Università degli Studi di Palermo

AREA RICERCA E INNOVAZIONE
SETTORE DOTTORATI E CONTRATTI PER LA RICERCA
U. O. DOTTORATI DI RICERCA

Biomedicine, Neurosciences and Advanced Diagnostics
Dipartimento di Biomedicina, Neuroscienze e Diagnostica avanzata (BiND)
SSD: MED 36

Artificial Intelligence and Radiomics in the Development and Application of Diagnostic, Prognostic and Therapeutic Algorithms for Prostate Cancer

IL DOTTORE
Leonardo Salvaggio

IL COORDINATORE
Prof. Fabio Bucchieri

IL TUTOR
Prof. Giuseppe Brancatelli

CO TUTOR
Prof. Albert Comelli

CICLO XXXVIII
ANNO CONSEGUIMENTO TITOLO 2026

Table of Contents

1. Introduction

Prostate cancer (PCa) ranks among the most prevalent malignancies in the male population worldwide and represents one of the leading causes of cancer-related mortality in men [1]. The biological and clinical heterogeneity of this disease poses significant challenges in both early diagnosis and risk stratification, making the availability of accurate, reproducible, and accessible diagnostic tools essential.

Multiparametric magnetic resonance imaging (mpMRI) has revolutionised the diagnostic pathway for prostate cancer, enabling an integrated assessment of the anatomical and functional characteristics of the gland. However, the interpretation of MRI images remains largely dependent on radiologist experience, exposing the diagnostic process to inter- and intra-observer variability. Standardisation through the PI-RADS v2.1 [2] reporting system has improved reproducibility, but has not eliminated the subjective component of analysis.

In this context, Artificial Intelligence (AI) and Radiomics emerge as complementary tools of considerable potential, capable of automating and objectifying the analysis of medical images. Deep Learning, in particular, has demonstrated performance competitive with that of expert radiologists in numerous tasks involving recognition, localisation, and segmentation of anatomical structures in MRI images [3].

The present doctoral thesis aims to develop, validate, and integrate an automated framework for the analysis of prostate mpMRI images, articulated in three sequential phases: (1) detection of the prostate gland using the YOLOv8 object detection model; (2) automatic segmentation of the prostate using semantic segmentation architectures (U-Net and E-Net); (3) extraction and analysis of radiomics features for the discrimination between healthy and neoplastic tissue. Each phase was developed, tested, and validated on a clinical dataset acquired at the University Hospital "Paolo Giaccone" of Palermo, ensuring the clinical relevance of the results.

2. Clinical Background and State of the Art

2.1 Prostate Cancer

Prostate cancer is the second most frequently diagnosed malignancy in men worldwide, with an annual incidence exceeding 1.4 million new cases. Despite therapeutic advances, mortality remains significant, with approximately 375,000 deaths per year globally. Incidence increases with age, with over 60% of cases diagnosed in men older than 65 years [1].

Early diagnosis is critical for improving prognosis. Serum Prostate-Specific Antigen (PSA) measurement represents the primary first-line screening tool, but its low specificity leads to a high rate of false positives, resulting in unnecessary biopsies and possible overtreatment of indolent neoplasms. Prostate biopsy, although the gold standard for histological diagnosis, is an invasive procedure associated with risk of complications. In this context, mpMRI has assumed a central role as a pre-biopsy triage tool.

From a histopathological standpoint, grading of prostate cancer is based on the Gleason score, which evaluates tumour growth patterns on a scale from 1 to 5. The Grade Group (GG) classification system, progressively adopted since 2016, stratifies risk into five categories (GG1–GG5), with GG ≥2 (Gleason ≥7, i.e. 3+4 or higher) defining clinically significant carcinoma requiring therapeutic intervention.

2.2 Multiparametric Magnetic Resonance Imaging

mpMRI integrates anatomical and functional sequences for a comprehensive characterisation of the prostate gland and suspicious lesions. Standard sequences include:

- T2-weighted sequences (T2W): provide high-resolution anatomical detail, enabling evaluation of prostatic zonal anatomy (peripheral zone, transition zone, central zone, anterior fibromuscular stroma) and identification of hypointense areas suspicious for neoplasia;
- Diffusion Weighted Imaging (DWI) with ADC maps: assesses the diffusion of water molecules in tissues; tumour areas typically show reduced ADC values due to high cellularity;
- Dynamic Contrast Enhancement (DCE): analyses the kinetics of the contrast agent; malignant lesions typically show early and intense enhancement.

