 |
| Decision Tree | 0.58 | 0.57 | 0.68 | 0.59 | 0.63 | 0.45 | 0.56 | 0.50 |
| Random Forest | 0.66 | 0.73 | 0.66 | 0.95 | 0.78 | 0.71 | 0.19 | 0.29 |
| SVM | 0.75 | 0.82 | 0.80 | 0.80 | 0.80 | 0.67 | 0.67 | 0.67 |
| LGBM | 0.71 | 0.74 | 0.72 | 0.86 | 0.78 | 0.67 | 0.44 | 0.53 |
| GradientBoosting | 0.72 | 0.74 | 0.73 | 0.86 | 0.79 | 0.68 | 0.48 | 0.57 |
| Ensemble 1 | 0.74 | 0.85 | 0.72 | 0.95 | 0.82 | 0.84 | 0.41 | 0.55 |
| Ensemble 2 | 0.74 | 0.82 | 0.72 | 0.95 | 0.82 | 0.84 | 0.41 | 0.55 |

Regarding Table 5.24, the results show the classification performance of different models using T2W sagittal images. Ensemble 1 achieved the highest AUC (0.85) with an accuracy of 0.74, indicating strong capability in distinguishing cancer from non-cancer cases. k-Nearest Neighbor and support vector machine also showed high overall accuracy and balanced class-specific performance. Tree-based models, such as Random Forest and Decision Tree, performed well for non-cancer precision but had lower recall for cancer, reflecting limitations in detecting positive cases. Overall, the results suggest that ensemble models and neural networks provide the most consistent and reliable classification outcomes for sagittal T2W images.

# Chapter 6

# Discussion, Conclusions, and Outlook

## 6.1 Discussion

### 6.1.1 Zonal Segmentation

Segmentation of prostate zones, particularly the peripheral zone, is critical for accurate diagnosis and effective management of prostate cancer. This study investigates three advanced neural network-based approaches, ensemble learning, MetaNet, and YOLO-V8, designed to address the challenges of zonal segmentation in prostate MRI scans from the Prostate158 dataset. While each method offers unique strengths, they share the overarching goal of improving segmentation accuracy and reliability, ultimately supporting better clinical decision-making and patient outcomes.

The ensemble learning approach employed in this study combines multiple U-Net-based models, effectively averaging their outputs to enhance segmentation performance in both the central gland and peripheral zone. This underscores the benefit of leveraging diverse models to accommodate anatomical variability in prostate MRI data.

MetaNet, adapted here for segmentation, represents a novel application of this

architecture. Its strong performance demonstrates the capacity of advanced neural networks to capture complex feature relationships in medical imaging. Although originally developed for other tasks, MetaNet's success in this context highlights its potential and encourages further exploration of state-of-the-art neural network architectures for medical image analysis.

The performance of three models, Ensemble Model, Meta-Net (Vanilla-Net + V-Net), and YOLO-V8, was assessed in both the central gland and peripheral zone regions using IoU and DSC metrics. In the central gland, YOLO-V8 achieved the highest performance, with an IoU of 80% and a DSC of 89%, followed by the Ensemble Model (IoU: 79.3%, DSC: 88.4%) and Meta-Net (IoU: 78%, DSC: 88%). In the peripheral zone, YOLO-V8 again led the results, attaining an IoU of 58% and DSC of 73%, while the Ensemble Model and Meta-Net showed similar performance, with IoUs of 54.5% and 54% and DSCs of 70.5% and 71%, respectively. Figure 6.1 provides a comparative overview of the DSC outcomes on the test set.

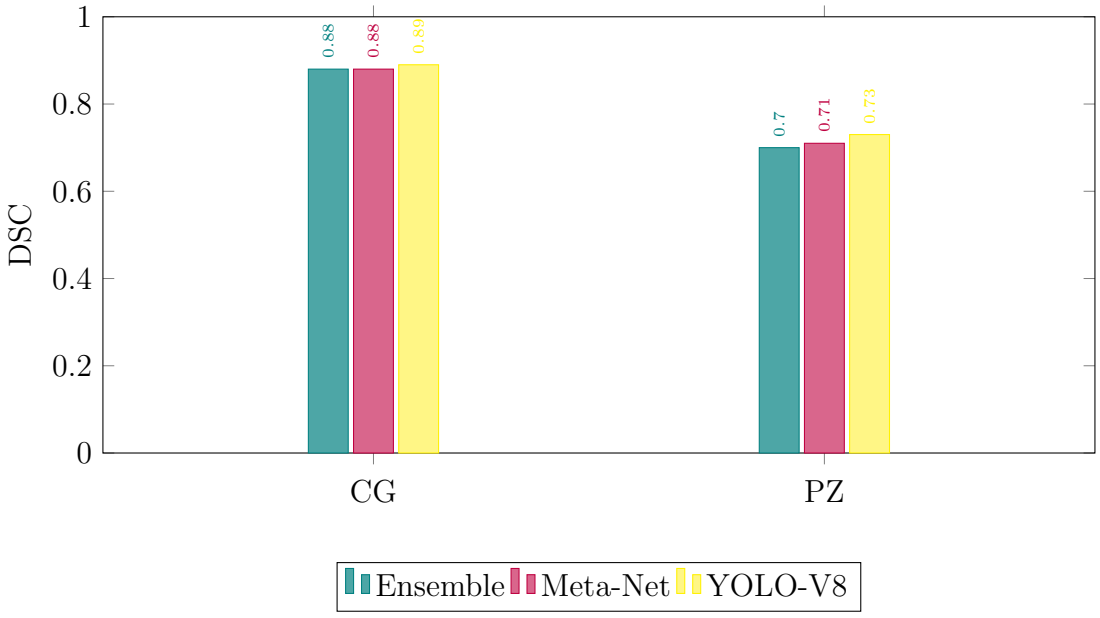



Figure 6.1: Comparison of DSC performance for Prostate Zone Segmentation on the test set

YOLO-V8 demonstrated strong potential, showcasing its ability to perform fast

and accurate prostate zone segmentation. Its architecture, optimized for segmentation rather than traditional object detection, offers a flexible approach applicable across various medical imaging tasks. Adapting YOLO for zonal segmentation highlights its versatility beyond conventional object detection applications.

The models were trained on a dataset of 90 MRI scans, with evaluation performed on a validation set of 25 scans and a test set of 24 scans. For central gland segmentation on the validation set, the ensemble learning method achieved an IoU of 80.4% and a DSC of 89.1%, while for peripheral zone segmentation, it reached an IoU of 60.6% and a DSC of 75.5%. On the test set, the ensemble method recorded an IoU of 79.3% and a DSC of 88.4% for the central gland, and an IoU of 54.5% and a DSC of 70.5% for the peripheral zone.

For Meta-Net, various network combinations were tested on both validation and test sets. The Vanilla-Net + V-Net combination performed best, achieving an IoU of 80% for the central gland and 58% for the peripheral zone, along with DSCs of 89% and 73% on the validation set. On the test set, this configuration maintained strong performance with an IoU of 78% and 54% for the central gland and the peripheral zone, respectively, and DSCs of 88% and 71%.

YOLO-V8 achieved an IoU of 81% for central gland and 60% for the peripheral zone on the validation set, with corresponding DSCs of 91% and 74%. On the test set, performance remained strong, with IoUs of 80% for the central gland and 58% for the peripheral zone and DSCs of 89% and 73%.

When compared with previous studies, it is important to note that our dataset is smaller than those used in most referenced works, which naturally affects performance metrics. Despite this limitation, our results demonstrate competitive efficacy in prostate zone segmentation. For instance, ENet achieved DSC scores of 91% for the whole gland, 87% for the transition zone, and 71% for the peripheral zone [110]. In comparison, our ensemble method reached 89.1% for the central gland and 75.5% for the peripheral zone on the validation set, showing particularly strong central gland segmentation.

Similarly, BASC-Net reported DSCs of 88.6% for the central gland and 79.9% for the peripheral zone [111]. Our results are closely aligned in central gland

performance, with slightly lower peripheral zone accuracy, which is notable given our smaller and less diverse dataset. Meta-Net also demonstrated competitive results with an IoU of 80% for the central gland and a DSC of 89%, further confirming the robustness of our approach.

These findings contribute to the growing body of literature on effective prostate segmentation techniques. They demonstrate that, even with a limited dataset, high-quality segmentation can be achieved through innovative model architectures and training strategies. Overall, our results are comparable to some of the best reported in the literature, highlighting the feasibility of strong segmentation performance under resource constraints. This suggests that future research can focus on optimizing segmentation methods without relying solely on large datasets, paving the way for effective prostate imaging solutions even when data availability is limited.

Figures 6.2 and 6.3 show a comparison of our proposed models with previous studies for central gland and peripheral zone segmentation, respectively. The figures illustrate model performance on the test set, focusing on DSC values, since most earlier studies did not report IoU.

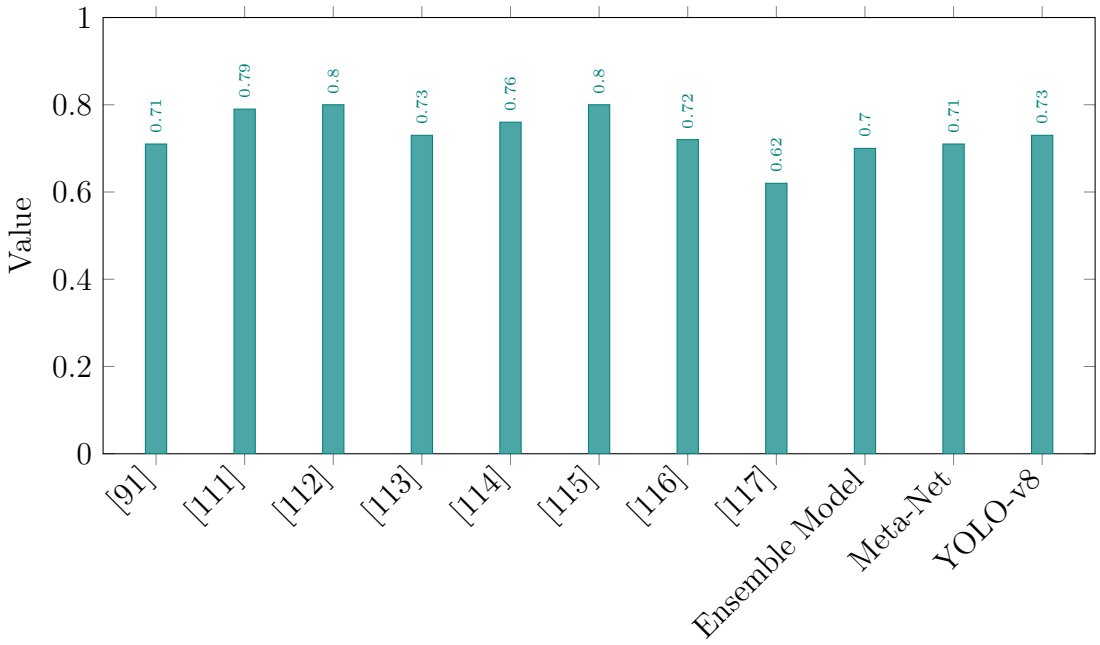



Figure 6.2: Comparison of the DSC results for CG segmentation, obtained from related works and the model of our study on the test set

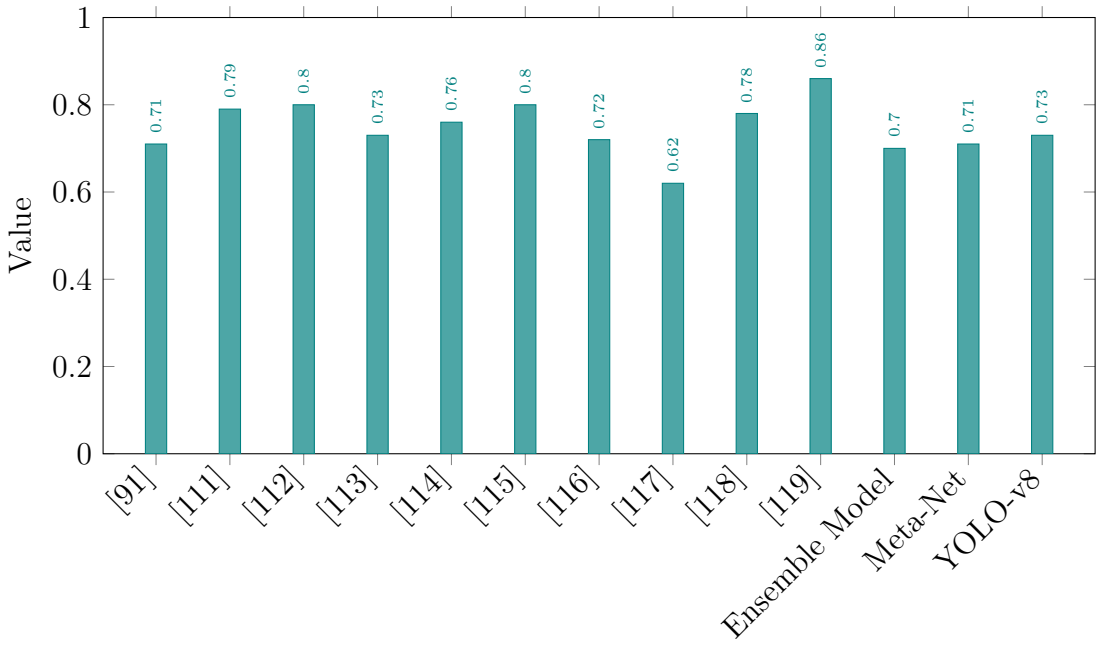



Figure 6.3: Comparison of the DSC results for PZ segmentation, obtained from related works and the model of our study on the test set

### 6.1.2 Lesion Segmentation

Segmentation of the lesion on prostate magnetic resonance imaging is a critical step in the early detection and diagnosis of clinically significant prostate cancer. High-precision segmentation ensures accurate localization of suspicious regions, facilitating biopsy targeting, treatment monitoring, and the development of computer-aided diagnostic tools. Owing to the heterogeneous appearance of prostate lesions, multi-parametric MRI (mpMRI) has become the standard approach to improve lesion characterization. In this study, three complementary MRI sequences, T2W, ADC, and DWI, were employed to enhance segmentation performance. Together, these modalities integrate anatomical and functional information, improving the differentiation between malignant and benign tissues.

In contrast to many existing studies that rely exclusively on publicly available datasets, we developed a custom dataset to more accurately reflect real-world

variations and challenges in prostate lesion segmentation. The dataset was meticulously annotated to ensure high-quality ground truth labels for both model training and evaluation.

To investigate the potential of DL for this task, four DL architectures, U-Net, Dense U-Net, Attention U-Net, and a custom LSTM U-Net, were explored for the segmentation of prostate lesions from mpMRI. These models were selected to evaluate how different architectural strategies capture spatial and contextual features. Two input strategies were used: training networks individually in single modalities (T2W, ADC, DWI) and multi-input U-Net–based models processing all three sequences together. ADC proved to be the most discriminative single modality, which led to the development of the LSTM U-Net incorporating a bidirectional ConvLSTM module at the bottleneck to better capture spatio-temporal dependencies and improve feature representation.

Since lesion masks were provided separately for each modality, a unified ground truth was created by merging the T2W, ADC, and DWI masks using a pixel-wise logical OR. This approach ensured that any voxel labeled as a lesion was included in at least one modality, integrating complementary information between sequences. To evaluate the generalization of the model, training was performed in two dataset subsets: one including all PI-RADS scores (3–5) and another limited to clinically significant lesions (4–5). This allowed for a performance assessment on both lower and higher-risk lesions.

In general, these findings demonstrate that the integration of multiple MRI sequences, the selection of appropriate network architectures, and the carefully designed ground truth masks are critical to achieving robust and accurate segmentation of prostate lesions. The results underscore the potential of DL, particularly architectures incorporating spatio-temporal modeling, to support improved detection and characterization of clinically significant prostate cancer.

The results of our lesion segmentation experiments reveal several key insights regarding the performance of different DL architectures and MRI sequences. When considering clinically significant lesions (PI-RADS 4–5), the highest segmentation accuracy was achieved using the ADC sequence, with Dense-U-Net reaching 69%,

followed closely by U-Net (68%) and LSTM U-Net (67%). DWI sequences produced slightly lower performance (up to 66%), while T2W alone yielded the lowest accuracy ( 48%). Interestingly, the combination of all three sequences did not consistently improve performance, with multi-sequence inputs achieving only moderate accuracy (41–58%). Similar trends were observed in the dataset, including all PI-RADS scores (3–5), highlighting the challenge of segmenting lower-grade and more heterogeneous lesions, which reduced overall performance across all architectures.

Comparing our findings with previous studies demonstrates both consistencies and differences. In line with our results, several works have reported superior performance using ADC images due to their higher lesion contrast. For instance, Simeth et al. [139] achieved Dice scores of 0.60 on internal datasets using ADC maps, and Gavade et al. [134] reached 67% Dice with U-Net and LSTM networks. Conversely, multi-sequence integration has shown variable results; some studies, such as Li et al. [141] and Li et al. [142], reported high DSCs of 84–92% when employing advanced attention-driven multi-scale networks, suggesting that more sophisticated fusion strategies or transformer-based approaches may better leverage complementary information from multiple sequences.

Several prior studies reported performances comparable to or slightly higher than ours on similar datasets. For example, Wu et al. [138] achieved Dice scores of 51–44% on Prostate158 and PI-CAI using selective multi-modal integration, while Wang et al. [94] reported a Dice of 56% using a dual-pyramid network. In contrast, high-performing approaches such as Shiri et al. [146] and Li et al. [142] employed multi-center datasets and attention mechanisms to reach DSCs of 0.84–0.92, demonstrating the potential benefits of large, diverse datasets and advanced architectural modules. Methods relying solely on T2W or less optimized fusion strategies, such as Adams et al. [123] and Zhong et al. [140], achieved lower Dice scores of 31–45%, reflecting the limitations of single-modality inputs for challenging prostate lesion segmentation tasks.

Our results, particularly the performance of ADC and LSTM U-Net architectures, suggest that the incorporation of recurrent modules and the aggregation of

contextual characteristics can improve the segmentation of small or irregular lesions. However, the limited gain from multi-sequence inputs in our study indicates that simple concatenation or naive fusion may not fully exploit complementary information across T2W, ADC, and DWI sequences. This suggests that attention mechanisms, modality-specific weighting, or transformer-based integration, as demonstrated by SAM-UNETR [133] and LeSMI [138], may be necessary to achieve the highest segmentation accuracy in prostate MRI.

In summary, our study confirms the importance of ADC imaging for accurate lesion segmentation and highlights the potential of LSTM-based architectures for capturing spatiotemporal dependencies. While our Dice scores are comparable to many published methods using similar datasets, there remains a performance gap compared to advanced attention- or transformer-based models trained on larger, multi-center datasets. These findings emphasize the need for further exploration of multi-sequence fusion strategies, recurrent and attention modules, and dataset diversity to improve the robustness and clinical applicability of automated prostate lesion segmentation.

Figures 6.4 and 6.5 show the comparison of our networks for lesion segmentation using the PI-RADS 4–5 and PI-RADS 3–5 datasets, respectively.