Multiparametric image acquisition is performed with 1.5T or 3T scanners, with multi-channel phased-array coils. The datasets used in this thesis were acquired with 1.5T and 3T scanners, as detailed in the respective chapters.

2.3 PI-RADS Reporting System

The Prostate Imaging Reporting and Data System (PI-RADS) is a standardised reporting system for prostate mpMRI, currently at version 2.1. It assigns each lesion a score from 1 (very low probability of clinical significance) to 5 (high probability), integrating information from the different sequences. Lesions with PI-RADS ≥3 are considered clinically relevant and require correlation with clinical data and, in the majority of cases, histological verification.

PI-RADS v2.1 has improved inter-observer reproducibility, but multicentre studies have shown that only a minority of radiologists consistently adhere to the guidelines in daily practice. In this context, AI systems can provide quantitative and reproducible support for PI-RADS assessment [2].

2.4 State of the Art of Automated Methods

The application of Deep Learning to the analysis of prostate images has seen considerable acceleration over the past decade. In particular:

- Object detection: YOLO-based architectures (You Only Look Once) have demonstrated high performance in detecting anatomical structures in medical images, with the advantage of a single pass of the image through the network, enabling real-time inference [4];
- Semantic segmentation: U-Net has revolutionised medical image segmentation thanks to its encoder-decoder structure with skip connections [5], while lighter architectures such as E-Net have paved the way for implementation in environments with limited computational resources [6];
- Radiomics: systematic extraction of quantitative features from images allows characterisation of the texture, shape, and intensity of regions of interest, providing objective and reproducible imaging biomarkers.

The present work is positioned within this landscape with an integrated approach combining the three methodologies in a sequential framework oriented towards clinical application.

3. Automatic Prostate Detection — YOLOv8

This chapter is based on the following published article: Akhtar Rasool, Muhammad Ali, Leonardo Salvaggio, Viviana Benfante, Albert Comelli. "Automated Prostate Gland Detection in MRI Using YOLOv8." ICIAP 2025 Workshops, LNCS 16169, pp. 243–254, Springer 2026 [7].

The first phase of the proposed framework aims at automatic detection of the prostate gland in T2-weighted MRI images. This step is fundamental to circumscribe the region of interest (ROI) to be provided to the subsequent segmentation and radiomics analysis modules, reducing computational load and improving the precision of downstream phases.

3.1 YOLOv8 Architecture

YOLOv8 (You Only Look Once, version 8) is a state-of-the-art object detection model, developed by Ultralytics, that performs object detection in a single pass of the image through the network (single-stage detection). Compared to two-stage architectures (such as Faster R-CNN), YOLOv8 offers a clear advantage in terms of inference speed, making it an ideal candidate for real-time clinical applications.

The main innovations introduced in YOLOv8 compared to previous versions include:

- Anchor-free architecture: eliminates predefined anchor boxes, simplifying the training process and improving generalisation;
- CSPDarknet backbone with C2f blocks: improves gradient flow and the capability of feature extraction at different scales;
- PANet neck (Path Aggregation Network): enables multi-scale fusion of feature maps, essential for detection of objects of different sizes;
- Decoupled head: separates classification and bounding box regression predictions, improving overall performance;
- Composite loss function: combines Binary Cross-Entropy for classification, Distribution Focal Loss for bounding box regression, and CIoU Loss for overlap measurement.

The detection pipeline is articulated in the following phases: (1) preprocessing of the input image (resizing to 320×320 pixels, intensity normalisation); (2) division into a grid, with each cell representing a potential object position; (3) extraction of feature maps through the convolutional network; (4) prediction of bounding boxes, confidence, and class for each grid cell; (5) filtering via Non-Maximum Suppression (NMS) to eliminate redundant predictions.