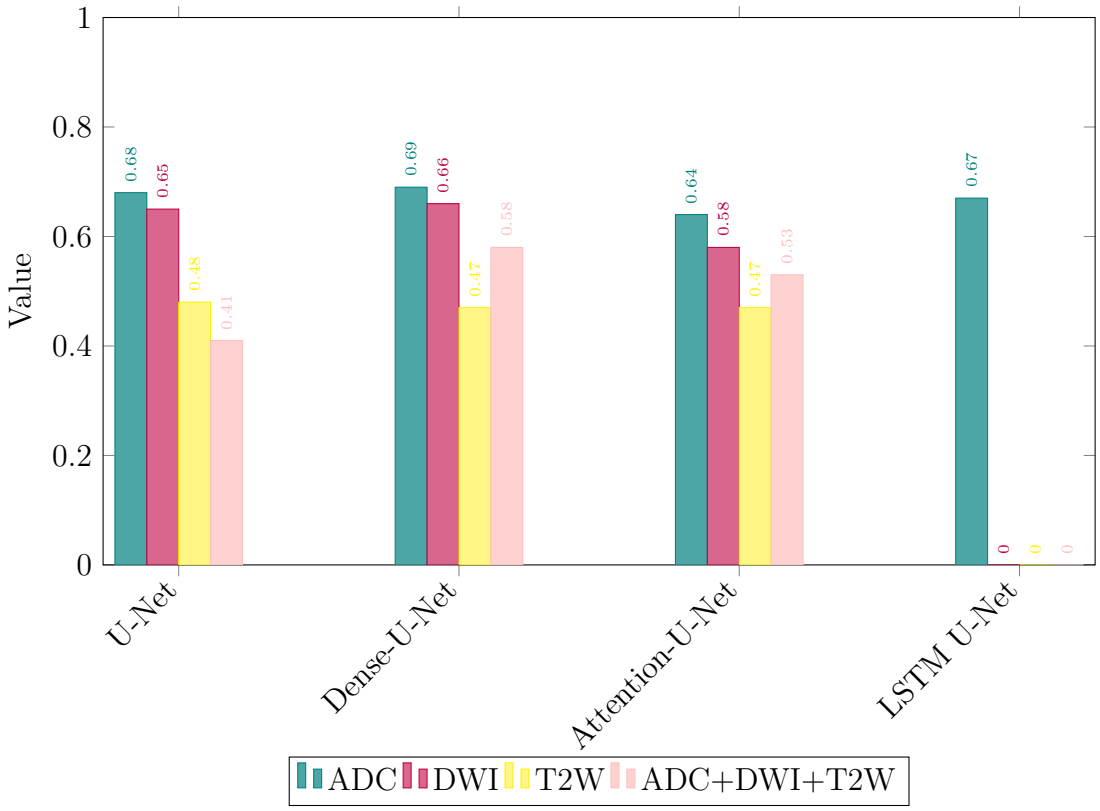



Figure 6.4: Lesion segmentation results using PI-RADS 4–5 dataset.

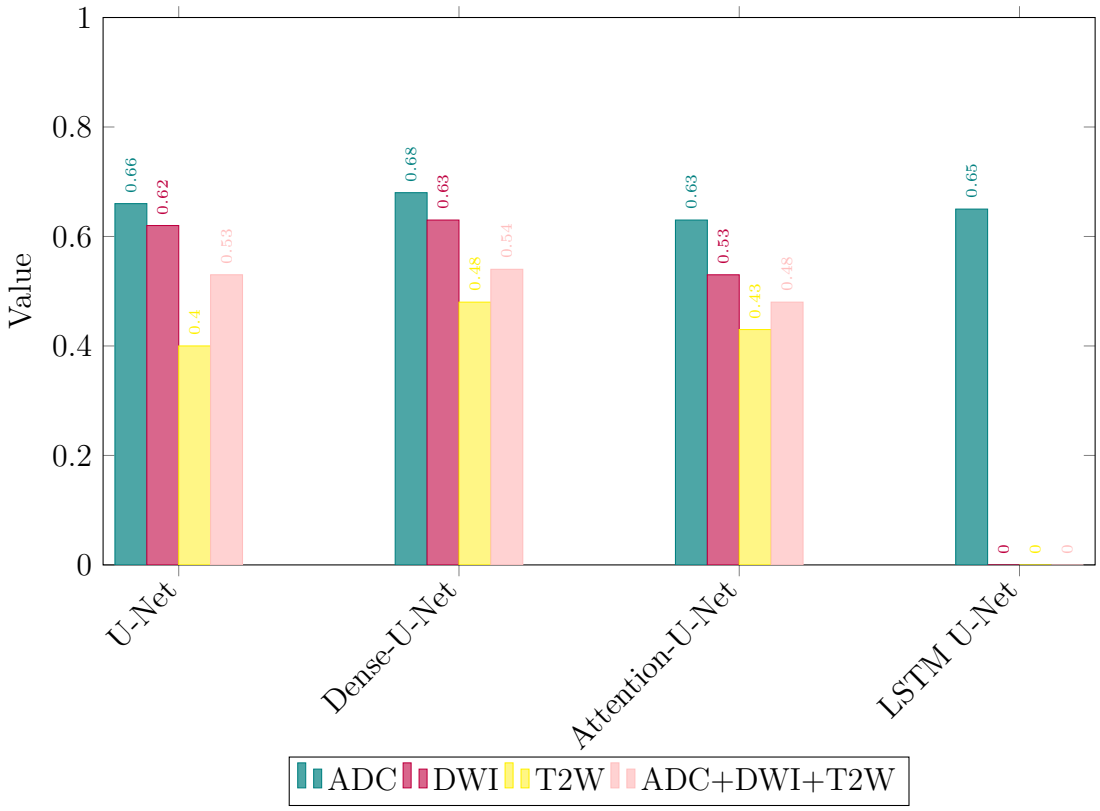



Figure 6.5: Lesion segmentation results using PI-RADS 3–5 dataset.

### 6.1.3 PI-RADS Classification

mp-MRI improves the detection of prostate tumors by integrating anatomical information from T2W images with quantitative data derived from DWI and dynamic contrast-enhanced (DCE) sequences [233]. The Prostate Imaging-Reporting and Data System (PI-RADS) v2, along with its updated version PI-RADS v2.1, offers a standardized framework for evaluating the likelihood of clinically significant prostate cancer based on lesion morphology and imaging characteristics, promoting consistency across different institutions [178]. Nevertheless, assigning a PI-RADS score still relies on subjective interpretation of qualitative lesion features, such as distinguishing heterogeneous T2W lesions (PI-RADS 3) from homogeneous ones (PI-RADS 4). This subjectivity introduces interobserver variability and potential disagreement among clinicians [212] [234].

In this study, we investigated several ML approaches for automated PI-RADS classification using radiomic features extracted from mpMRI sequences, including T2W, DWI, and ADC maps. Across the evaluated methods, models based on ADC features consistently outperformed those using DWI or T2W alone, with ensemble and gradient boosting models achieving the highest overall accuracy and AUC values.

Notably, PI-RADS 5 lesions were reliably classified across all models, whereas PI-RADS 3 lesions remained the most challenging, reflecting the intrinsic difficulty in differentiating intermediate-risk lesions. Figure 6.6 presents a comparison of AUCs for ML models using Feature Set 1. ADC-derived features consistently achieved the highest AUCs across most models, whereas T2W-based features generally performed the least effectively. Combining features from multiple sequences improved performance relative to single-sequence models, particularly for XGBoost, Gradient Boosting, and ensemble approaches. Ensemble models demonstrated competitive or superior AUCs compared to individual classifiers when using merged feature sets.

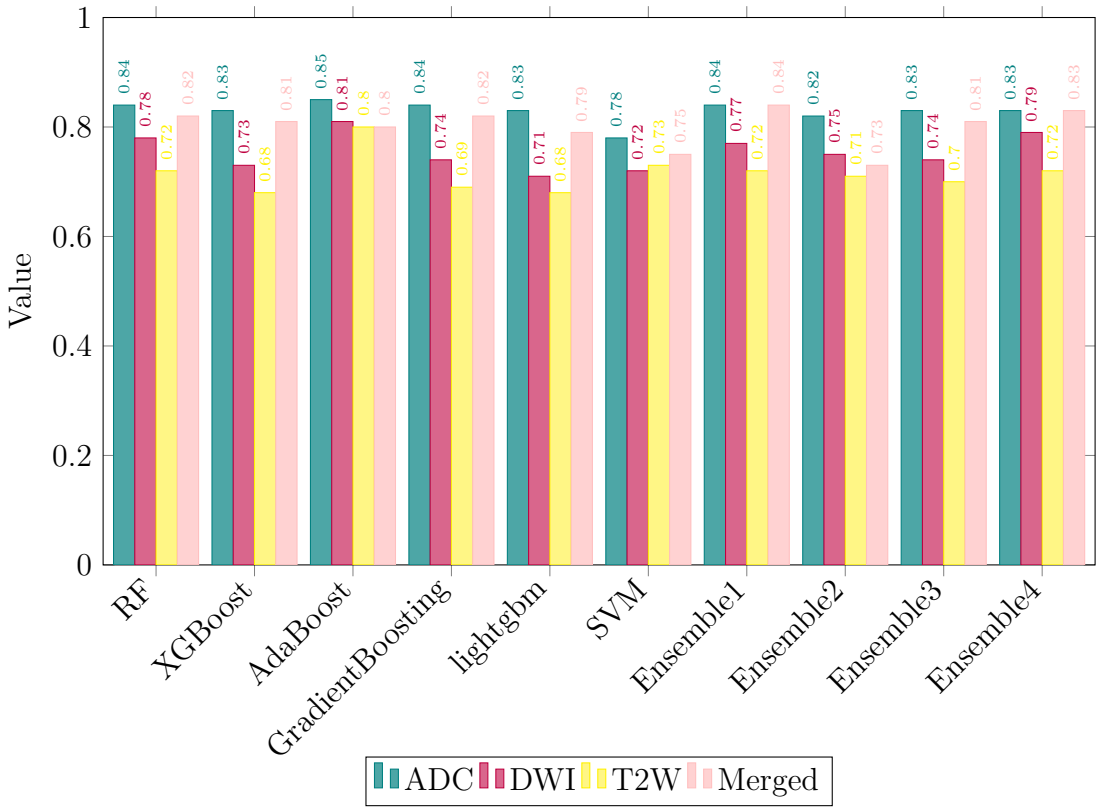



Figure 6.6: Comparison of the AUCs achieved by all models in Approach 1 using Feature Set 1 for PI-RADS classification

As illustrated in Figure 6.7, which compares the AUC values of all models under Approach 1 with Feature Set 2 for PI-RADS classification, ADC demonstrates consistently superior performance across the majority of models. The merged feature set also yields strong results, especially within ensemble methods such as Ensemble 1 and Ensemble 4. In contrast, DWI and T2W achieve comparatively lower and more fluctuating AUCs, reflecting weaker discriminative capability. The use of ensemble strategies generally enhances both robustness and classification accuracy. Overall, the figure underscores the leading role of ADC and the advantages of ensemble modeling.

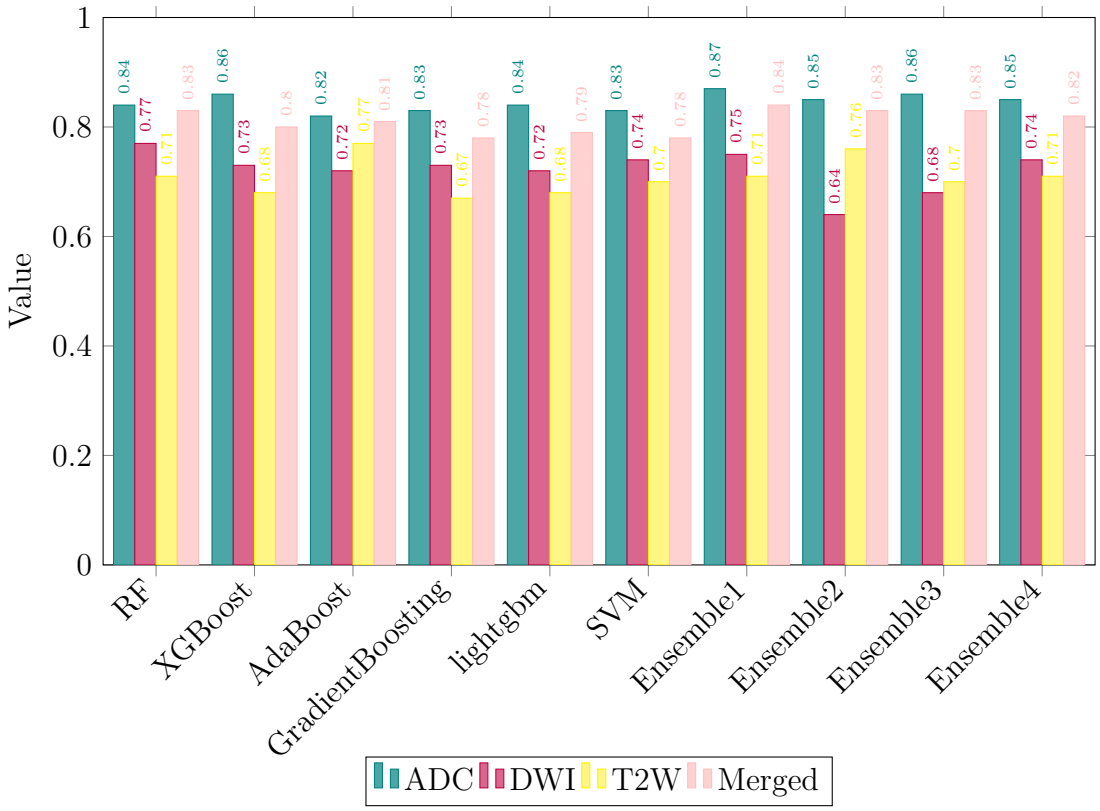



Figure 6.7: Comparison of the AUCs achieved by all models in Approach 1 using Feature Set 2 for PI-RADS classification

Figures 6.8 and 6.9 illustrate the AUC values obtained by all models in Approach 2 with Feature Set 1 and Feature Set 2, respectively. Model performance appears generally consistent across both sets, with most AUCs concentrated near 0.80. In Figure 6.8 (Feature Set 1), GradientBoosting and Ensemble 4 achieve the highest scores, whereas in Figure 6.9 (Feature Set 2), performance is more evenly distributed, although Ensemble 2 shows a marked decline. In both scenarios, support vector machine performs poorly, indicating limited suitability for ADC-driven inputs. Overall, the findings suggest that while both feature sets are effective, Feature Set 1 provides slightly better model separation and peak outcomes, especially for boosting and ensemble approaches.

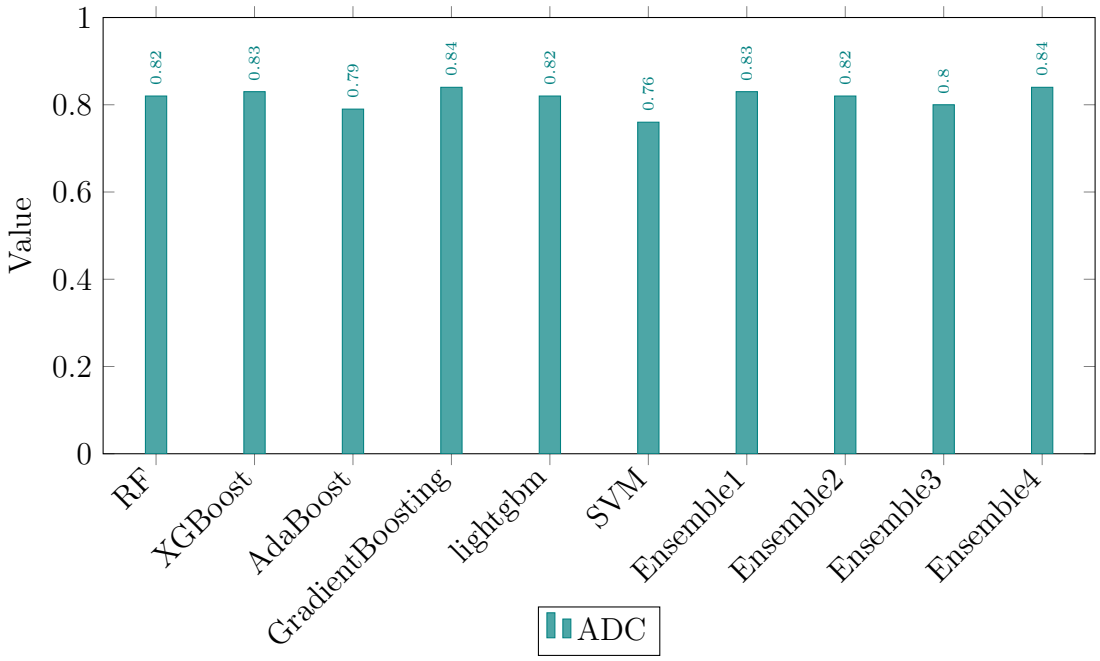



Figure 6.8: Comparison of the AUCs achieved by all models in Approach 2 using Feature Set 1 for PI-RADS classification

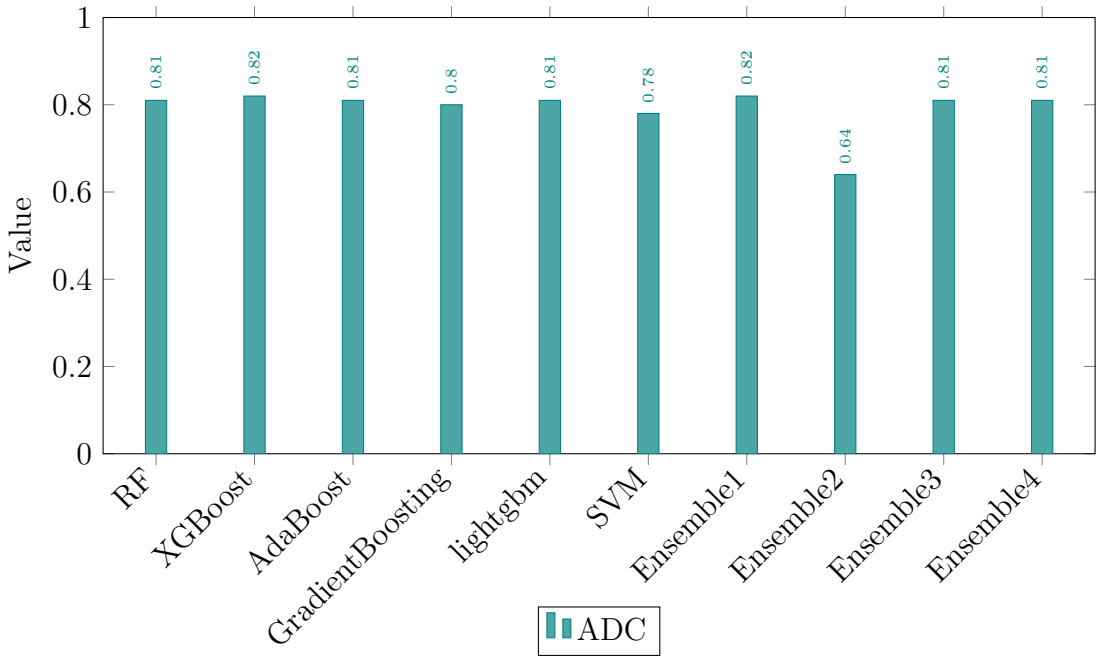



Figure 6.9: Comparison of the AUCs achieved by all models in Approach 2 using Feature Set 2 for PI-RADS classification

Figure 6.10 presents the AUCs of different models in Approach 3 using ADC features for PI-RADS classification. Tree-based methods such as RF and XGBoost consistently achieve strong results, with AUCs exceeding 0.83. Ensemble 4 outperforms all other models, highlighting the effectiveness of its combination strategy in this approach. Conversely, support vector machine and Ensemble 1 demonstrate the lowest performance, both falling below 0.74. These findings indicate that while some ensemble methods may not generalize well, a well-designed ensemble like Ensemble 4 can substantially enhance predictive accuracy.