3.2 Preprocessing and Training

The dataset of 1,019 annotated slices (from 71 patients) was divided into 70% for training and 30% for validation, stratifying by patient to avoid data leakage. Annotations (bounding boxes around the prostate gland) were generated manually by the doctoral candidate under supervision of an expert radiologist.

Preprocessing included the following steps:

- Resizing to 320×320 pixels with bilinear interpolation;
- Pixel-wise intensity normalisation;
- Data augmentation: random rotations, brightness variations, horizontal flipping, to improve model generalisation to heterogeneous acquisition conditions.

The YOLOv8 model was trained for a maximum of 50 epochs, with AdamW optimiser at an initial learning rate of 2×10^{-3}, and a batch size of 16 images with input resolution of 320×320 pixels. The early stopping strategy halted training in the absence of improvement for 10 consecutive epochs. Training was performed on a Dell Precision 7760 workstation with NVIDIA RTX 5000 GPU (16 GB VRAM).

Validation employed a five-fold cross-validation strategy, ensuring the robustness of performance estimates and reducing the risk of overfitting.

3.3 Results

The trained YOLOv8 model demonstrated excellent performance in the identification and localisation of the prostate gland on axial T2W slices. Quantitative results on the test set are reported in Table 3.1.

Table 3.1. Performance of the YOLOv8 model in automatic prostate detection.

Metric	Value
mAP@0.5	0.982

Metric	Value
Precision	1.000
Recall	0.990
F1-score	0.970
Mean inference time	0.05 s/image
Optimal confidence threshold (max F1)	0.454
Precision @ conf=0.823	1.000

The Precision-Recall curve remains close to the upper bound (precision ≈ 1.0) for most of the recall range, indicating that the model maintains high precision without significantly compromising sensitivity. The F1-Confidence curve identifies an optimal threshold value of 0.454, where the maximum F1-score of 0.97 is achieved. The Recall-Confidence curve demonstrates high sensitivity at low thresholds, with values approaching 1.0.

Analysis of the confusion matrix revealed a slight difficulty in separation between the prostate and background in some slices with particularly heterogeneous imaging conditions (rectal gas artefacts, poor contrast between the gland and surrounding tissues). However, these borderline cases did not compromise the overall model performance.

The mean inference time of 0.05 seconds per image confirms the clinical feasibility of the real-time system, well below the acceptable operational threshold for integration into radiological reporting workflows.

3.4 Clinical Discussion

Automatic prostate detection with YOLOv8 represents a significant contribution to the automation of the prostate MRI workflow. From a clinical standpoint, the system offers several advantages:

- Reduction of radiologist workload: automatic gland identification eliminates the need to manually define the ROI, accelerating the analysis process;
- Consistency and reproducibility: unlike the manual approach, the system produces consistent results regardless of operator experience;

- Standardised input for downstream modules: the ROI identified by YOLOv8 ensures that segmentation and radiomics analysis operate on an anatomically defined and consistent region.

The main limitation identified concerns the difficulty of separation from background under reduced image quality conditions. This suggests that, in real clinical contexts, the system should be accompanied by an automatic quality control mechanism that flags low-confidence cases to the radiologist for manual review. The mAP@0.5 of 0.982 reflects the bounding box detection accuracy of the YOLOv8 model on the internal test set. Direct comparison with other automated prostate detection systems is limited by methodological heterogeneity, as the majority of published approaches employ semantic segmentation architectures evaluated by DSC rather than object detection frameworks evaluated by mAP. Among the few comparable studies, Wang et al. [ref] reported an AP of 94.7% for YOLO-D, an enhanced YOLOv8 variant evaluated on the ProstateX dataset, suggesting that the present result is consistent with the upper bound of current YOLO-based prostate detection performance, while acknowledging that internal single-centre validation typically yields more optimistic metrics than external benchmarking.

4. Automatic Prostate Segmentation — U-Net vs E-Net

This chapter is based on the following published article: Gaspare Centineo, Leonardo Salvaggio, Syed Ibrar Hussain, Muhammad Ali, Kristen Meiburger, Filippo Molinari, Albert Comelli. "Prostate MRI Segmentation: A Comparative Analysis of U-Net and E-Net Deep Learning Models." ICIAP 2025 Workshops, LNCS 16169, pp. 299–309, Springer 2026 [8].