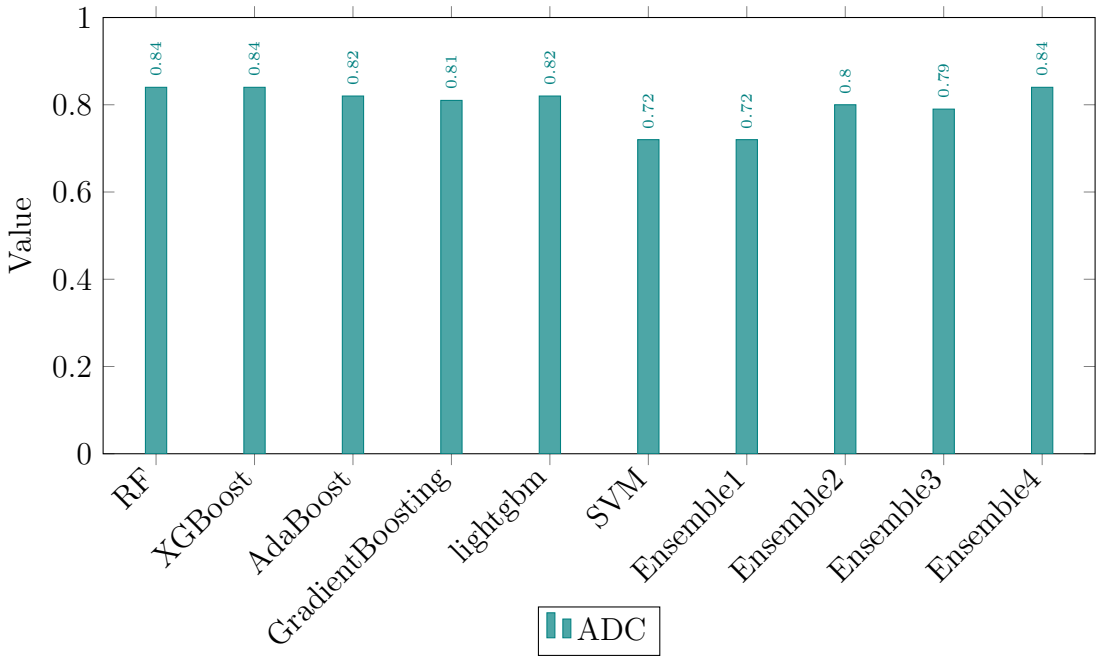


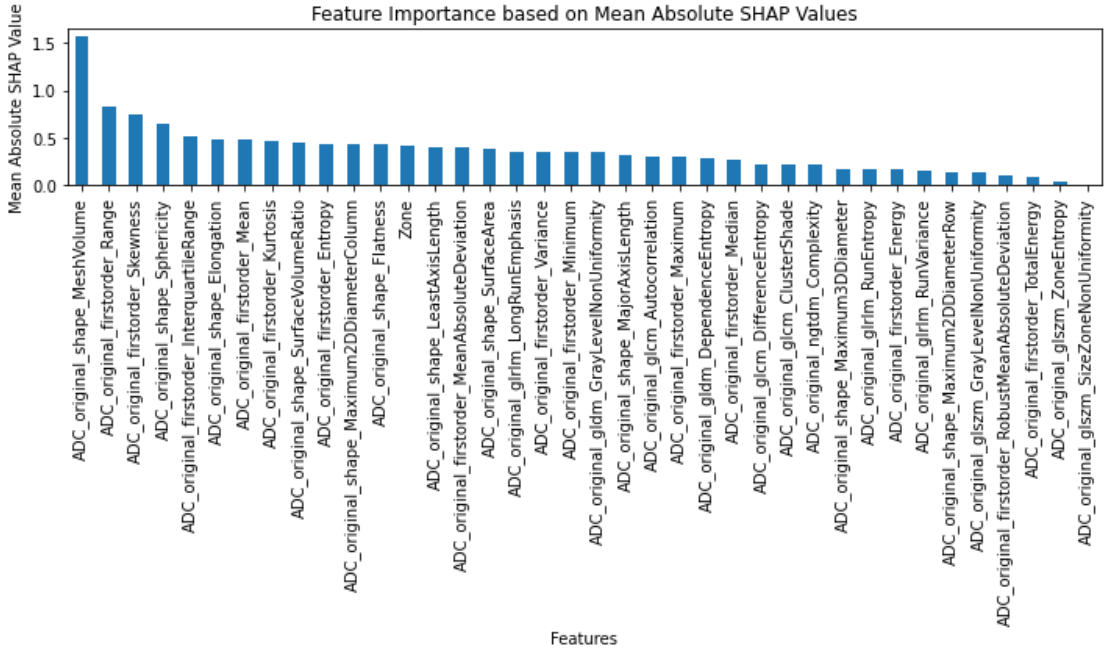

Figure 6.10: Comparison of the AUCs achieved by all models in Approach 3

To understand feature contributions and enhance model interpretability, SHAP (SHapley Additive exPlanations) was utilized. SHAP provides a unified framework for explaining model predictions, rooted in cooperative game theory, by assigning each feature a quantifiable importance for individual predictions [235]. Its consistency and local accuracy have made it a widely adopted standard for model explainability. However, applying SHAP to complex ensemble models, particularly stacked ensembles, can be computationally demanding or impractical due to their intricate structure and limited direct SHAP support [236]. Consequently, XGBoost, chosen for its consistently high AUCs and its central role within the ensemble models, was selected for SHAP analysis.

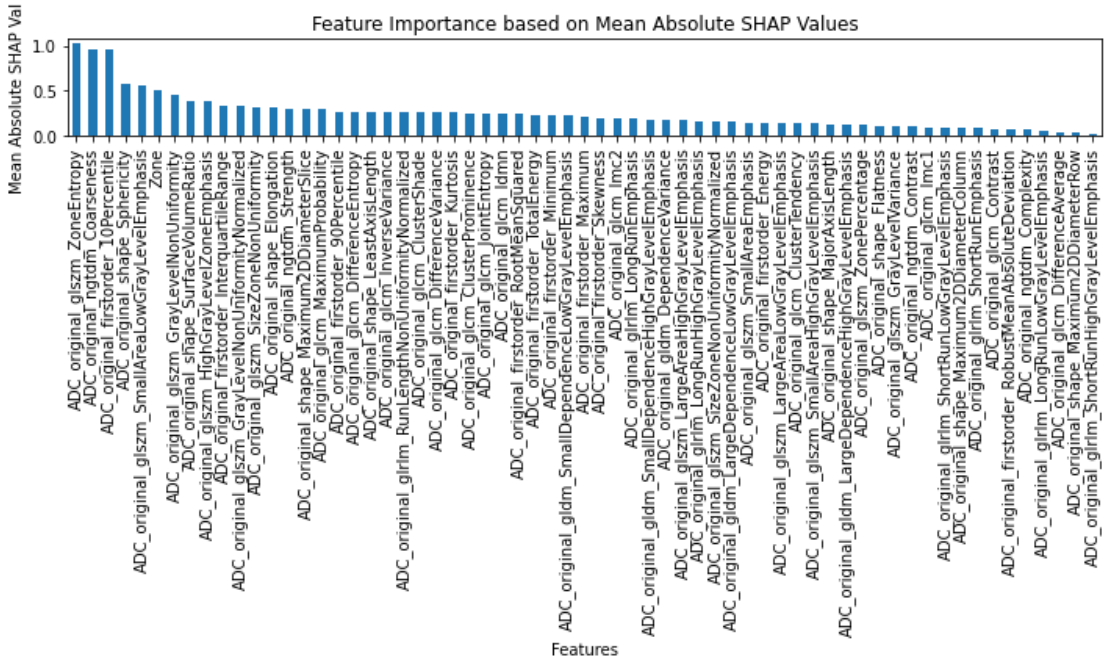
Figures 6.11 and 6.12 present the SHAP-based interpretability results obtained from XGBoost using ADC features in Approach 1. Figure 6.11 displays feature importance plots derived from the mean absolute SHAP values for both Feature Set 1 and Feature Set 2, highlighting the features that contribute most on average to model predictions. Figure 6.12 depicts the mean SHAP value plots, capturing both the direction and magnitude of each feature's impact on the model output

across all classes for the same feature sets.

These visualizations summarize the average effect of each feature on the model's predictions, effectively illustrating overall feature importance within the classification task.

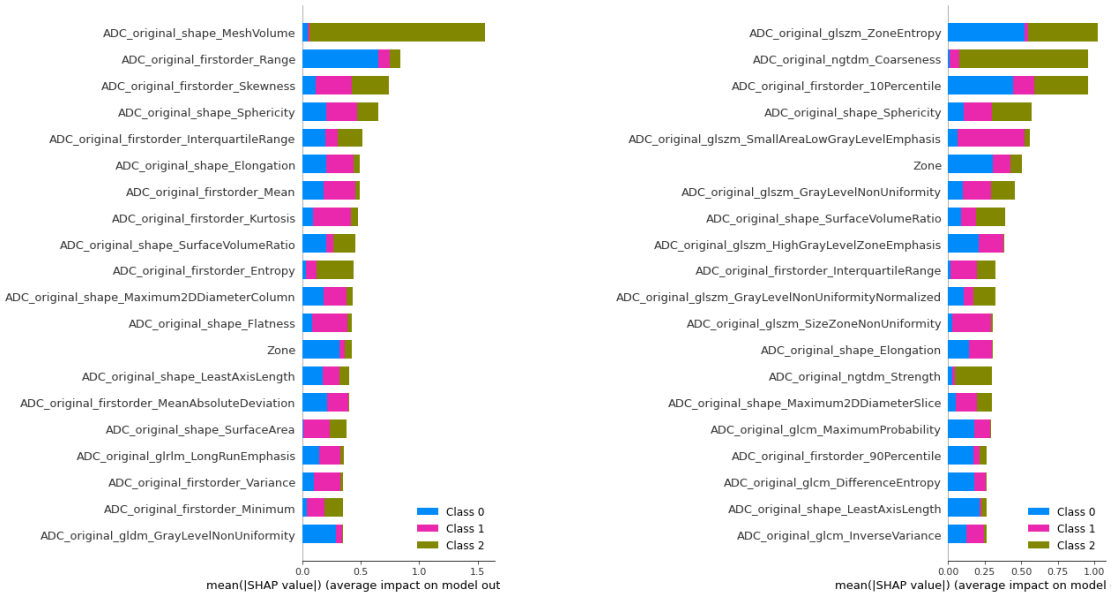



(a)




(b)

Figure 6.11: *Feature importance plots based on the mean absolute SHAP values, in a) approach 1 using feature set 1, and b) approach 1 using feature set 2*




(a) (b)

Figure 6.12: Feature importance plots based on the mean SHAP value based on the classes (Class 0: PI-RADS 3, Class 1: PI-RADS 4, and Class 3: PI-RADS 5), in a) approach 1 using feature set 1, and b) approach 1 using feature set 2

Based on Figure 6.12.a, for Class 0 (PI-RADS 3), MeshVolume and Range exhibit the strongest positive influence, highlighting their key role in identifying this category. For Class 1 (PI-RADS 4), features such as Skewness, Sphericity, and Elongation show notable impact, suggesting that the model leverages shape irregularities to distinguish this group. In Class 2 (PI-RADS 5), Skewness and Mean are the most influential, reflecting intensity-related characteristics as primary indicators. Certain features, including Skewness and Range, contribute meaningfully across all classes, though with varying magnitudes. Overall, the model utilizes both shape and intensity-based features differently across classes to support PI-RADS classification.

According to Figure 6.12.b, features associated with Class 2 (PI-RADS 5) have the highest SHAP impact, with texture metrics such as ZoneEntropy and Coarseness being particularly influential. Features for Class 1 (PI-RADS 4), including first-order and small-area emphasis metrics, show moderate importance. Class 0 (PI-RADS 3) features exert lower but consistent effects across variables. These

results indicate that texture features derived from ADC images dominate model predictions, whereas shape and first-order features contribute less, emphasizing the critical role of texture heterogeneity in differentiating PI-RADS categories.

These observations are consistent with prior studies. Previous work has shown that DL models can achieve moderate agreement with radiologist-assigned PI-RADS scores, with the highest reliability observed for PI-RADS 5 lesions and lower performance for intermediate categories such as PI-RADS 3 [149] [150]. Radiomics-based approaches incorporating ADC and T2W features have demonstrated improved detection of clinically significant prostate cancer, particularly in high-risk lesions, with reported accuracies ranging from 75% to 84% and AUCs between 0.83 and 0.99 [151] [152] [153] [154] [156]. Our findings align with these results, reinforcing the utility of radiomic features and ensemble learning in enhancing automated classification of prostate lesions.

Importantly, while ADC-based models performed well, the persistent difficulty in accurately classifying PI-RADS 3 lesions highlights a key limitation of current automated pipelines. This challenge echoes prior reports noting that intermediate-risk lesions often require nuanced radiological judgment and clinical context for reliable stratification [149] [150] [151]. Therefore, automated models should be regarded as complementary tools to expert radiologist assessment rather than standalone solutions.

Overall, our results demonstrate that integrating radiomic features from multiple MRI sequences and employing ensemble learning techniques can enhance PI-RADS classification. High-risk lesions (PI-RADS 5) are identified reliably, whereas intermediate categories remain challenging. This underscores the need for further refinement in feature extraction, model training, and the integration of clinical data to achieve robust, clinically applicable performance.

This study has some limitations. First, the dataset contained a relatively limited number of MRI cases, which may constrain the generalizability of the findings. Second, PI-RADS 3 lesions remain particularly difficult to classify due to their intermediate and often ambiguous imaging characteristics. These lesions frequently

share overlapping features with both lower and higher PI-RADS categories, making accurate detection and classification challenging even with advanced ML approaches.

### 6.1.4 Prostate Cancer Detection

Figure 6.13 shows the comparison of accuracy and AUC values across different classifiers evaluated on ADC images

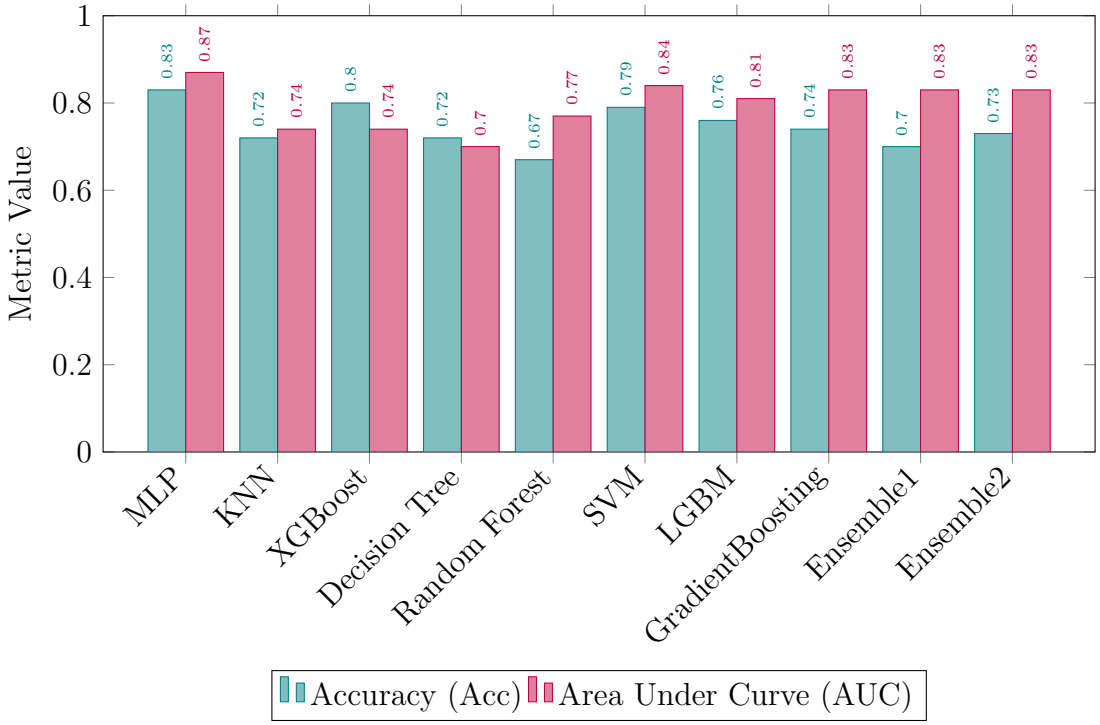



Figure 6.13: Comparison of Acc and AUC for different classifiers using ADC images.

The performance of the classifiers on ADC images varied, with multilayer perceptron achieving the highest accuracy (0.83) and AUC (0.87) among individual models. Support vector machine also showed strong performance, with an accuracy of 0.79 and an AUC of 0.84, closely followed by LGBM (accuracy 0.76, AUC 0.81) and Gradient Boosting (accuracy 0.74, AUC 0.83). Traditional classifiers such as k-Nearest Neighbor and Decision Tree yielded lower accuracy (0.72 each)

and relatively moderate AUC values (0.74 and 0.70, respectively). Random Forest showed a discrepancy between accuracy (0.67) and AUC (0.77), suggesting it may tend to misclassify some samples despite good overall ranking capability. Interestingly, ensemble approaches (Ensemble1 and Ensemble2) did not outperform the best individual models, reaching accuracy values of 0.70–0.73 and AUC values of 0.83. Overall, multilayer perceptron and support vector machine appear to be the most reliable classifiers for ADC-based evaluation, balancing both accuracy and discriminative ability.

Figure 6.14 illustrates the Acc and AUC for various classifiers evaluated on DWI images.

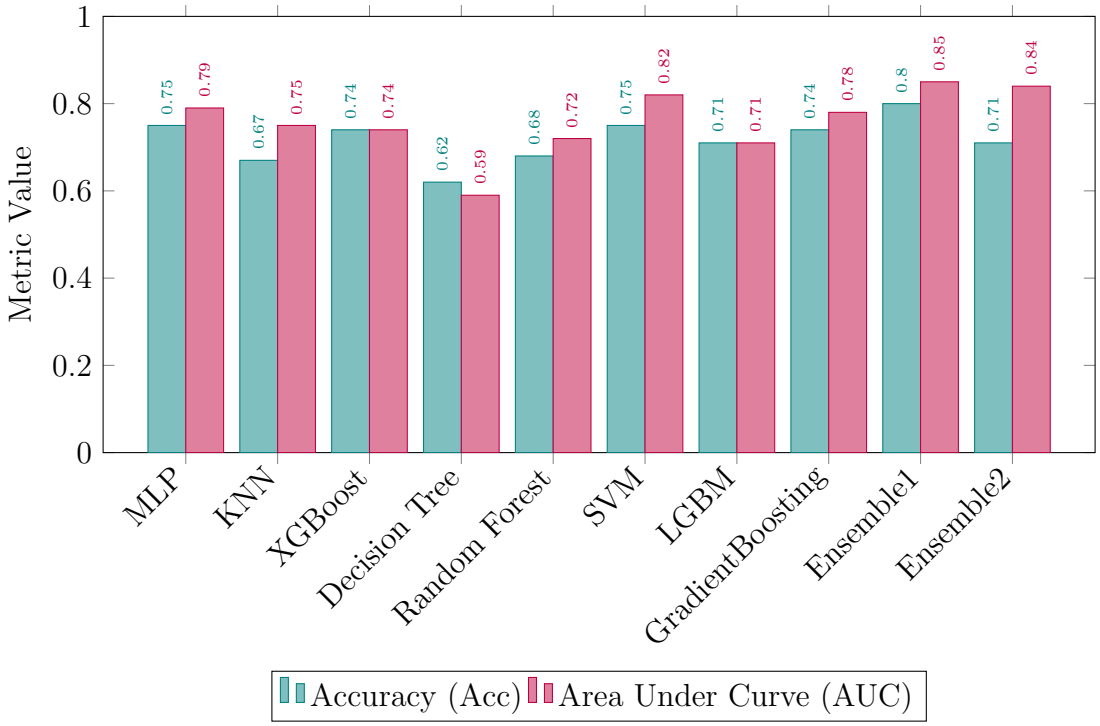



Figure 6.14: Comparison of Accuracy and AUC for different classifiers using DWI images.