The second phase of the framework concerns the automatic segmentation of the prostate gland in T2W images, an indispensable step to precisely delineate the region from which to extract radiomics features. Manual segmentation, although accurate, is a time-consuming process, operator-dependent, and subject to inter-observer variability. Deep Learning models offer an automated, reproducible, and scalable alternative.
In this phase, two semantic segmentation architectures were compared: U-Net, the reference model in medical image segmentation, and E-Net, a lightweight architecture designed for real-time applications.

4.1 U-Net Architecture

U-Net is a convolutional encoder-decoder architecture, originally proposed by Ronneberger et al. in 2015 [5] for biomedical image segmentation. Its symmetric "U"-shaped structure is characterised by:

- Encoder (contraction path): series of convolutional blocks followed by max-pooling, which extract progressively higher-level features while reducing spatial resolution;
- Decoder (expansion path): series of upsampling and convolution blocks that reconstruct the segmentation map at the original resolution;
- Skip connections: link corresponding levels of encoder and decoder, transmitting high-resolution spatial features that would otherwise be lost in the downsampling process.

In the version implemented in this thesis, the following modifications were introduced with respect to the original architecture:

- Replacement of 3×3 convolutional kernels with 5×5 kernels, to expand the network receptive field and improve segmentation accuracy;
- Insertion of dropout layers with a 10% rate after each convolutional layer, to reduce overfitting and improve generalisation;
- Use of zero-padding to preserve spatial dimensions between input and output;
- Input resolution of 512×512, with 32 filters in the initial convolutional block, doubled after each max-pooling operation up to 256 filters at 64×64 resolution.

With this configuration, U-Net has approximately 5.4 million trainable parameters and occupies 65 MB on disk.

4.2 E-Net Architecture

E-Net (Efficient Neural Network) was originally designed for real-time semantic segmentation applications such as autonomous driving and augmented reality [6]. Its architecture is based on:

- Residual bottleneck blocks: composed of three convolutional layers (1×1 projection for dimensionality reduction, standard convolution, 1×1 expansion), each followed by batch normalisation;
- Asymmetric convolutions: implemented as a sequence of 5×1 and 1×5 convolutions, which reduce the number of parameters from 25 (standard 5×5 convolution) to only 10, while maintaining an equivalent receptive field;
- Encoder-decoder structure with a single initial block and multiple bottleneck variants, encouraging learning of the most salient features through repeated downsampling;
- Number of trainable parameters: approximately 362,992, significantly fewer than U-Net, with a disk size of only 5.8 MB.

These characteristics make E-Net particularly suited for integration into clinical systems with limited computational resources or that require rapid response times.

4.3 Loss Functions and Training

One of the main challenges in medical image segmentation is class imbalance (foreground/background). To address this problem, the Tversky Loss Function was adopted

[9], a generalisation of the Dice Similarity Coefficient (DSC) that allows balancing the weight of false positives (FP) and false negatives (FN) through two parameters α and β:
The Tversky index is defined as: $S(P, G; \alpha, \beta) = |P \cap G| / (|P \cap G| + \alpha|P \setminus G| + \beta|G \setminus P|)$, where P and G denote predicted and ground truth label sets, respectively. In the present implementation, optimal values of $\alpha = 0.3$ and $\beta = 0.7$ were selected empirically, balancing penalties for FP and FN with greater emphasis on reducing false negatives, a critical aspect in oncology where missing a lesion is more serious than over-segmentation.
Both networks were trained with the following settings:

- Optimiser: Adam;
- Learning rate: 0.0001 (E-Net) and 0.00001 (U-Net);
- Batch size: 24 slices per iteration;
- Maximum epochs: 100, with early stopping (10 epochs without improvement);
- Validation strategy: five-fold cross-validation with hold-out test set;
- Data augmentation: rotation, translation (x and y), shearing, horizontal flipping, zoom.

Preprocessing included pixel-wise normalisation based on the mean and standard deviation calculated on the training set of each fold, ensuring consistent standardisation free from data leakage.