For DWI images, the ensemble approaches showed the strongest performance, with Ensemble1 achieving the highest AUC (0.85) and the highest accuracy (0.80), while Ensemble2 also performed well (accuracy 0.71, AUC 0.84). Among individual

models, multilayer perceptron and support vector machine maintained competitive results, both reaching an accuracy of 0.75, with support vector machine slightly outperforming multilayer perceptron in AUC (0.82 vs. 0.79). Gradient Boosting and XGBoost demonstrated moderate performance, with accuracies of 0.74 and 0.74, respectively. The Decision Tree classifier again exhibited the lowest performance among all classifiers, with an accuracy of 0.62 and an AUC of 0.59. Overall, these results indicate that, unlike ADC images, where multilayer perceptron led, ensemble methods are particularly effective for DWI data, providing improved discriminative capability and robustness.

Figure 6.15 presents the Acc and AUC values for different classifiers evaluated on T2W axial images.

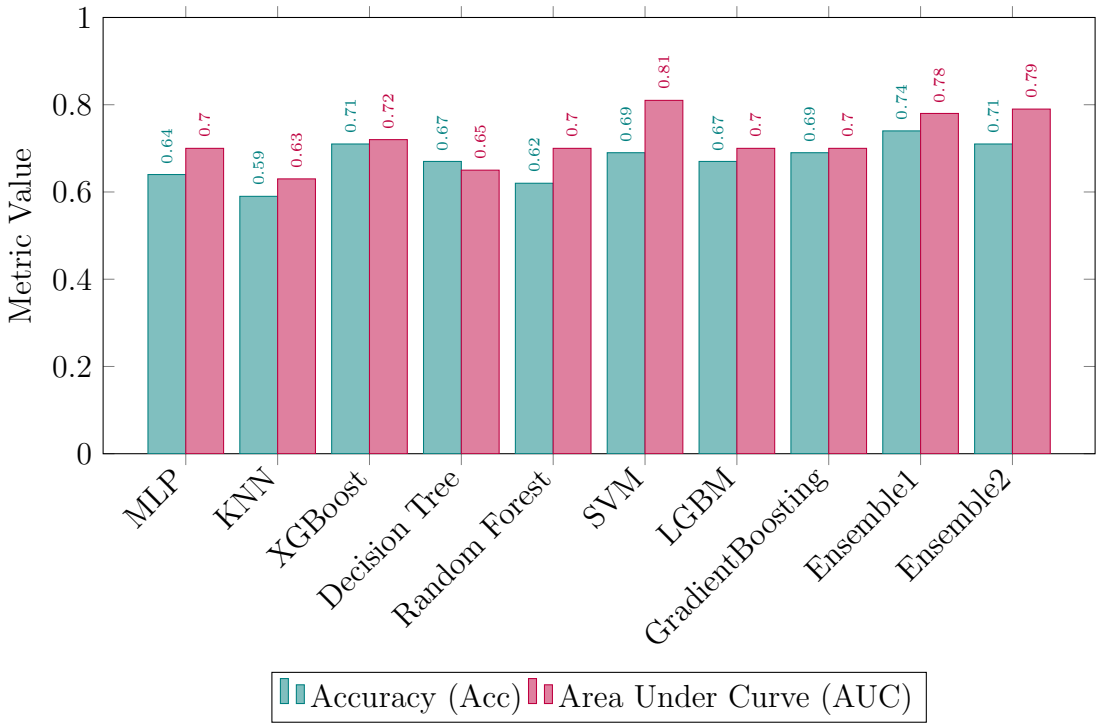



Figure 6.15: Comparison of Accuracy and AUC for different classifiers using T2W Axial images.

When evaluating T2W axial images, ensemble models again demonstrated

strong performance, with Ensemble1 and Ensemble2 achieving the highest accuracies (0.74 and 0.71) and AUCs (0.78 and 0.79), indicating reliable overall discriminative ability. Among single classifiers, the support vector machine stood out with a high AUC of 0.81, despite a moderate accuracy of 0.69, suggesting effective ranking of positive and negative cases even if some misclassifications occurred. XGBoost also performed relatively well (accuracy 0.71, AUC 0.72), while Decision Tree and Random Forest showed lower accuracy (0.67 and 0.62), though their AUC values remained moderate. Overall, T2W images appear to favor ensemble and support vector machine models, highlighting that combining predictions or using robust margin-based classifiers can improve performance on this modality.

Figure 6.16 displays the accuracy and AUC metrics for various classifiers applied to T2W coronal images.

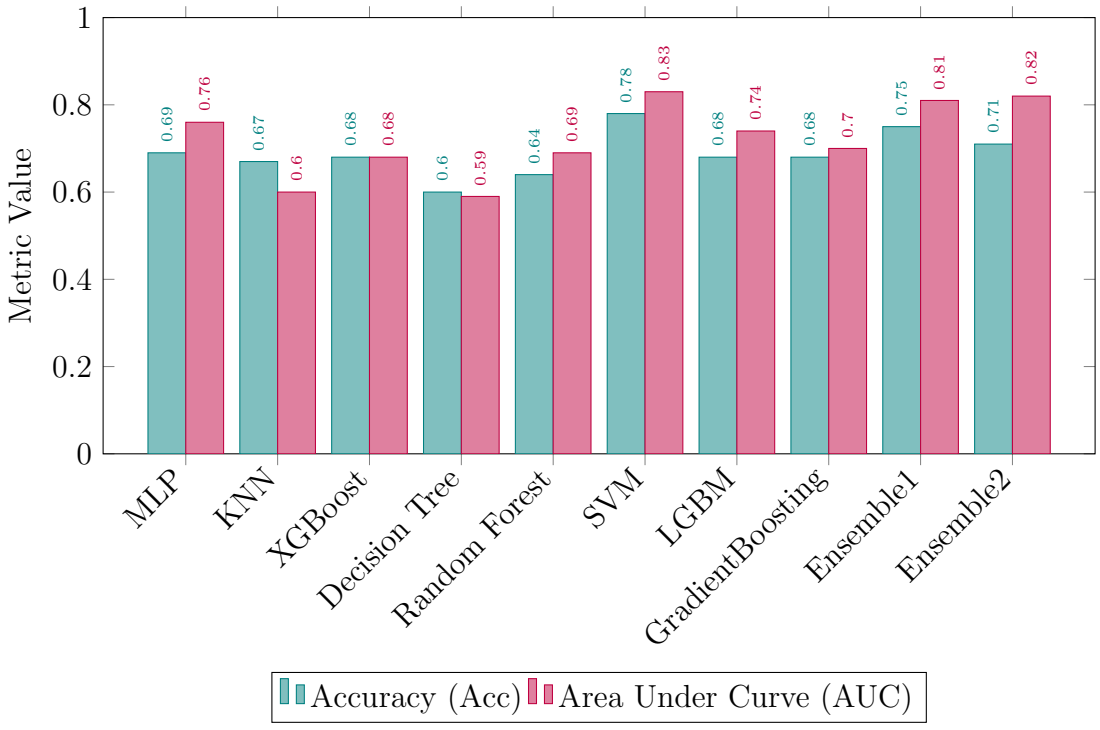



Figure 6.16: Comparison of Accuracy and AUC for different classifiers using T2W Coronal images.

For T2W coronal images, the support vector machine exhibited the highest

individual performance, achieving an accuracy of 0.78 and a strong AUC of 0.83, indicating robust discriminative ability. Ensemble models also performed very well, with Ensemble1 and Ensemble2 reaching accuracies of 0.75 and 0.71 and AUCs of 0.81 and 0.82, respectively, highlighting the benefit of combining multiple classifiers. Other individual models, such as multilayer perceptron and LGBM, showed moderate performance (accuracy  0.68–0.69, AUC 0.74–0.76), while Decision Tree and k-Nearest Neighbor lagged in both accuracy and AUC. Overall, these results suggest that for T2W coronal images, margin-based classifiers like support vector machine, as well as ensemble strategies, provide the most reliable predictive performance.

Figure 6.17 presents the Acc and AUC values for various classifiers evaluated on T2W sagittal images.

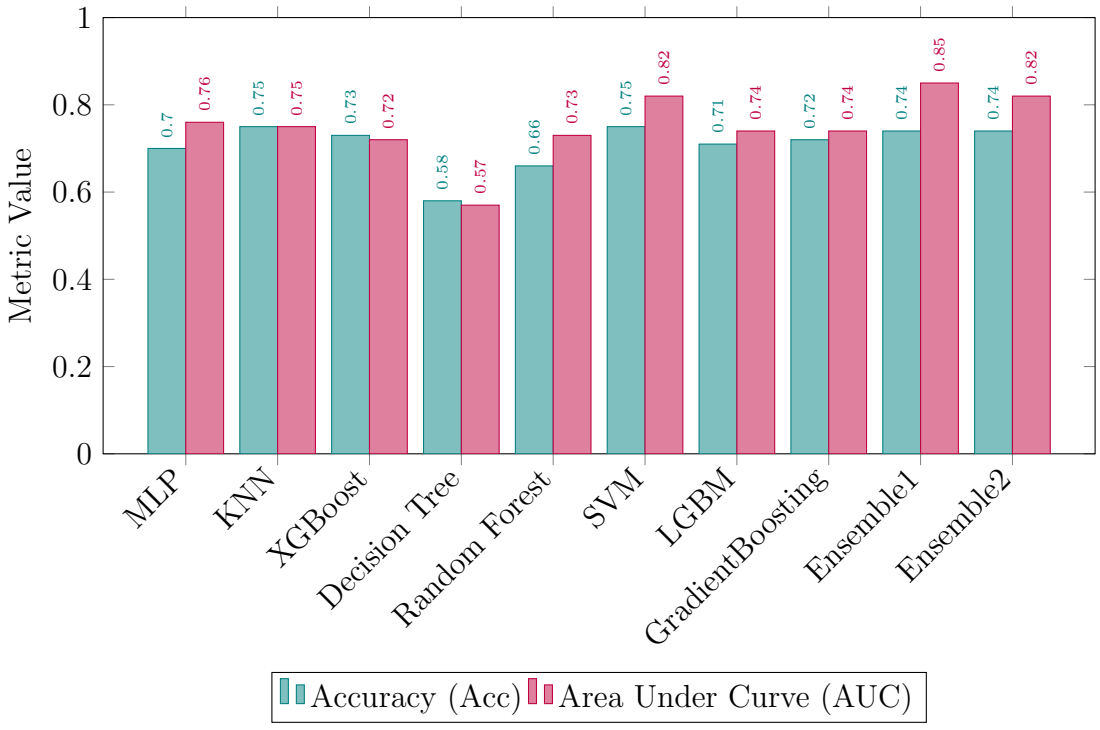



Figure 6.17: Comparison of Accuracy and AUC for different classifiers using T2W Sagittal images.

In the case of T2W sagittal images, both support vector machine and k-Nearest

Neighbor demonstrated strong individual performance, each reaching an accuracy of 0.75, with Support Vector Machine showing a higher AUC of 0.82 compared to k-Nearest Neighbor's 0.75. Ensemble1 achieved the highest AUC overall (0.85) while maintaining a solid accuracy of 0.74, indicating that combining multiple models can enhance discriminative capability. Multilayer perceptron and XGBoost also provided moderate results (accuracy 0.70–0.73, AUC 0.72–0.76), whereas Decision Tree lagged (accuracy 0.58, AUC 0.57). Overall, these findings suggest that, similar to T2W coronal images, margin-based classifiers such as support vector machine and ensemble methods yield the most reliable predictions for sagittal plane images.

In comparison with previous studies for prostate cancer detection, our work demonstrates that despite using a relatively small dataset, classical ML methods and ensemble strategies can achieve competitive performance for prostate cancer classification. For instance, our best-performing models using ADC and DWI images achieved an accuracy of 0.83 with an AUC of 0.871 (multilayer perceptron on ADC) and an accuracy of 0.80 with an AUC of 0.85 (Ensemble 1 on DWI). These results are encouraging when considering that many recent studies relied on substantially larger datasets and DL architectures.

For example, Karagoz et al.[165] employed a 3D nnU-Net on more than 1,500 training cases and reported an AUROC of 0.888–0.889, while Taguelmimt et al.[164] combined bp-MRI with clinical data and achieved an AUC of 0.82±0.03. Similarly, Wang et al.[170] developed a multi-input convolution neural network trained on 650 cases, achieving an accuracy of 82.5%. In contrast, studies such as Chakraborty et al.[166]. Islam et al.[172] reported very high accuracies (98%) using a convolutional neural network-based transfer learning approach (ResNet, VGG). Still, these were obtained with datasets of several hundred to several thousand images, which are not directly comparable to our smaller dataset.

Compared to earlier studies based on ML, such as Litjens et al.[168], who used a support vector machine in 31 patients and achieved an accuracy of 0.76, our results are comparable or even superior in terms of AUC, especially when considering inputs in single-modality. Moreover, our findings align with Provenzano et al.[171], who highlighted the sensitivity of ML models to dataset variability,

reporting AUCs between 0.82 and 0.99 when trained on multi-institutional data but showing performance drops on external datasets.

Overall, while DL methods with large-scale datasets (e.g., ResNet, convolution neural network ensembles) clearly achieve higher performance, our results demonstrate that classical ML approaches can still yield robust classification accuracy and AUC values in smaller datasets. Importantly, our study highlights the feasibility of using simpler models that require fewer computational resources, while achieving accuracy and AUC values comparable to some early DL works. This suggests that with larger data sets or multimodal data integration (combining ADC, DWI, and T2W), our proposed approaches could further improve and bridge the performance gap with state-of-the-art DL methods.

## 6.2 Conclusions and Outlook

This thesis presented a comprehensive AI-based framework for prostate cancer diagnosis, which includes prostate zonal segmentation, lesion segmentation, PI-RADS classification, and real cancer detection using mp-MRI and clinical data. The primary objective was to develop a multi-stage, clinically relevant workflow that can assist radiologists, reduce variability, and provide an educational tool for medical trainees.

In zonal segmentation, the proposed DL model demonstrated a high accuracy in delineating the peripheral and transition zone, which is critical to localizing lesions and guiding subsequent analysis. Compared to previous studies that often used a single MRI sequence or relied on small datasets, this work integrated all three mpMRI sequences and manually verified masks under the supervision of an expert radiologist, ensuring robust and clinically valid results.

For lesion segmentation, the framework effectively identified suspicious regions within the prostate, allowing evaluation of the size, location, and proximity of the lesion to critical structures. Although high PI-RADS lesions were accurately

highlighted, the study emphasized that not all high PI-RADS lesions are malignant, highlighting the ongoing need for biopsy confirmation and clinical interpretation. The segmentation results demonstrated competitive performance compared to state-of-the-art models reported in the literature, even though the dataset was relatively small.

The PI-RADS classification module further contributed to diagnostic precision by extracting key imaging features capable of differentiating lesions across PI-RADS categories. Integrating AI in this step improved consistency and reproducibility, reducing inter-observer variability common in manual scoring. This shows the potential of AI to support standardized evaluation and increase the confidence of radiographers.

In the final stage, real prostate cancer classification combined imaging outputs with selected clinical features, such as patient age, prostate-specific antigen level, and biopsy results. This integration enabled more reliable identification of malignant cases, reflecting a multi-modal decision-making process that aligns closely with real-world clinical practice. Despite the limited dataset, the model achieved promising predictive performance, suggesting that with larger, multi-center data, accuracy and generalizability could be further enhanced.

A notable contribution of this thesis is the end-to-end workflow, where the output from each stage is directly fed into the next, mimicking clinical reasoning. Additionally, the development of a user-friendly 3D Slicer plugin for prostate zonal segmentation ensures that the models can be easily applied in both research and clinical settings, serving as a training and decision-support tool. This approach addresses a significant gap in previous studies, which often focus on isolated tasks rather than integrated practical solutions.

Limitations of this work include the relatively small dataset size, restricted clinical features, and lack of multi-center external validation. Variability in mpMRI acquisition protocols also posed challenges, and while manual annotations under expert supervision mitigated this, further standardization would enhance robustness.

Future work should focus on expanding the dataset involving multiple institutions, incorporating additional clinical and laboratory features, and exploring ensemble or hybrid DL models to improve predictive performance. Prospective validation in clinical workflows is also critical to assess real-world utility, and potential extension to other imaging modalities could broaden applicability.

In addition, future work could explore Federated Learning approaches to enable collaborative model training and sharing without the need to exchange sensitive patient data. Specifically, our prostate zonal segmentation and lesion segmentation models could be deployed across multiple hospitals or clinical centers, allowing each institution to contribute to model improvement while keeping their datasets locally stored. This approach would enhance model generalizability, reduce potential biases, and maintain patient privacy, which is particularly important in multi-center studies. To facilitate practical deployment, we have developed a user-friendly 3D Slicer interface that allows clinicians and researchers to easily use, share, and test the models in their local environments. This interface simplifies the integration of federated learning workflows, making it easy for other centers to adopt the models without requiring advanced programming expertise. In general, leveraging federated learning in combination with our intuitive 3D Slicer platform has the potential to accelerate collaborative AI development, promote standardized prostate MRI analysis, and expand the impact of our models in real-world clinical settings.

In conclusion, this thesis provides a multi-stage AI framework for prostate cancer diagnosis that combines imaging, clinical data, and advanced ML. It shows potential to improve clinical decision-making, reduce diagnostic workload, improve resident training, and support personalized patient care. The methodology, tools, and insights presented in this work lay the foundation for future research and clinical applications in AI-assisted prostate cancer detection.

# Bibliography


[1] J. Guo, H. Lou, H. Chen, H. Liu, J. Gu, L. Bi, and X. Duan, "A new detection algorithm for alien intrusion on highway," Scientific Reports, vol. 13, jul 2023.

[2] F. Bray, M. Laversanne, H. Sung, J. Ferlay, R. L. Siegel, I. Soerjomataram, and A. Jemal, "Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA: A Cancer Journal for Clinicians, vol. 74, no. 3, pp. 229–263, 2024. [Online]. Available: https://doi.org/10.3322/caac.21834

[3] P. Rawla, "Epidemiology of Prostate Cancer." World journal of oncology, vol. 10, no. 2, pp. 63–89, apr 2019.

[4] S. C. Statistics, "Prostate cancer statistics," https://seer.cancer.gov/statfacts/html/prost.html, 2025, accessed: 2025-09-30.

[5] N. H. G. R. I. (NHGRI), "Prostate cancer background information," https://www.genome.gov/10005585/prostate-cancer-background-information, 2012, last updated May 23, 2012; Accessed: 2025-09-30.

[6] L. Geddes, "Genetics may help explain black men's high prostate cancer risk, say scientists," https://www.theguardian.com/society/2024/feb/15/genetics-may-help-explain-black-mens-high-prostate-cancer-risk-say-scientists, Feb 2024, accessed: 2025-09-30.

[7] S. Ghai and M. A. Haider, "Multiparametric-MRI in diagnosis of prostate cancer."

Indian journal of urology : IJU : journal of the Urological Society of India, vol. 31, no. 3, pp. 194–201, 2015.

[8] M. C. Cabarrus and A. C. Westphalen, "Multiparametric magnetic resonance imaging of the prostate-a basic tutorial." Translational andrology and urology, vol. 6, no. 3, pp. 376–386, jun 2017.

[9] C. Wu, S. Montagne, D. Hamzaoui, N. Ayache, H. Delingette, and R. Renard-Penna, "Automatic segmentation of prostate zonal anatomy on MRI: a systematic review of the literature," Insights into Imaging, vol. 13, no. 1, p. 202, 2022. [Online]. Available: https://doi.org/10.1186/s13244-022-01340-2

[10] O. J. Pellicer-Valero, J. L. Marenco Jiménez, V. Gonzalez-Perez, J. L. Casanova Ramón-Borja, I. Martín García, M. Barrios Benito, P. Pelechano Gómez, J. Rubio-Briones, M. J. Rupérez, and J. D. Martín-Guerrero, "Deep learning for fully automatic detection, segmentation, and Gleason grade estimation of prostate cancer in multiparametric magnetic resonance images," Scientific Reports, vol. 12, no. 1, p. 2975, 2022. [Online]. Available: https://doi.org/10.1038/s41598-022-06730-6

[11] B. Turkbey, A. B. Rosenkrantz, M. A. Haider, A. R. Padhani, G. Villeirs, K. J. Macura, C. M. Tempany, P. L. Choyke, F. Cornud, D. J. Margolis, H. C. Thoeny, S. Verma, J. Barentsz, and J. C. Weinreb, "Prostate Imaging Reporting and Data System Version 2.1: 2019 Update of Prostate Imaging Reporting and Data System Version 2." European urology, vol. 76, no. 3, pp. 340–351, sep 2019.

[12] P. Schelb, S. Kohl, J. P. Radtke, M. Wiesenfarth, P. Kickingereder, S. Bickelhaupt, T. A. Kuder, A. Stenzinger, M. Hohenfellner, H.-P. Schlemmer, K. H. Maier-Hein, and D. Bonekamp, "Classification of Cancer at Prostate MRI: Deep Learning versus Clinical PI-RADS Assessment,"

Radiology, vol. 293, no. 3, pp. 607–617, oct 2019. [Online]. Available: https://doi.org/10.1148/radiol.2019190938

[13] A. Ayranci, U. Caglar, H. B. Yazili, F. S. Erdal, A. Erbin, O. Sarilar, and F. Ozgor, "PSA Density and Lesion Volume: Key Factors in Avoiding Unnecessary Biopsies for PI-RADS 3 Lesions." The Prostate, vol. 85, no. 4, pp. 385–390, mar 2025.

[14] Y.-H. Wang, C. Liang, F.-P. Zhu, T.-R. Zhou, J. Li, Z.-J. Wang, and B.-J. Liu, "Improving the understanding of PI-RADS in practice: characters of PI-RADS 4 and 5 lesions with negative biopsy." Asian journal of andrology, vol. 25, no. 2, pp. 217–222, 2023.

[15] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2020." CA: a cancer journal for clinicians, vol. 70, no. 1, pp. 7–30, jan 2020.

[16] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA: A Cancer Journal for Clinicians, vol. 68, no. 6, pp. 394–424, nov 2018. [Online]. Available: https://doi.org/10.3322/caac.21492

[17] H. Quon, A. Loblaw, and R. Nam, "Dramatic increase in prostate cancer cases by 2021." BJU international, vol. 108, no. 11, pp. 1734–1738, dec 2011.

[18] ——, "Dramatic increase in prostate cancer cases by 2021." BJU international, vol. 108, no. 11, pp. 1734–1738, dec 2011.

[19] A. M. D. Wolf, R. C. Wender, R. B. Etzioni, I. M. Thompson, A. V. D'Amico, R. J. Volk, D. D. Brooks, C. Dash, I. Guessous, K. Andrews, C. DeSantis, and R. A. Smith, "American Cancer Society guideline for the early detection of prostate cancer: update 2010." CA: a cancer journal for clinicians, vol. 60, no. 2, pp. 70–98, 2010.

[20] J. Cuzick, M. A. Thorat, G. Andriole, O. W. Brawley, P. H. Brown, Z. Culig, R. A. Eeles, L. G. Ford, F. C. Hamdy, L. Holmberg, D. Ilic, T. J. Key,

C. La Vecchia, H. Lilja, M. Marberger, F. L. Meyskens, L. M. Minasian, C. Parker, H. L. Parnes, S. Perner, H. Rittenhouse, J. Schalken, H.-P. Schmid, B. J. Schmitz-Dräger, F. H. Schröder, A. Stenzl, B. Tombal, T. J. Wilt, and A. Wolk, "Prevention and early detection of prostate cancer." The Lancet. Oncology, vol. 15, no. 11, pp. e484–92, oct 2014.

[21] A. Fernandez-Quilez, M. Germán Borda, G. Leonardo Carreño, N. Castellanos-Perilla, H. Soennesyn, K. Oppedal, and S. Reidar Kjosavik, "Prostate cancer screening and socioeconomic disparities in Mexican older adults." pp. 121–122, 2020.

[22] "The Prostate Gland," 2025. [Online]. Available: https://teachmeanatomy.info/pelvis/%0Athe-male-reproductive-system/prostate-gland

[23] J. E. McNeal, E. A. Redwine, F. S. Freiha, and T. A. Stamey, "Zonal distribution of prostatic adenocarcinoma. Correlation with histologic pattern and direction of spread." The American journal of surgical pathology, vol. 12, no. 12, pp. 897–906, dec 1988.

[24] J. E. McNeal, "The zonal anatomy of the prostate." The Prostate, vol. 2, no. 1, pp. 35–49, 1981.

[25] M. S. Powell, R. Li, H. Dai, M. Sayeeduddin, T. M. Wheeler, and G. E. Ayala, "Neuroanatomy of the normal prostate." The Prostate, vol. 65, no. 1, pp. 52–57, sep 2005.

[26] E. F. Alves, B. L. M. de Freitas Ribeiro, W. S. Costa, C. B. M. Gallo, and F. J. B. Sampaio, "Histological and quantitative analyzes of the stromal and acinar components of normal human prostate zones." The Prostate, vol. 78, no. 4, pp. 289–293, mar 2018.

[27] X. Yu, R. Liu, L. Song, W. Gao, X. Wang, and Y. Zhang, "Differences in the pathogenetic characteristics of prostate cancer in the transitional and peripheral zones and the possible molecular biological mechanisms." Frontiers in oncology, vol. 13, p. 1165732, 2023.

[28] J. E. McNeal, "Normal histology of the prostate." The American journal of surgical pathology, vol. 12, no. 8, pp. 619–633, aug 1988.

[29] J. V. Tricoli, M. Schoenfeldt, and B. A. Conley, "Detection of Prostate Cancer and Predicting Progression: Current and Future Diagnostic Markers," Clinical Cancer Research, vol. 10, no. 12, pp. 3943–3953, 2004. [Online]. Available: https://doi.org/10.1158/1078-0432.CCR-03-0200

[30] F. S. Leach, M. S. Koh, Y. W. Chan, S. Bark, R. Ray, R. A. Morton, and A. T. Remaley, "Prostate specific antigen as a clinical biomarker for prostate cancer: what's the take home message?" Cancer Biology & Therapy, vol. 4, no. 4, pp. 371–375, Apr. 2005.

[31] A. Krilaviciute, R. Kaaks, P. Seibold, M. de Vrieze, J. Lakes, J. P. Radtke, M. Kuczyk, N. N. Harke, J. Debus, C. A. Fink, K. Herkommer, J. E. Gschwend, V. H. Meissner, A. Benner, G. Kristiansen, B. Hadaschik, C. Arsov, L. Schimmöller, G. Antoch, F. L. Giesel, M. Makowski, F. Wacker, H.-P. Schlemmer, N. Becker, and P. Albers, "Risk-adjusted Screening for Prostate Cancer—Defining the Low-risk Group by Data from the PROBASE Trial," European Urology, vol. 86, no. 6, pp. 493–500, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0302283824023583

[32] S. Garg, A. Sachdeva, M. Peeters, and J. McClements, "Point-of-Care Prostate Specific Antigen Testing: Examining Translational Progress toward Clinical Implementation." ACS sensors, vol. 8, no. 10, pp. 3643–3658, oct 2023.

[33] V. M. Sundaresan, S. Smani, P. Rajwa, J. Renzulli, P. C. Sprenkle, I. Y. Kim, and M. S. Leapman, "Prostate-specific antigen screening for prostate cancer: Diagnostic performance, clinical thresholds, and strategies for refinement." Urologic oncology, vol. 43, no. 1, pp. 41–48, jan 2025.

[34] J. Panzone, T. Byler, G. Bratslavsky, and H. Goldberg, "Transrectal Ultrasound in Prostate Cancer: Current Utilization, Integration with mpMRI, HIFU and Other Emerging Applications." Cancer management and research, vol. 14, pp. 1209–1228, 2022.

[35] B. P. Rai, C. Mayerhofer, B. K. Somani, P. Kallidonis, U. Nagele, and T. Tokas, "Magnetic Resonance Imaging/Ultrasound Fusion-guided Transperineal Versus Magnetic Resonance Imaging/Ultrasound Fusion-guided Transrectal Prostate Biopsy-A Systematic Review." European urology oncology, vol. 4, no. 6, pp. 904–913, dec 2021.

[36] D. F. Gleason, "Histologic grading of prostate cancer: A perspective," Human Pathology, vol. 23, no. 3, pp. 273–279, 1992. [Online]. Available: https://www.sciencedirect.com/science/article/pii/004681779290108F

[37] J. Gordetsky and J. Epstein, "Grading of prostatic adenocarcinoma: current state and prognostic implications." Diagnostic pathology, vol. 11, p. 25, mar 2016.

[38] K. K. Hodge, J. E. McNeal, and T. A. Stamey, "Ultrasound guided transrectal core biopsies of the palpably abnormal prostate." The Journal of urology, vol. 142, no. 1, pp. 66–70, jul 1989.

[39] K. Eichler, S. Hempel, J. Wilby, L. Myers, L. M. Bachmann, and J. Kleijnen, "Diagnostic value of systematic biopsy methods in the investigation of prostate cancer: a systematic review." The Journal of urology, vol. 175, no. 5, pp. 1605–1612, may 2006.

[40] A. Manseck, M. Froehner, S. Oehlschlaeger, O. Hakenberg, K. Friedrich, F. Theissig, and M. P. Wirth, "Is systematic sextant biopsy suitable for the detection of clinically significant prostate cancer?" Urologia internationalis, vol. 65, no. 2, pp. 80–83, 2000.

[41] K. G. Fink, G. Hutarew, W. Lumper, A. Jungwirth, O. Dietze, and N. T. Schmeller, "Prostate cancer detection with two sets

of ten-core compared with two sets of sextant biopsies," Urology, vol. 58, no. 5, pp. 735–739, 2001. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0090429501013528

[42] A. V. Taira, G. S. Merrick, R. W. Galbreath, H. Andreini, W. Taubenslag, R. Curtis, W. M. Butler, E. Adamovich, and K. E. Wallner, "Performance of transperineal template-guided mapping biopsy in detecting prostate cancer in the initial and repeat biopsy setting." Prostate cancer and prostatic diseases, vol. 13, no. 1, pp. 71–77, mar 2010.

[43] H. C. Demirel and J. W. Davis, "Multiparametric magnetic resonance imaging: Overview of the technique, clinical applications in prostate biopsy and future directions." Turkish journal of urology, vol. 44, no. 2, pp. 93–102, mar 2018.

[44] P. Dirix, S. Van Bruwaene, H. Vandeursen, and F. Deckers, "Magnetic resonance imaging sequences for prostate cancer triage: two is a couple, three is a crowd?" pp. S476–S479, dec 2019.

[45] A. Shukla-Dave and H. Hricak, "Role of MRI in prostate cancer detection." NMR in biomedicine, vol. 27, no. 1, pp. 16–24, jan 2014.

[46] A. Stabile, F. Giganti, A. B. Rosenkrantz, S. S. Taneja, G. Villeirs, I. S. Gill, C. Allen, M. Emberton, C. M. Moore, and V. Kasivisvanathan, "Multiparametric MRI for prostate cancer diagnosis: current status and future directions." Nature reviews. Urology, vol. 17, no. 1, pp. 41–61, jan 2020.

[47] A. Grenabo Bergdahl, U. Wilderäng, G. Aus, S. Carlsson, J.-E. Damber, M. Frånlund, K. Geterud, A. Khatami, A. Socratous, J. Stranne, M. Hellström, and J. Hugosson, "Role of Magnetic Resonance Imaging in Prostate Cancer Screening: A Pilot Study Within the Göteborg Randomised Screening Trial." European urology, vol. 70, no. 4, pp. 566–573, oct 2016.

[48] D. Eldred-Evans, P. Burak, M. J. Connor, E. Day, M. Evans, F. Fiorentino, M. Gammon, F. Hosking-Jervis, N. Klimowska-Nassar, W. McGuire, A. R.

Padhani, A. T. Prevost, D. Price, H. Sokhi, H. Tam, M. Winkler, and H. U. Ahmed, "Population-Based Prostate Cancer Screening With Magnetic Resonance Imaging or Ultrasonography: The IP1-PROSTAGRAM Study." JAMA oncology, vol. 7, no. 3, pp. 395–402, mar 2021.

[49] T. Tamada, A. Kido, A. Yamamoto, M. Takeuchi, Y. Miyaji, T. Moriya, and T. Sone, "Comparison of Biparametric and Multiparametric MRI for Clinically Significant Prostate Cancer Detection With PI-RADS Version 2.1." Journal of magnetic resonance imaging : JMRI, vol. 53, no. 1, pp. 283–291, jan 2021.

[50] B. Israël, M. van der Leest, M. Sedelaar, A. R. Padhani, P. Zámecnik, and J. O. Barentsz, "Multiparametric Magnetic Resonance Imaging for the Detection of Clinically Significant Prostate Cancer: What Urologists Need to Know. Part 2: Interpretation," European Urology, vol. 77, no. 4, pp. 469–480, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0302283819308231

[51] C. Hoeks, J. Barentsz, T. Hambrock, D. Yakar, D. Somford, S. Heijmink, T. Scheenen, P. Vos, H. Huisman, I. van Oort, J. Witjes, A. Heerschap, and J. Fütterer, "Prostate Cancer: Multiparametric MR Imaging for Detection, Localization, and Staging," Radiology, vol. 261, pp. 46–66, oct 2011.

[52] "PI-RADS Version 2," 2015. [Online]. Available: https://www.acr.org/

[53] S. Verma, S. Sarkar, J. Young, R. Venkataraman, X. Yang, A. Bhavsar, N. Patil, J. Donovan, and K. Gaitonde, "Evaluation of the impact of computed high b-value diffusion-weighted imaging on prostate cancer detection." Abdominal radiology (New York), vol. 41, no. 5, pp. 934–945, may 2016.

[54] M. D. Blackledge, M. O. Leach, D. J. Collins, and D.-M. Koh, "Computed diffusion-weighted MR imaging may improve tumor detection." Radiology, vol. 261, no. 2, pp. 573–581, nov 2011.

[55] P. De Visschere, N. Lumen, P. Ost, K. Decaestecker, E. Pattyn, and G. Villeirs, "Dynamic contrast-enhanced imaging has limited added value over T2-weighted imaging and diffusion-weighted imaging when using PI-RADSv2 for diagnosis of clinically significant prostate cancer in patients with elevated PSA." Clinical radiology, vol. 72, no. 1, pp. 23–32, jan 2017.

[56] M. Frydenberg, P. D. Stricker, and K. W. Kaye, "Prostate cancer diagnosis and management," pp. 1681–1687.

[57] L. A. M. Simmons, A. Kanthabalan, M. Arya, T. Briggs, D. Barratt, S. C. Charman, A. Freeman, J. Gelister, D. Hawkes, Y. Hu, C. Jameson, N. McCartan, C. M. Moore, S. Punwani, N. Ramachandran, J. van der Meulen, M. Emberton, and H. U. Ahmed, "The PICTURE study: diagnostic accuracy of multiparametric MRI in men requiring a repeat prostate biopsy," British Journal of Cancer, vol. 116, no. 9, pp. 1159–1165, 2017. [Online]. Available: https://doi.org/10.1038/bjc.2017.57

[58] H. U. Ahmed, A. El-Shater Bosaily, L. C. Brown, R. Gabe, R. Kaplan, M. K. Parmar, Y. Collaco-Moraes, K. Ward, R. G. Hindley, A. Freeman, A. P. Kirkham, R. Oldroyd, C. Parker, and M. Emberton, "Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study," The Lancet, vol. 389, no. 10071, pp. 815–822, feb 2017. [Online]. Available: https://doi.org/10.1016/S0140-6736(16)32401-1

[59] D. F. Gleason, "Classification of prostatic carcinomas." Cancer chemotherapy reports, vol. 50, no. 3, pp. 125–128, mar 1966.

[60] J. I. Epstein, L. Egevad, M. Amin, B. Delahunt, J. R. Srigley, P. A. Humphrey, T. Al Hussain, F. Algaba, M. Aron, D. Berman, D. Berney, F. Brimo, D. Cao, J. Cheville, D. Clouston, M. Colecchia, E. Compérat, I. Cunha, A. Marzo, and G. Rodrigues, "The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma: definition of grading patterns and proposal for a

new grading system." American Journal of Surgical Pathology, vol. 40, pp. 244–252, feb 2016.