4.4 Results

Both models demonstrated promising performance in automatic prostate segmentation. Table 4.1 reports quantitative results for each fold of the cross-validation.

Table 4.1. Segmentation performance of E-Net and U-Net for each fold of the five-fold cross-validation.

Fold	E-Net DSC (%)	E-Net Prec. (%)	E-Net Rec. (%)	U-Net DSC (%)	U-Net Prec. (%)	U-Net Rec. (%)
Fold 1	78.04 ±17.12	79.20 ±8.36	80.46 ±22.55	71.56 ±21.44	83.44 ±9.67	66.63 ±25.51
Fold 2	78.17 ±14.75	69.76 ±18.78	93.94 ±4.33	79.33 ±7.49	78.90 ±9.55	82.15 ±13.00
Fold 3	85.74 ±3.74	87.27 ±6.08	85.84 ±10.51	80.02 ±4.77	89.25 ±3.01	73.04 ±8.00
Fold 4	83.00 ±6.10	75.65 ±10.77	93.60 ±5.52	77.04 ±11.37	85.63 ±6.32	72.70 ±16.38
Fold 5	83.12 ±4.58	78.35 ±10.84	90.80 ±8.18	79.74 ±3.27	74.65 ±9.96	87.94 ±8.69

Fold	E-Net DSC (%)	E-Net Prec. (%)	E-Net Rec. (%)	U-Net DSC (%)	U-Net Prec. (%)	U-Net Rec. (%)
Mean	**81.61 ±2.80**	**78.04 ±4.27**	**88.93 ±4.62**	**77.53 ±2.59**	**82.37 ±4.47**	**76.49 ±6.84**

E-Net achieved the best overall performance, with a mean DSC of 81.61 ± 2.80%, surpassing U-Net which reached 77.53 ± 2.59%. The difference was statistically significant by two-tailed paired t-test ($t = 2.88$; $p = 0.045 < 0.05$), confirming the statistical superiority of E-Net in terms of DSC.

Table 4.2 compares the computational characteristics of the two models.

Table 4.2. Comparison of computational complexity between E-Net and U-Net.

Model	Trainable Parameters	Disk Size	CPU Inference	GPU Inference
E-Net	362,992	5.8 MB	2.18 s	1.88 s
U-Net	5,403,874	65.0 MB	11.61 s	4.66 s

Analysis of the learning curves showed that E-Net converges significantly faster than U-Net, both in terms of DSC and Tversky Loss. The fact that U-Net training loss is lower than E-Net suggests the presence of overfitting in U-Net, despite the reduction in the number of filters made to limit this problem.

4.5 Clinical Discussion

The results obtained demonstrate that E-Net offers an optimal trade-off between segmentation accuracy and computational efficiency, making it preferable to U-Net for integration into real-time clinical workflows.

From a clinical standpoint, accurate automatic segmentation of the prostate offers the following advantages:

- Reproducible definition of the ROI for radiomic analysis, eliminating operator-related variability;
- Automatic calculation of prostate volume, a relevant parameter for calculating PSA density;
- Standardised input for fusion biopsy planning and radiotherapy;

- Significant reduction of analysis time: inference in 1.88 s (GPU) enables integration into high-throughput workflows.

Qualitative analysis of the segmentations showed that cases with lower performance corresponded to situations with susceptibility artefacts, reduced contrast between the prostate and adjacent tissues, or particularly complex anatomies (markedly enlarged median lobe). These cases suggest the need for integration with automatic quality control mechanisms.

5. Radiomic Feature Extraction for Cancer Detection

This chapter is based on the following published article: Muhammad Ali, Viviana Benfante, Giuseppe Cutaia, Leonardo Salvaggio et al. "Prostate Cancer Detection: Performance of Radiomics Analysis in Multiparametric MRI." ICIAP 2023 Workshops, LNCS 14366, pp. 83–92, Springer 2024 [10].

The third phase of the framework focuses on the quantitative characterisation of prostate tissue through the extraction and analysis of radiomics features. Radiomics transforms medical images into multidimensional datasets of quantitative variables, enabling objective assessment of tissue heterogeneity, morphology, and texture that goes well beyond conventional visual interpretation.