[61] P. Lambin, E. Rios-Velazquez, R. Leijenaar, S. Carvalho, R. G. P. M. van Stiphout, P. Granton, C. M. L. Zegers, R. Gillies, R. Boellard, A. Dekker, and H. J. W. L. Aerts, "Radiomics: Extracting more information from medical images using advanced feature analysis," European Journal of Cancer, vol. 48, no. 4, pp. 441–446, mar 2012. [Online]. Available: https://doi.org/10.1016/j.ejca.2011.11.036

[62] H. J. W. L. Aerts, E. R. Velazquez, R. T. H. Leijenaar, C. Parmar, P. Grossmann, S. Carvalho, J. Bussink, R. Monshouwer, B. Haibe-Kains, D. Rietveld, F. Hoebers, M. M. Rietbergen, C. R. Leemans, A. Dekker, J. Quackenbush, R. J. Gillies, and P. Lambin, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," Nature Communications, vol. 5, no. 1, p. 4006, 2014. [Online]. Available: https://doi.org/10.1038/ncomms5006

[63] A. Zwanenburg, M. Vallières, M. A. Abdalah, H. J. W. L. Aerts, V. Andrearczyk, A. Apte, S. Ashrafinia, S. Bakas, R. J. Beukinga, R. Boellaard, M. Bogowicz, L. Boldrini, I. Buvat, G. J. R. Cook, C. Davatzikos, A. Depeursinge, M.-C. Desseroit, N. Dinapoli, C. V. Dinh, S. Echegaray, I. El Naqa, A. Y. Fedorov, R. Gatta, R. J. Gillies, V. Goh, M. Götz, M. Guckenberger, S. M. Ha, M. Hatt, F. Isensee, P. Lambin, S. Leger, R. T. H. Leijenaar, J. Lenkowicz, F. Lippert, A. Losnegård, K. H. Maier-Hein, O. Morin, H. Müller, S. Napel, C. Nioche, F. Orlhac, S. Pati, E. A. G. Pfaehler, A. Rahmim, A. U. K. Rao, J. Scherer, M. M. Siddique, N. M. Sijtsema, J. Socarras Fernandez, E. Spezi, R. J. H. M. Steenbakkers, S. Tanadini-Lang, D. Thorwarth, E. G. C. Troost, T. Upadhaya, V. Valentini, L. V. van Dijk, J. van Griethuysen, F. H. P. van Velden, P. Whybra, C. Richter, and S. Löck, "The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping,"

Radiology, vol. 295, no. 2, pp. 328–338, mar 2020. [Online]. Available: https://doi.org/10.1148/radiol.2020191145

[64] V. Kumar, Y. Gu, S. Basu, A. Berglund, S. A. Eschrich, M. B. Schabath, K. Forster, H. J. W. L. Aerts, A. Dekker, D. Fenstermacher, D. B. Goldgof, L. O. Hall, P. Lambin, Y. Balagurunathan, R. A. Gatenby, and R. J. Gillies, "Radiomics: the process and the challenges," Magnetic Resonance Imaging, vol. 30, no. 9, pp. 1234–1248, 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0730725X12002202

[65] J. J. M. van Griethuysen, A. Fedorov, C. Parmar, A. Hosny, N. Aucoin, V. Narayan, R. G. H. Beets-Tan, J.-C. Fillion-Robin, S. Pieper, and H. J. W. L. Aerts, "Computational Radiomics System to Decode the Radiographic Phenotype," Cancer Research, vol. 77, no. 21, pp. e104–e107, oct 2017. [Online]. Available: https://doi.org/10.1158/0008-5472.CAN-17-0339

[66] C. Parmar, P. Grossmann, D. Rietveld, M. M. Rietbergen, P. Lambin, and H. J. W. L. Aerts, "Radiomic Machine-Learning Classifiers for Prognostic Biomarkers of Head and Neck Cancer," Frontiers in Oncology, vol. Volume 5 -, 2015. [Online]. Available: https://www.frontiersin.org/journals/oncology/articles/10.3389/fonc.2015.00272

[67] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning," Nature, vol. 518, no. 7540, pp. 529–533, 2015. [Online]. Available: https://doi.org/10.1038/nature14236

[68] M. Moravčík, M. Schmid, N. Burch, V. Lisý, D. Morrill, N. Bard, T. Davis, K. Waugh, M. Johanson, and M. Bowling, "DeepStack: Expert-level artificial intelligence in heads-up no-limit poker," Science, vol. 356, no. 6337, pp. 508–513, may 2017. [Online]. Available: https://doi.org/10.1126/science.aam6960

[69] W. Xiong, J. Droppo, X. Huang, F. Seide, M. L. Seltzer, A. Stolcke, D. Yu, and G. Zweig, "Toward Human Parity in Conversational Speech Recognition," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 25, no. 12, pp. 2410–2423, 2017.

[70] S. D. Pendleton, H. Andersen, X. Du, X. Shen, M. Meghjani, Y. H. Eng, D. Rus, and M. H. Ang, "Perception, Planning, Control, and Coordination for Autonomous Vehicles," 2017.

[71] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, "Mastering the game of Go with deep neural networks and tree search," Nature, vol. 529, no. 7587, pp. 484–489, 2016. [Online]. Available: https://doi.org/10.1038/nature16961

[72] K. Grace, J. Salvatier, A. Dafoe, B. Zhang, and O. Evans, "Viewpoint: When Will AI Exceed Human Performance? Evidence from AI Experts," Journal of Artificial Intelligence Research, vol. 62, pp. 729–754, jul 2018.

[73] A. Hosny, C. Parmar, J. Quackenbush, L. H. Schwartz, and H. J. W. L. Aerts, "Artificial intelligence in radiology," Nature Reviews Cancer, vol. 18, no. 8, pp. 500–510, 2018. [Online]. Available: https://doi.org/10.1038/s41568-018-0016-5

[74] R. Castellino, "Computer aided detection (CAD): An overview," Cancer imaging : the official publication of the International Cancer Imaging Society, vol. 5, pp. 17–19, feb 2005.

[75] Q. Sun, X. Lin, Y. Zhao, L. Li, K. Yan, D. Liang, D. Sun, and Z.-C. Li, "Deep Learning vs. Radiomics for Predicting Axillary Lymph Node Metastasis of Breast Cancer Using Ultrasound Images: Don't Forget the Peritumoral Region," Frontiers in Oncology, vol. 10, p. 53, jan 2020.

[76] D. Shen, G. Wu, and H.-I. Suk, "Deep Learning in Medical Image Analysis," Annual Review of Biomedical Engineering, vol. 19, no. Volume 19, 2017, pp. 221–248, 2017. [Online]. Available: https://www.annualreviews.org/content/journals/10.1146/annurev-bioeng-071516-044442

[77] S. Benjamens, P. Dhunnoo, and B. Meskó, "The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database," npj Digital Medicine, vol. 3, no. 1, p. 118, 2020. [Online]. Available: https://doi.org/10.1038/s41746-020-00324-0

[78] P. Papadimitroulas, L. Brocki, N. Christopher Chung, W. Marchadour, F. Vermet, L. Gaubert, V. Eleftheriadis, D. Plachouris, D. Visvikis, G. C. Kagadis, and M. Hatt, "Artificial intelligence: Deep learning in oncological radiomics and challenges of interpretability and data harmonization." Physica medica : PM : an international journal devoted to the applications of physics to mec vol. 83, pp. 108–121, mar 2021.

[79] R. Paul, S. H. Hawkins, Y. Balagurunathan, M. B. Schabath, R. J. Gillies, L. O. Hall, and D. B. Goldgof, "Deep Feature Transfer Learning in Combination with Traditional Features Predicts Survival Among Patients with Lung Adenocarcinoma." Tomography (Ann Arbor, Mich.), vol. 2, no. 4, pp. 388–395, dec 2016.

[80] J.-Z. Cheng, D. Ni, Y.-H. Chou, J. Qin, C.-M. Tiu, Y.-C. Chang, C.-S. Huang, D. Shen, and C.-M. Chen, "Computer-Aided Diagnosis with Deep Learning Architecture: Applications to Breast Lesions in US Images and Pulmonary Nodules in CT Scans," Scientific Reports, vol. 6, no. 1, p. 24454, 2016. [Online]. Available: https://doi.org/10.1038/srep24454

[81] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016. [Online]. Available: https://www.deeplearningbook.org/

[82] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and

C. I. Sánchez, "A survey on deep learning in medical image analysis," Medical Image Analysis, vol. 42, pp. 60–88, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841517301135

[83] A. Khan et al., "Artificial intelligence in prostate cancer imaging: a systematic review of deep learning for detection, characterization, and prediction," European Radiology, 2024.

[84] A. Saha et al., "Artificial intelligence applications in prostate cancer imaging: current status and future perspectives," Insights into Imaging, vol. 14, no. 1, p. 24, 2023.

[85] O. S. Tătaru, M. D. Vartolomei, J. J. Rassweiler, O. Virgil, G. Lucarelli, F. Porpiglia, D. Amparore, M. Manfredi, G. Carrieri, U. Falagario, D. Terracciano, O. de Cobelli, G. M. Busetto, F. D. Giudice, and M. Ferro, "Artificial Intelligence and Machine Learning in Prostate Cancer Patient Management—Current Trends and Future Perspectives," 2021.

[86] J. C. Weinreb, J. O. Barentsz, P. L. Choyke, F. Cornud, M. A. Haider, K. J. Macura, D. Margolis, M. D. Schnall, F. Shtern, C. M. Tempany, H. C. Thoeny, and S. Verma, "PI-RADS Prostate Imaging - Reporting and Data System: 2015, Version 2." European urology, vol. 69, no. 1, pp. 16–40, jan 2016.

[87] D. Hamzaoui, S. Montagne, B. Granger, A. Allera, M. Ezziane, A. Luzurier, R. Quint, M. Kalai, N. Ayache, H. Delingette, and R. Renard-Penna, "Prostate volume prediction on MRI: tools, accuracy and variability," European Radiology, vol. 32, no. 7, pp. 4931–4941, 2022. [Online]. Available: https://doi.org/10.1007/s00330-022-08554-4

[88] T. Matsugasumi, A. Fujihara, S. Ushijima, M. Kanazawa, Y. Yamada, T. Shiraishi, F. Hongo, K. Kamoi, K. Okihara, A. L. de Castro Abreu,

M. Oishi, T. Shin, S. Palmer, I. S. Gill, and O. Ukimura, "Morphometric analysis of prostate zonal anatomy using magnetic resonance imaging: impact on age-related changes in patients in Japan and the USA." BJU international, vol. 120, no. 4, pp. 497–504, oct 2017.

[89] S. Montagne, D. Hamzaoui, A. Allera, M. Ezziane, A. Luzurier, R. Quint, M. Kalai, N. Ayache, H. Delingette, and R. Renard-Penna, "Challenge of prostate MRI segmentation on T2-weighted images: inter-observer variability and impact of prostate morphology." Insights into imaging, vol. 12, no. 1, p. 71, jun 2021.

[90] O. Zavala-Romero, A. L. Breto, I. R. Xu, Y.-C. C. Chang, N. Gautney, A. Dal Pra, M. C. Abramowitz, A. Pollack, and R. Stoyanova, "Segmentation of prostate and prostate zones using deep learning : A multi-MRI vendor analysis." Strahlentherapie und Onkologie : Organ der Deutschen Rontgengesellschaft ... [et al], vol. 196, no. 10, pp. 932–942, oct 2020.

[91] R. Cuocolo, A. Stanzione, A. Castaldo, D. R. De Lucia, and M. Imbriaco, "Quality control and whole-gland, zonal and lesion annotations for the PROSTATEx challenge public dataset," European Journal of Radiology, vol. 138, p. 109647, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0720048X21001273

[92] R. Meng, X. Zhang, S. Huang, Y. Gu, G. Liu, G. Wu, N. Wang, K. Sun, and D. Shen, "NaMa: neighbor-aware multi-modal adaptive learning for prostate tumor segmentation on anisotropic MR images," in Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence and Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence and Fourteenth Symposium on Educational Advances in Artificial Intelligence, ser. AAAI'24/IAAI'24/EAAI'24. AAAI Press, 2024. [Online]. Available: https://doi.org/10.1609/aaai.v38i5.28215

[93] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation BT - Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015," N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.

[94] W. Wang, B. Pan, Y. Ai, G. Li, Y. Fu, and Y. Liu, "ParaCM-PNet: A CNN-tokenized MLP combined parallel dual pyramid network for prostate and prostate cancer segmentation in MRI," Computers in Biology and Medicine, vol. 170, p. 107999, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482524000830

[95] R. Mayeux and Y. Stern, "Epidemiology of Alzheimer disease," Cold Spring Harbor perspectives in medicine, vol. 2, no. 8, p. a006239, aug 2012. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/22908189https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3405821/

[96] G. D. Thomas, "Machine learning research: Four current directions," Artificial Intelligence, Magazine, vol. 18, no. 4, pp. 97–136, 1997.

[97] Z.-H. Zhou, J. Wu, and W. Tang, "Ensembling neural networks: Many could be better than all," Artificial Intelligence, vol. 137, no. 1, pp. 239–263, 2002. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S000437020200190X

[98] A. Verikas, A. Lipnickas, K. Malmqvist, M. Bacauskiene, and A. Gelzinis, "Soft combination of neural classifiers: A comparative study," Pattern Recognition Letters, vol. 20, no. 4, pp. 429–444, 1999. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167865599000124

[99] R. Hao, K. Namdar, L. Liu, M. A. Haider, and F. Khalvati, "A Comprehensive Study of Data Augmentation Strategies for Prostate Cancer Detection in Diffusion-Weighted MRI Using Convolutional Neural Networks." Journal of digital imaging, vol. 34, no. 4, pp. 862–876, aug 2021.

[100] W. Zhang, Q. Guo, Y. Zhu, M. Wang, T. Zhang, G. Cheng, Q. Zhang, and H. Ding, "Cross-institutional evaluation of deep learning and radiomics models in predicting microvascular invasion in hepatocellular carcinoma: validity, robustness, and ultrasound modality efficacy comparison," Cancer Imaging, vol. 24, no. 1, p. 142, 2024. [Online]. Available: https://doi.org/10.1186/s40644-024-00790-9

[101] A. Hasse, J. Bertini, S. Foxley, Y. Jeong, A. Javed, and T. J. Carroll, "Application of a novel T1 retrospective quantification using internal references (T1-REQUIRE) algorithm to derive quantitative T1 relaxation maps of the brain," International Journal of Imaging Systems and Technology, vol. 32, no. 6, pp. 1903–1915, 2022. [Online]. Available: https://doi.org/10.1002/ima.22768

[102] K. Fatania, F. Mohamud, A. Clark, M. Nix, S. C. Short, J. O'Connor, A. F. Scarsbrook, and S. Currie, "Intensity standardization of MRI prior to radiomic feature extraction for artificial intelligence research in glioma—a systematic review," European Radiology, vol. 32, no. 10, pp. 7014–7025, 2022. [Online]. Available: https://doi.org/10.1007/s00330-022-08807-2

[103] L. G. Nyúl and J. K. Udupa, "On standardizing the MR image intensity scale." Magnetic resonance in medicine, vol. 42, no. 6, pp. 1072–1081, dec 1999.

[104] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J.-C. Fillion-Robin, S. Pujol, C. Bauer, D. Jennings, F. Fennessy, M. Sonka, J. Buatti, S. Aylward, J. V. Miller, S. Pieper, and R. Kikinis, "3D Slicer as an image computing platform for the Quantitative Imaging Network." Magnetic resonance imaging, vol. 30, no. 9, pp. 1323–1341, nov 2012.

[105] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard,

T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, "Array programming with NumPy," Nature, vol. 585, no. 7825, pp. 357–362, 2020. [Online]. Available: https://doi.org/10.1038/s41586-020-2649-2

[106] J. D. Hunter, "Matplotlib: A 2D Graphics Environment," Computing in Science Engineering, vol. 9, no. 3, pp. 90–95, 2007.

[107] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, and X. Zheng, "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems," mar 2016.

[108] Z. Yaniv, B. C. Lowekamp, H. J. Johnson, and R. Beare, "SimpleITK Image-Analysis Notebooks: a Collaborative Environment for Education and Reproducible Research." Journal of digital imaging, vol. 31, no. 3, pp. 290–303, jun 2018.

[109] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, and S. Chintala, PyTorch: An Imperative Style, High-Performance Deep Learning Library, dec 2019.

[110] R. Cuocolo, A. Comelli, A. Stefano, V. Benfante, N. Dahiya, A. Stanzione, A. Castaldo, D. R. De Lucia, A. Yezzi, and M. Imbriaco, "Deep Learning Whole-Gland and Zonal Prostate Segmentation on a Public MRI Dataset." Journal of magnetic resonance imaging : JMRI, vol. 54, no. 2, pp. 452–459, aug 2021.

[111] W. Kou, H. Marshall, and B. Chiu, "Boundary-aware semantic clustering network for segmentation of prostate zones from T2-weighted MRI." Physics in medicine and biology, vol. 69, no. 17, aug 2024.

[112] J. Mitura, R. Jóźwiak, J. Mycka, I. Mykhalevych, M. Gonet, P. Sobecki, T. Lorenc, and K. Tupikowski, "Ensemble Deep Learning Models for Segmentation of Prostate Zonal Anatomy and Pathologically Suspicious Areas BT - Medical Image Understanding and Analysis," M. H. Yap, C. Kendrick, A. Behera, T. Cootes, and R. Zwiggelaar, Eds. Cham: Springer Nature Switzerland, 2024, pp. 217–231.

[113] L. Xu, G. Zhang, D. Zhang, J. Zhang, X. Zhang, X. Bai, L. Chen, Q. Peng, R. Jin, L. Mao, X. Li, Z. Jin, and H. Sun, "Development and clinical utility analysis of a prostate zonal segmentation model on T2-weighted imaging: a multicenter study," Insights into Imaging, vol. 14, no. 1, p. 44, 2023. [Online]. Available: https://doi.org/10.1186/s13244-023-01394-w

[114] Y. Yan, R. Liu, H. Chen, L. Zhang, and Q. Zhang, "CCT-Unet: A U-Shaped Network Based on Convolution Coupled Transformer for Segmentation of Peripheral and Transition Zones in Prostate MRI," IEEE Journal of Biomedical and Health Informatics, vol. 27, no. 9, pp. 4341–4351, 2023.

[115] A. Jimenez-Pastor, R. Lopez-Gonzalez, B. Fos-Guarinos, F. Garcia-Castro, M. Wittenberg, A. Torregrosa-Andrés, L. Marti-Bonmati, M. Garcia-Fontes, P. Duarte, J. P. Gambini, L. K. Bittencourt, F. C. Kitamura, V. K. Venugopal, V. Mahajan, P. Ros, E. Soria-Olivas, and A. Alberich-Bayarri, "Automated prostate multi-regional segmentation in magnetic resonance using fully convolutional neural networks," European Radiology, vol. 33, no. 7, pp. 5087–5096, 2023. [Online]. Available: https://doi.org/10.1007/s00330-023-09410-9

[116] M. Kaneko, G. E. Cacciamani, Y. Yang, V. Magoulianitis, J. Xue, J. Yang, J. Liu, M. S. L. Lenon, P. Mohamed, D. H. Hwang, K. Gill, M. Aron, V. Duddalwar, S. L. Palmer, C.-C. Jay Kuo, A. L. Abreu, I. Gill, and C. L. Nikias, "MP09-05AUTOMATED PROSTATE GLAND AND PROSTATE ZONES SEGMENTATION USING A NOVEL MRI-BASED MACHINE LEARNING FRAMEWORK AND

CREATION OF SOFTWARE INTERFACE FOR USERS ANNOTATION," Journal of Urology, vol. 209, no. Supplement 4, 2023. [Online]. Available: https://journals.lww.com/auajuro/fulltext/2023/04001/mp09_05_automated_prostate_gland_and_prostate.219.aspx

[117] M. Baldeon-Calisto, Z. Wei, S. Abudalou, Y. Yilmaz, K. Gage, J. Pow-Sang, and Y. Balagurunathan, "A multi-object deep neural network architecture to detect prostate anatomy in T2-weighted MRI: Performance evaluation," Frontiers in Nuclear Medicine, vol. 2, 2023. [Online]. Available: https://www.frontiersin.org/journals/nuclear-medicine/articles/10.3389/fnume.2022.1083245

[118] N. Aldoj, F. Biavati, F. Michallek, S. Stober, and M. Dewey, "Automatic prostate and prostate zones segmentation of magnetic resonance images using DenseNet-like U-net," Scientific Reports, vol. 10, aug 2020.

[119] F. Zabihollahy, N. Schieda, S. Krishna Jeyaraj, and E. Ukwatta, "Automated segmentation of prostate zonal anatomy on T2-weighted (T2W) and apparent diffusion coefficient (ADC) map MR images using U-Nets." Medical physics, vol. 46, no. 7, pp. 3078–3090, jul 2019.

[120] W. Li, B. Zheng, Q. Shen, X. Shi, K. Luo, Y. Yao, X. Li, S. Lv, J. Tao, and Q. Wei, "Adaptive window adjustment with boundary DoU loss for cascade segmentation of anatomy and lesions in prostate cancer using bpMRI," Neural Networks, vol. 181, p. 106831, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S089360802400755X

[121] C. Yan, F. Liu, Y. Peng, Y. Zhao, J. He, and R. Wang, "3D convolutional network with edge detection for prostate gland and tumor segmentation on T2WI and ADC," Biomedical Signal Processing and Control, vol. 90, p. 105883, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809423013162

[122] C. Thipkasorn, S. Chaichulee, T. Bejrananda, and T. Tubtawee, "Cascaded Architecture for Segmenting Prostate Cancer Lesions in Biparametric

MRI," in 2024 21st International Joint Conference on Computer Science and Software Engineering (JCSSE), 2024, pp. 167–173.

[123] L. C. Adams, M. R. Makowski, G. Engel, M. Rattunde, F. Busch, P. Asbach, S. M. Niehues, S. Vinayahalingam, B. van Ginneken, G. Litjens, and K. K. Bressem, "Prostate158 - An expert-annotated 3T MRI dataset and algorithm for prostate cancer detection," Computers in Biology and Medicine, vol. 148, p. 105817, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482522005789

[124] S. Hong, S. H. Kim, B. Yoo, and J. Y. Kim, "Deep Learning Algorithm for Tumor Segmentation and Discrimination of Clinically Significant Cancer in Patients with Prostate Cancer," pp. 7275–7285, 2023.

[125] Y. Liu, Y. Zhu, W. Wang, B. Zheng, X. Qin, and P. Wang, "Multi-scale discriminative network for prostate cancer lesion segmentation in multiparametric MR images." Medical physics, vol. 49, no. 11, pp. 7001–7015, nov 2022.

[126] R. Alkadi, F. Taher, A. El-Baz, and N. Werghi, "A Deep Learning-Based Approach for the Detection and Localization of Prostate Cancer in T2 Magnetic Resonance Images." Journal of digital imaging, vol. 32, no. 5, pp. 793–807, oct 2019.

[127] Z. A. Eidex, T. Wang, Y. Lei, M. Axente, O. O. Akin-Akintayo, O. A. A. Ojo, A. A. Akintayo, J. Roper, J. D. Bradley, T. Liu, D. M. Schuster, and X. Yang, "MRI-based prostate and dominant lesion segmentation using cascaded scoring convolutional neural network." Medical physics, vol. 49, no. 8, pp. 5216–5224, aug 2022.

[128] Z. Dai, E. Carver, C. Liu, J. Lee, A. Feldman, W. Zong, M. Pantelic, M. Elshaikh, and N. Wen, "Segmentation of the Prostatic Gland and the Intraprostatic Lesions on Multiparametic Magnetic Resonance Imaging Using Mask Region-Based Convolutional Neural Networks," Advances in

Radiation Oncology, vol. 5, no. 3, pp. 473–481, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S245210942030021X

[129] S. Kohl, D. Bonekamp, H.-P. Schlemmer, K. Yaqubi, M. Hohenfellner, B. Hadaschik, J.-P. Radtke, and K. H. Maier-Hein, "Adversarial Networks for the Detection of Aggressive Prostate Cancer." 2017. [Online]. Available: http://arxiv.org/abs/1702.08014

[130] Y. Artan, M. A. Haider, D. L. Langer, and I. S. Yetik, "Semi-supervised prostate cancer segmentation with multispectral MRI," in 2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2010, pp. 648–651.

[131] X. Liu, D. L. Langer, M. A. Haider, Y. Yang, M. N. Wernick, and I. S. Yetik, "Prostate cancer segmentation with simultaneous estimation of Markov random field parameters and class." IEEE transactions on medical imaging, vol. 28, no. 6, pp. 906–915, jun 2009.

[132] Z. Dai, I. Jambor, P. Taimen, M. Pantelic, M. Elshaikh, A. Dabaja, C. Rogers, O. Ettala, P. J. Boström, H. J. Aronen, H. Merisaari, and N. Wen, "Prostate cancer detection and segmentation on MRI using non-local mask R-CNN with histopathological ground truth." Medical physics, vol. 50, no. 12, pp. 7748–7763, dec 2023.

[133] J. A. Alzate-Grisales, A. Mora-Rubio, F. García-García, R. Tabares-Soto, and M. D. L. Iglesia-Vayá, "SAM-UNETR: Clinically Significant Prostate Cancer Segmentation Using Transfer Learning From Large Model," IEEE Access, vol. 11, pp. 118 217–118 228, 2023.

[134] A. B. Gavade, R. Nerli, N. Kanwal, P. A. Gavade, S. S. Pol, and S. T. Rizvi, "Automated Diagnosis of Prostate Cancer Using mpMRI Images: A Deep Learning Approach for Clinical Decision Support," 2023.

[135] D. D. Gunashekar, L. Bielak, L. Hägele, B. Oerther, M. Benndorf, A.-L. Grosu, T. Brox, C. Zamboglou, and M. Bock, "Explainable AI for CNN-based prostate tumor segmentation in multi-parametric MRI correlated to whole mount histopathology." Radiation oncology (London, England), vol. 17, no. 1, p. 65, apr 2022.

[136] Y. Lin, E. C. Yilmaz, M. J. Belue, S. A. Harmon, J. Tetreault, T. E. Phelps, K. M. Merriman, L. Hazen, C. Garcia, D. Yang, Z. Xu, N. S. Lay, A. Toubaji, M. J. Merino, D. Xu, Y. M. Law, S. Gurram, B. J. Wood, P. L. Choyke, P. A. Pinto, and B. Turkbey, "Evaluation of a Cascaded Deep Learning–based Algorithm for Prostate Lesion Detection at Biparametric MRI," Radiology, vol. 311, no. 2, p. e230750, may 2024. [Online]. Available: https://doi.org/10.1148/radiol.230750

[137] L. Wang, R. Sun, X. Wei, J. Chen, S. Jia, G. Wu, and S. Nie, "Enhancing prostate cancer segmentation on multiparametric magnetic resonance imaging with background information and gland masks," Medical Physics, vol. 51, no. 11, pp. 8179–8191, nov 2024. [Online]. Available: https://doi.org/10.1002/mp.17346

[138] M. Wu, F. Li, Y. Tao, Y. Zhang, S. Wang, P. Burstein, X. Mu, and J. Zhu, Lesion-Guided Selective Multi-Modal Integration for Prostate Cancer Segmentation and Pi-R jan 2024.

[139] J. Simeth, J. Jiang, A. Nosov, A. Wibmer, M. Zelefsky, N. Tyagi, and H. Veeraraghavan, "Deep learning-based dominant index lesion segmentation for MR-guided radiation therapy of prostate cancer," Medical Physics, vol. 50, no. 8, pp. 4854–4870, 2023. [Online]. Available: https://doi.org/10.1002/mp.16320

[140] J. Zhong, L. H. Staib, R. Venkataraman, and J. A. Onofrey, "Integrating Prostate Specific Antigen Density Biomarker Into Deep Learning Prostate MRI Lesion Segmentation Models," in 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI), 2023, pp. 1–5.

[141] Y. Li, Y. Wu, M. Huang, Y. Zhang, and Z. Bai, "Attention-guided multi-scale learning network for automatic prostate and tumor segmentation on MRI," Computers in Biology and Medicine, vol. 165, p. 107374, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482523008399

[142] Y. Li, M. Huang, Y. Zhang, S. Feng, J. Chen, and Z. Bai, "A dual attention-guided 3D convolution network for automatic segmentation of prostate and tumor," Biomedical Signal Processing and Control, vol. 85, p. 104755, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S174680942300188X

[143] C.-C. Lai, H.-K. Wang, F.-N. Wang, Y.-C. Peng, T.-P. Lin, H.-H. Peng, and S.-H. Shen, "Autosegmentation of Prostate Zones and Cancer Regions from Biparametric Magnetic Resonance Images by Using Deep-Learning-Based Neural Networks," 2021.

[144] P. Mehta, M. Antonelli, S. Singh, N. Grondecka, E. W. Johnston, H. U. Ahmed, M. Emberton, S. Punwani, and S. Ourselin, "AutoProstate: Towards Automated Reporting of Prostate MRI for Prostate Cancer Assessment Using Deep Learning," 2021.

[145] L. Zhu, G. Gao, Y. Zhu, C. Han, X. Liu, D. Li, W. Liu, X. Wang, J. Zhang, and X.-y. Wang, "Fully automated detection and localization of clinically significant prostate cancer on MR images using a cascaded convolutional neural network," Frontiers in Oncology, vol. 12, sep 2022.

[146] I. Shiri, E. Showkatian, R. Mohammadi, B. Razeghi, S. Bagheri, G. Hajianfar, Y. Salimi, M. Amini, M. G. Oghli, S. Ferdowsi, S. Voloshynovskiy, and H. Zaidi, "Collaborative Multi-Institutional Prostate Lesion Segmentation from MR images Using Deep Federated Learning Framework," in 2022 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2022, pp. 1–3.

[147] S. Mehralivand, D. Yang, S. A. Harmon, D. Xu, Z. Xu, H. Roth, S. Masoudi, T. H. Sanford, D. Kesani, N. S. Lay, M. J. Merino, B. J. Wood, P. A. Pinto, P. L. Choyke, and B. Turkbey, "A Cascaded Deep Learning–Based Artificial Intelligence Algorithm for Automated Lesion Detection and Classification on Biparametric Prostate Magnetic Resonance Imaging," Academic Radiology, vol. 29, no. 8, pp. 1159–1168, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1076633221003779

[148] J. M. Jaen-Lorites, S. Ruiz-Espana, T. Pineiro-Vidal, J. M. Santabarbara, A. M. Maceira, and D. Moratal, "Multiclass Classification of Prostate Tumors Following an MR Image Analysis-Based Radiomics Approach." Annual International Conference of the IEEE Engineering in Medicine and Biology Society. I vol. 2022, pp. 1436–1439, jul 2022.

[149] T. Sanford, S. A. Harmon, E. B. Turkbey, D. Kesani, S. Tuncer, M. Madariaga, C. Yang, J. Sackett, S. Mehralivand, P. Yan, S. Xu, B. J. Wood, M. J. Merino, P. A. Pinto, P. L. Choyke, and B. Turkbey, "Deep-Learning-Based Artificial Intelligence for PI-RADS Classification to Assist Multiparametric Prostate MRI Interpretation: A Development Study," Journal of Magnetic Resonance Imaging, vol. 52, no. 5, pp. 1499–1507, nov 2020. [Online]. Available: https://doi.org/10.1002/jmri.27204

[150] S. Y. Youn, M. H. Choi, D. H. Kim, Y. J. Lee, H. Huisman, E. Johnson, T. Penzkofer, I. Shabunin, D. J. Winkel, P. Xing, D. Szolar, R. Grimm, H. von Busch, Y. Son, B. Lou, and A. Kamen, "Detection and PI-RADS classification of focal lesions in prostate MRI: Performance comparison between a deep learning-based algorithm (DLA) and radiologists with various levels of experience," European Journal of Radiology, vol. 142, p. 109894, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0720048X21003752

[151] K. S. Zhang, P. Schelb, S. Kohl, J. P. Radtke, M. Wiesenfarth, L. Schimmöller, T. A. Kuder, A. Stenzinger, M. Hohenfellner, H.-P.

Schlemmer, K. Maier-Hein, and D. Bonekamp, "Improvement of PI-RADS-dependent prostate cancer classification by quantitative image assessment using radiomics or mean ADC," Magnetic Resonance Imaging, vol. 82, pp. 9–17, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0730725X21001016

[152] P. A. Bonaffini, E. De Bernardi, A. Corsi, P. N. Franco, D. Nicoletta, R. Muglia, G. Perugini, M. Roscigno, M. Occhipinti, L. F. Da Pozzo, and S. Sironi, "Towards the Definition of Radiomic Features and Clinical Indices to Enhance the Diagnosis of Clinically Significant Cancers in PI-RADS 4 and 5 Lesions," 2023.

[153] J. Wang, C.-J. Wu, M.-L. Bao, J. Zhang, X.-N. Wang, and Y.-D. Zhang, "Machine learning-based analysis of MR radiomics can help to improve the diagnostic performance of PI-RADS v2 in clinically relevant prostate cancer." European radiology, vol. 27, no. 10, pp. 4082–4090, oct 2017.

[154] M. Li, L. Yang, Y. Yue, J. Xu, C. Huang, and B. Song, "Use of Radiomics to Improve Diagnostic Performance of PI-RADS v2.1 in Prostate Cancer." Frontiers in oncology, vol. 10, p. 631831, 2020.

[155] J. Bao, X. Qiao, Y. Song, Y. Su, L. Ji, J. Shen, G. Yang, H. Shen, X. Wang, and C. Hu, "Prediction of clinically significant prostate cancer using radiomics models in real-world clinical practice: a retrospective multicenter study." Insights into imaging, vol. 15, no. 1, p. 68, feb 2024.

[156] G. Nketiah, M. R. S. Sunoqrot, E. Sandsmark, S. Langørgen, K. Selnæs, H. Bertilsson, M. Elschot, and T. Bathen, Deep Radiomics Detection of Clinically Significant Prostate Cancer on Multicenter MRI: Init oct 2024.

[157] V. Brancato, M. Aiello, L. Basso, S. Monti, L. Palumbo, G. Di Costanzo, M. Salvatore, A. Ragozzino, and C. Cavaliere, "Evaluation of a multiparametric MRI radiomic-based approach for stratification of equivocal PI-RADS

3 and upgraded PI-RADS 4 prostatic lesions." Scientific reports, vol. 11, no. 1, p. 643, jan 2021.

[158] D. Singh, V. Kumar, C. J. Das, A. Singh, and A. Mehndiratta, "Machine learning-based analysis of a semi-automated PI-RADS v2.1 scoring for prostate cancer." Frontiers in oncology, vol. 12, p. 961985, 2022.

[159] B. Oerther, H. Engel, C. Wilpert, A. Nedelcu, A. Sigle, R. Grimm, H. von Busch, C. L. Schlett, F. Bamberg, M. Benndorf, J. Herrmann, K. Nikolaou, B. Amend, C. Bolenz, C. Kloth, M. Beer, and D. Vogele, "Multi-Center Benchmarking of a Commercially Available Artificial Intelligence Algorithm for Prostate Imaging Reporting and Data System (PI-RADS) Score Assignment and Lesion Detection in Prostate MRI." Cancers, vol. 17, no. 5, feb 2025.

[160] C. Yang, B. Li, Y. Luan, S. Wang, Y. Bian, J. Zhang, Z. Wang, B. Liu, X. Chen, M. Hacker, Z. Li, X. Li, and Z. Wang, "Deep learning model for the detection of prostate cancer and classification of clinically significant disease using multiparametric MRI in comparison to PI-RADs score," Urologic Oncology: Seminars and Original Investigations, vol. 42, no. 5, pp. 158.e17–158.e27, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1078143924000231

[161] X. Wei, J. Xu, S. Zhong, J. Zou, Z. Cheng, Z. Ding, and X. Zhou, "Diagnostic value of combining PI-RADS v2.1 with PSAD in clinically significant prostate cancer," Abdominal Radiology, vol. 47, no. 10, pp. 3574–3582, 2022. [Online]. Available: https://doi.org/10.1007/s00261-022-03592-4

[162] Y. Kan, Q. Zhang, J. Hao, W. Wang, J. Zhuang, J. Gao, H. Huang, J. Liang, G. Marra, G. Calleris, M. Oderda, X. Zhao, P. Gontero, and H. Guo, "Clinico-radiological characteristic-based machine learning in reducing unnecessary prostate biopsies of PI-RADS 3 lesions with dual validation," European Radiology, vol. 30, no. 11, pp. 6274–6284, 2020. [Online]. Available: https://doi.org/10.1007/s00330-020-06958-8

[163] S. J. Hectors, C. Chen, J. Chen, J. Wang, S. Gordon, M. Yu, B. Al Hussein Al Awamlh, M. R. Sabuncu, D. J. A. Margolis, and J. C. Hu, "Magnetic Resonance Imaging Radiomics-Based Machine Learning Prediction of Clinically Significant Prostate Cancer in Equivocal PI-RADS 3 Lesions," Journal of Magnetic Resonance Imaging, vol. 54, no. 5, pp. 1466–1473, 2021. [Online]. Available: https://doi.org/10.1002/jmri.27692

[164] K. Taguelmimt, G. Andrade-Miranda, H. Harb, T. T. Thanh, H.-P. Dang, B. Malavaud, and J. Bert, "Towards more reliable prostate cancer detection: Incorporating clinical data and uncertainty in MRI deep learning," Computers in Biology and Medicine, vol. 194, p. 110440, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482525007917

[165] A. Karagoz, D. Alis, M. E. Seker, G. Zeybel, M. Yergin, I. Oksuz, and E. Karaarslan, "Anatomically guided self-adapting deep neural network for clinically significant prostate cancer detection on bi-parametric MRI: a multi-center study," Insights into Imaging, vol. 14, no. 1, p. 110, 2023. [Online]. Available: https://doi.org/10.1186/s13244-023-01439-0

[166] S. Chakraborty, S. S. Rathor, S. Verma, A. Raj, A. Vikram, P. K. Mallick, and A. Kumar, "Deep Learning ResNet Model for Prostate Cancer Classification: A Comparative Study of CNNs," in 2025 International Conference on Intelligent and Cloud Computing (ICoICC), 2025, pp. 1–7.

[167] R. Alfano, D. Soetemans, E. Gibson, M. Gaed, M. Moussa, J. Gomez-Lemus, J. Chin, S. Pautler, and A. Ward, Development of a computer aided diagnosis model for prostate cancer classification on multi- feb 2018.

[168] "Evaluation of prostate segmentation algorithms for MRI: the PROMISE12 challenge." Medical image analysis, vol. 18, no. 2, pp. 359–373, feb 2014.

[169] X. Zhong, R. Cao, S. Shakeri, F. Scalzo, Y. Lee, D. R. Enzmann, H. H. Wu, S. S. Raman, and K. Sung, "Deep transfer learning-based prostate

cancer classification using 3 Tesla multi-parametric MRI," Abdominal Radiology, vol. 44, no. 6, pp. 2030–2039, 2019. [Online]. Available: https://doi.org/10.1007/s00261-018-1824-5

[170] Y. Wang and M. Wang, "Selecting proper combination of mpMRI sequences for prostate cancer classification using multi-input convolutional neuronal network," Physica Medica, vol. 80, pp. 92–100, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1120179720302568

[171] D. Provenzano, O. Melnyk, D. Imtiaz, B. McSweeney, D. Nemirovsky, M. Wynne, M. Whalen, Y. J. Rao, M. Loew, and S. Haji-Momenian, "Machine Learning Algorithm Accuracy Using Single- versus Multi-Institutional Image Data in the Classification of Prostate MRI Lesions," 2023.

[172] R. Islam, A. Imran, and M. F. Rabbi, "Prostate Cancer Detection from MRI Using Efficient Feature Extraction with Transfer Learning," Prostate Cancer, vol. 2024, no. 1, p. 1588891, 2024. [Online]. Available: https://doi.org/10.1155/2024/1588891

[173] N. M. Al-Zidi and D. Vasumathi, "Identifying Optimal Multiparametric MRI Sequence Combinations for Prostate Cancer Classification: An Integrated Deep Feature and Machine Learning Approach," in 2024 5th IEEE Global Conference for Advancement in Technology (GCAT), 2024, pp. 1–7.

[174] A. Horasan and A. Güneş, "Advancing Prostate Cancer Diagnosis: A Deep Learning Approach for Enhanced Detection in MRI Images," 2024.

[175] Y. Li, J. Wynne, J. Wang, J. Roper, C.-W. Chang, A. B. Patel, J. Shelton, T. Liu, H. Mao, and X. Yang, "MRI-based prostate cancer classification using 3D efficient capsule network," Medical Physics, vol. 51, no. 7, pp. 4748–4758, 2024. [Online]. Available: https://doi.org/10.1002/mp.16975

[176] R. M. Y. Baazeem, "Efficient Classification of Prostate Cancer Using Artificial Intelligence Techniques," SN Computer Science, vol. 5, no. 4, p. 389, 2024. [Online]. Available: https://doi.org/10.1007/s42979-024-02745-0

[177] D. Sowmya, S. A. Bhavani, V. V. S. Sasank, and T. S. Rao, "Prostate cancer classification using adaptive swarm Intelligence based deep attention neural network," Biomedical Signal Processing and Control, vol. 96, p. 106654, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809424007122

[178] J. O. Barentsz, J. C. Weinreb, S. Verma, H. C. Thoeny, C. M. Tempany, F. Shtern, A. R. Padhani, D. Margolis, K. J. Macura, M. A. Haider, F. Cornud, and P. L. Choyke, "Synopsis of the PI-RADS v2 Guidelines for Multiparametric Prostate Magnetic Resonance Imaging and Recommendations for Use." pp. 41–49, jan 2016.

[179] E. Gibson, W. Li, C. Sudre, L. Fidon, D. I. Shakir, G. Wang, Z. Eaton-Rosen, R. Gray, T. Doel, Y. Hu, T. Whyntie, P. Nachev, M. Modat, D. C. Barratt, S. Ourselin, M. J. Cardoso, and T. Vercauteren, "NiftyNet: a deep-learning platform for medical imaging," Computer Methods and Programs in Biomedicine, vol. 158, pp. 113–122, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169260717311823

[180] Chen Peng, "MetaNet: Network Analysis for Omics Data," 2023. [Online]. Available: https://github.com/Asa12138/MetaNet

[181] C. Zhang, Y. Song, S. Liu, S. Lill, C. Wang, Z. Tang, Y. You, Y. Gao, A. Klistorner, M. Barnett, and W. Cai, MS-GAN: GAN-Based Semantic Segmentation of Multiple Sclerosis Lesions in Brain Magnet dec 2018.

[182] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.

[183] O. Oktay, J. Schlemper, L. L. Folgoc, M. J. Lee, M. P. Heinrich, K. Misawa, K. Mori, S. G. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention u-net: Learning where to look

for the pancreas," ArXiv, vol. abs/1804.03999, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:4861068

[184] V. Mnih, N. Heess, A. Graves, and K. Kavukcuoglu, "Recurrent Models of Visual Attention," Advances in Neural Information Processing Systems, vol. 3, jun 2014.

[185] N. Abraham and N. M. Khan, "A novel focal tversky loss function with improved attention u-net for lesion segmentation," in 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), 2019, pp. 683–687.

[186] H. Chen, Y. Wang, J. Guo, and D. Tao, "VanillaNet: the Power of Minimalism in Deep Learning," in Advances in Neural Information Processing Systems, A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, Eds., vol. 36. Curran Associates, Inc., 2023, pp. 7050–7064. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2023/file/16336d94a5ffca8de019087ab7fe403f-Paper-Conference.pdf

[187] F. Milletari, N. Navab, and S. A. Ahmadi, "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," in 2016 Fourth International Conference on 3D Vision (3DV), 2016, pp. 565–571.

[188] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation BT - Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016," S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. Unal, and W. Wells, Eds. Cham: Springer International Publishing, 2016, pp. 424–432.

[189] M. Buscema, "Self-Reflexive Networks," Substance use misuse, vol. 33, pp. 409–438, feb 1998.

[190] ——, "MetaNet: the theory of independent judges." Substance use misuse, vol. 33, no. 2, pp. 439–461, jan 1998.

[191] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779–788. [Online]. Available: http://doi.ieeecomputersociety.org/10.1109/CVPR.2016.91

[192] L. Zhao and S. Li, "Object Detection Algorithm Based on Improved YOLOv3," Electronics, vol. 9, p. 537, mar 2020.

[193] M. Hussain, "YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection," Machines and Tooling, vol. 11, p. 677, jun 2023.

[194] J. Jocher, G.; Chaurasia, A.; Qiu, "YOLO by Ultralytics," 2023. [Online]. Available: https://github.com/ultralytics/ultralytics

[195] A. Sahafi, A. Koulaouzidis, and M. Lalinia, "Polypoid Lesion Segmentation Using YOLO-V8 Network in Wireless Video Capsule Endoscopy Images," 2024.

[196] G. Jocher, A. Chaurasia, A. Stoken, J. Borovec, NanoCode012, Y. Kwon, K. Michael, TaoXie, J. Fang, Imyhxy, Lorna, Yifu), C. Wong, A. V, D. Montes, Z. Wang, C. Fati, J. Nadar, Laughing, UnglvKitDe, V. Sonck, Tkianai, YxNONG, P. Skalski, A. Hogan, D. Nair, M. Strobel, and M. Jain, "ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation," nov 2022. [Online]. Available: https://doi.org/10.5281/zenodo.7347926

[197] H. Ren, C. Ren, Z. Guo, G. Zhang, X. Luo, Z. Ren, H. Tian, W. Li, H. Yuan, L. Hao, J. Wang, and M. Zhang, "A novel approach for automatic segmentation of prostate and its lesion regions on magnetic resonance imaging," Frontiers in Oncology, vol. 13, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fonc.2023.1095353

[198] P. Schelb, A. A. Tavakoli, T. Tubtawee, T. Hielscher, J.-P. Radtke, M. Görtz, V. Schütz, T. A. Kuder, L. Schimmöller, A. Stenzinger, M. Hohenfellner, H.-P. Schlemmer, and D. Bonekamp, "Comparison of Prostate MRI Lesion Segmentation Agreement Between Multiple Radiologists and a Fully Automatic Deep Learning System." RoFo : Fortschritte auf dem Gebiete der Rontgenstrahlen und der Nuklearmedizin, vol. 193, no. 5, pp. 559–573, may 2021.

[199] S. M. Rezaeijo, S. Jafarpoor Nesheli, M. Fatan Serj, and M. J. Tahmasebi Birgani, "Segmentation of the prostate, its zones, anterior fibromuscular stroma, and urethra on the MRIs and multimodality image fusion using U-Net model." Quantitative imaging in medicine and surgery, vol. 12, no. 10, pp. 4786–4804, oct 2022.

[200] S. Cai, Y. Tian, H. Lui, H. Zeng, Y. Wu, and G. Chen, "Dense-UNet: a novel multiphoton in vivo cellular image segmentation model based on a convolutional neural network." Quantitative imaging in medicine and surgery, vol. 10, no. 6, pp. 1275–1285, jun 2020.

[201] S. Fouladi, L. Di Palma, F. Darvizeh, D. Fazzini, A. Maiocchi, S. Papa, G. Gianini, and M. Alì, "Neural network models for prostate zones segmentation in magnetic resonance imaging," Information, vol. 16, no. 3, 2025. [Online]. Available: https://www.mdpi.com/2078-2489/16/3/186

[202] R. C. B. Z. M. T. N. Tekli, Management of Digital EcoSystems, 2024. [Online]. Available: http://link.springer.com/book/10.1007/978-3-031-51643-6

[203] S. N. Almuayqil, R. Arnous, N. Sakr, and M. M. Fadel, "A new hybrid model for segmentation of the skin lesion based on residual attention u-net," Computers, Materials and Continua, vol. 75, no. 3, pp. 5177–5192, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1546221823008172

[204] M. D. Alahmadi, "Multiscale attention u-net for skin lesion segmentation," IEEE Access, vol. 10, pp. 59 145–59 154, 2022.

[205] D.-M. Koh and D. J. Collins, "Diffusion-weighted MRI in the body: applications and challenges in oncology." AJR. American journal of roentgenology, vol. 188, no. 6, pp. 1622–1635, jun 2007.

[206] M. C. DeLano, T. G. Cooper, J. E. Siebert, M. J. Potchen, and K. Kuppusamy, "High-b-value diffusion-weighted MR imaging of adult brain: image contrast and apparent diffusion coefficient map features." AJNR. American journal of neuroradiology, vol. 21, no. 10, pp. 1830–1836, 2000.

[207] L. G. Nyúl, J. K. Udupa, and X. Zhang, "New variants of a method of MRI scale standardization." pp. 143–150, feb 2000.

[208] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: application to breast MR images," IEEE Transactions on Medical Imaging, vol. 18, no. 8, pp. 712–721, 1999.

[209] R. Kikinis, S. Pieper, and K. Vosburgh, "3D Slicer: A Platform for Subject-Specific Image Analysis, Visualization, and Clinical Support," in Intraoper Imaging Image-Guided Ther, jan 2014, vol. 3, pp. 277–289.

[210] R. T. Shinohara, E. M. Sweeney, J. Goldsmith, N. Shiee, F. J. Mateen, P. A. Calabresi, S. Jarso, D. L. Pham, D. S. Reich, and C. M. Crainiceanu, "Statistical normalization techniques for magnetic resonance imaging," NeuroImage: Clinical, vol. 6, pp. 9–19, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S221315821400117X

[211] M. Eklund, F. Jäderling, A. Discacciati, M. Bergman, M. Annerstedt, M. Aly, A. Glaessgen, S. Carlsson, H. Grönberg, and T. Nordström, "MRI-Targeted or Standard Biopsy in Prostate Cancer Screening." The New England journal of medicine, vol. 385, no. 10, pp. 908–920, sep 2021.

[212] A. B. Rosenkrantz, L. A. Ginocchio, D. Cornfeld, A. T. Froemming, R. T. Gupta, B. Turkbey, A. C. Westphalen, J. S. Babb, and D. J. Margolis, "Interobserver Reproducibility of the PI-RADS Version 2 Lexicon: A Multicenter Study of Six Experienced Prostate Radiologists," Radiology, vol. 280, no. 3, pp. 793–804, apr 2016. [Online]. Available: https://doi.org/10.1148/radiol.2016152542

[213] R. J. Gillies, P. E. Kinahan, and H. Hricak, "Radiomics: Images Are More than Pictures, They Are Data," Radiology, vol. 278, no. 2, pp. 563–577, nov 2015. [Online]. Available: https://doi.org/10.1148/radiol.2015151169

[214] J. O. Barentsz, J. Richenberg, R. Clements, P. Choyke, S. Verma, G. Villeirs, O. Rouviere, V. Logager, and J. J. Fütterer, "ESUR prostate MR guidelines 2012." European radiology, vol. 22, no. 4, pp. 746–757, apr 2012.

[215] C. Parmar, P. Grossmann, J. Bussink, P. Lambin, and H. J. W. L. Aerts, "Machine Learning methods for Quantitative Radiomic Biomarkers," Scientific Reports, vol. 5, no. 1, p. 13087, 2015. [Online]. Available: https://doi.org/10.1038/srep13087

[216] R. Kohavi, "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection," vol. 14, mar 2001.

[217] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, "Scikit-learn: Machine Learning in Python," J. Mach. Learn. Res., vol. 12, no. null, pp. 2825–2830, nov 2011.

[218] R. Tibshirani, "Regression shrinkage selection via the LASSO," Journal of the Royal Statistical Society Series B, vol. 73, pp. 273–282, jun 2011.

[219] P. Cornford, R. C. N. van den Bergh, E. Briers, T. Van den Broeck, O. Brunckhorst, J. Darraugh, D. Eberli, G. De Meerleer, M. De Santis, A. Farolfi, G. Gandaglia, S. Gillessen, N. Grivas, A. M. Henry, M. Lardas,

G. J. L. H. van Leenders, M. Liew, E. Linares Espinos, J. Oldenburg, I. M. van Oort, D. E. Oprea-Lager, G. Ploussard, M. J. Roberts, O. Rouvière, I. G. Schoots, N. Schouten, E. J. Smith, J. Stranne, T. Wiegel, P.-P. M. Willemse, and D. Tilki, "EAU-EANM-ESTRO-ESUR-ISUP-SIOG Guidelines on Prostate Cancerx2014;2024 Update. Part I: Screening, Diagnosis, and Local Treatment with Curative Intent," European Urology, vol. 86, no. 2, pp. 148–163, aug 2024. [Online]. Available: https://doi.org/10.1016/j.eururo.2024.03.027

[220] R. Cuocolo, M. B. Cipullo, A. Stanzione, L. Ugga, V. Romeo, L. Radice, A. Brunetti, and M. Imbriaco, "Machine learning applications in prostate cancer magnetic resonance imaging." European radiology experimental, vol. 3, no. 1, p. 35, aug 2019.

[221] A. L. Pappula, S. Rasheed, G. Mirzaei, R. C. Petreaca, and R. A. Bouley, "A Genome-Wide Profiling of Glioma Patients with an IDH1 Mutation Using the Catalogue of Somatic Mutations in Cancer Database," 2021.

[222] "Introduction to MLP neural networks BT - Neural Networks in Multidimensional Domains: Fundamentals and New Trends in Modelling and Control," P. Arena, L. Fortuna, G. Muscato, and M. G. Xibilia, Eds. London: Springer London, 1998, pp. 3–11. [Online]. Available: https://doi.org/10.1007/BFb0047684

[223] A. Pinkus, "Approximation theory of the MLP model in neural networks," Acta Numerica, vol. 8, pp. 143–195, 1999. [Online]. Available: https://www.cambridge.org/core/product/18072C558C8410C4F92A82BCC8FC8CF9

[224] L. Breiman, "Random Forests," Machine Learning, vol. 45, pp. 5–32, oct 2001.

[225] A. Dobra, Decision Tree Classification. New York, NY: Springer New York, 2018, pp. 1017–1022. [Online]. Available: https://doi.org/10.1007/978-1-4614-8265-9_554

[226] C. Cortes and V. Vapnik, "Support-vector networks," Chem. Biol. Drug Des., vol. 297, pp. 273–297, jan 2009.

[227] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ser. KDD '16. New York, NY, USA: Association for Computing Machinery, 2016, pp. 785–794. [Online]. Available: https://doi.org/10.1145/2939672.2939785

[228] Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," Journal of Computer and System Sciences, vol. 55, no. 1, pp. 119–139, 1997. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S002200009791504X

[229] J. Friedman, "Greedy Function Approximation: A Gradient Boosting Machine," The Annals of Statistics, vol. 29, nov 2000.

[230] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, LightGBM: A Highly Efficient Gradient Boosting Decision Tree, dec 2017.

[231] T. Cover and P. Hart, "Nearest neighbor pattern classification," IEEE Transactions on Information Theory, vol. 13, no. 1, pp. 21–27, 1967.

[232] T. Seidl, "Nearest Neighbor Classification BT - Encyclopedia of Database Systems," L. LIU and M. T. ÖZSU, Eds. Boston, MA: Springer US, 2009, pp. 1885–1890. [Online]. Available: https://doi.org/10.1007/978-0-387-39940-9_561

[233] B. Turkbey, H. Mani, V. Shah, A. R. Rastinehad, M. Bernardo, T. Pohida, Y. Pang, D. Daar, C. Benjamin, Y. L. McKinney, H. Trivedi, C. Chua, G. Bratslavsky, J. H. Shih, W. M. Linehan, M. J. Merino, P. L. Choyke, and P. A. Pinto, "Multiparametric 3T prostate magnetic resonance imaging to detect cancer: histopathological correlation using prostatectomy specimens processed in customized magnetic resonance imaging based molds." The Journal of urology, vol. 186, no. 5, pp. 1818–1824, nov 2011.

[234] B. G. Muller, J. H. Shih, S. Sankineni, J. Marko, S. Rais-Bahrami, A. K. George, J. J. M. C. H. de la Rosette, M. J. Merino, B. J. Wood, P. Pinto, P. L. Choyke, and B. Turkbey, "Prostate Cancer: Interobserver Agreement and Accuracy with the Revised Prostate Imaging Reporting and Data System at Multiparametric MR Imaging," Radiology, vol. 277, no. 3, pp. 741–750, jun 2015. [Online]. Available: https://doi.org/10.1148/radiol.2015142818

[235] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Proceedings of the 31st International Conference on Neural Information Processing Systems, ser. NIPS'17. Red Hook, NY, USA: Curran Associates Inc., 2017, pp. 4768–4777.

[236] K. Aas, M. Jullum, and A. Løland, "Explaining individual predictions when features are dependent: More accurate approximations to Shapley values," Artificial Intelligence, vol. 298, p. 103502, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0004370221000539

# Appendix A

# Embedding a Customized Plugin in 3D-Slicer for Prostate Zonal Segmentation

Figure A.1 shows a screenshot of our plugin's interface in use. The plugin we developed provides an intuitive and user-friendly interface for zonal segmentation of prostate MRI scans. With just one click on the "Apply" button, the plugin performs automated predictions slide by slide across the MRI volume. This allows efficient and accurate segmentation without manual effort for initial labeling. To enhance flexibility, the plugin includes a set of manual editing tools, such as:

- Paint and Brush tools for refining segmentation areas,
- An Eraser for removing incorrect predictions.

Navigation through the MRI slices is made easy with arrow buttons, allowing the operator or radiologist to move seamlessly between images. After reviewing and adjusting the segmentation, users can save their modifications directly within the plugin. Each anatomical zone is clearly distinguished using two classes with different colors:

- Central Gland
- Peripheral Zone

These zones can be analyzed and edited separately, giving the user full control

and precision during the segmentation process. Overall, the plugin is designed to be highly accessible, requiring minimal training, and is tailored to meet the practical needs of radiologists and medical imaging professionals. Its clear layout and easy-to-use features make it a powerful tool for clinical and research environments alike.

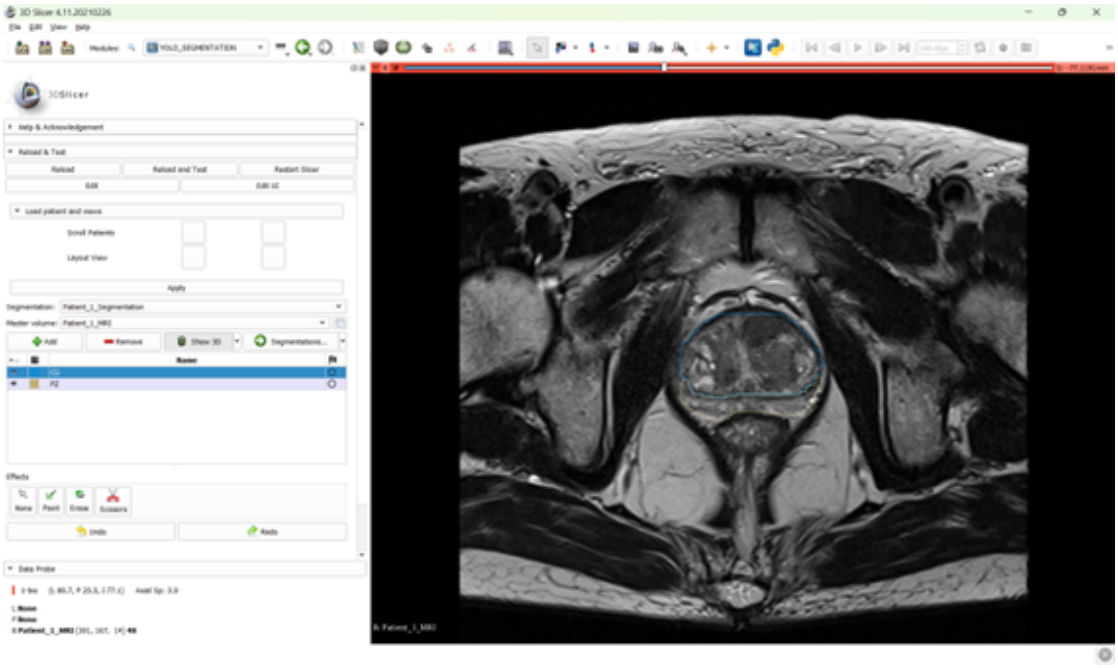


Figure A.1: *A screenshot of our plugin's interface*