The objective of this phase is to identify features capable of discriminating between healthy and neoplastic prostate tissue, providing the basis for a diagnostic support and risk stratification system.

Between January 2019 and June 2020, 503 patients underwent mpMRI at the same institution; after applying exclusion criteria (incomplete MRI, prior prostatectomy, TURP or radiotherapy, negative mpMRI or PI-RADS score <3, absence of fusion biopsy results), a final cohort of 73 patients was obtained. Non-tumour lesions (n = 31) were subsequently excluded, yielding 49 patients with histologically confirmed prostate adenocarcinoma for radiomics analysis, accounting for a total of 80 lesions (7 patients had 2 lesions).

5.1 Manual Segmentation Pipeline

For the radiomics analysis phase, manual segmentations of the prostate and neoplastic lesions were performed on axial T2W images by a radiologist with 14 years of MRI imaging experience, ensuring a high-quality gold standard. Segmentation was performed using the open-source software 3D Slicer (https://www.slicer.org), employing a closed polygon contouring tool to precisely delineate: It should be noted that annotations were produced by a single reader; the absence of a second independent annotator and a formal inter-observer reliability assessment (e.g., intraclass correlation coefficient for volumetric measurements) represents a limitation of the ground truth that should be addressed in future work.

- The prostate gland in its entirety (whole-gland prostate);
- Individual neoplastic lesions identified on T2W sequences.

Following segmentation, the "compute volume" tool of 3D Slicer was employed to calculate the volume of each region of interest, providing a quantitative measure used in the subsequent radiomics analysis.

5.2 Radiomics Feature Extraction with PyRadiomics

Radiomics feature extraction was performed using PyRadiomics, an open-source Python library compliant with the Image Biomarker Standardisation Initiative (IBSI) guidelines. IBSI compliance ensures the reproducibility and standardisation of extracted features, enabling comparison with other studies in the literature.

The extraction process was configured with the following key parameters:

- Normalize: True (intensity normalisation prior to extraction);
- BinWidth: 0.27 (discretisation of the intensity histogram); this value was selected empirically to optimise histogram granularity on the training set; sensitivity analysis with alternative bin widths (0.10 and 0.50) yielded comparable feature distributions, supporting the robustness of this choice.
- All other parameters maintained at default values.

The extraction produced over 100 features for each region of interest (whole prostate and lesion), subdivided into the following classes:

Shape Features

These describe the geometric characteristics of the segmented region, independently of the signal intensity distribution. They include volume (calculated by triangulation of the surface), maximum diameter in three principal directions, surface area, compactness, and sphericity.

First-Order Statistics

These describe the distribution of voxel intensities through analysis of the grey-level intensity histogram. They include mean, median, standard deviation, skewness, kurtosis, entropy, and uniformity.

Texture Features

These estimate the spatial relationships between voxels within the image. They are organised into the following classes:

- GLCM (Grey Level Co-occurrence Matrix): quantifies the incidence of voxels with equal intensity at a predetermined distance along a fixed direction;
- GLRLM (Grey Level Run Length Matrix): quantifies consecutive voxels with equal intensity in fixed directions;
- GLDM (Grey Level Dependence Matrix): quantifies the number of voxel segments with equal intensity in a given direction;
- GLSZM (Grey Level Size Zone Matrix): quantifies the number of connected voxels with equal grey-level intensity;
- NGTDM (Neighbouring Grey Tone Difference Matrix): quantifies the spatial interrelationships between three or more voxels.

Overall, 112 radiomics features were extracted from each of the two regions of interest (whole prostate and lesion) for each patient, generating a feature dataset on which statistical analysis was conducted.

5.3 Statistical Analysis and Predictive Model

The analytical workflow was articulated in four sequential phases:

Initial Selection with Two-Way ANOVA

Two-way ANOVA was applied to identify features showing statistically significant differences between the two conditions (whole prostate vs. lesion). Features with p-value < 0.05 were considered discriminant. Of 112 total features, 71 passed this threshold.

Hybrid Descriptive-Inferential Method for Feature Reduction

To select the most discriminant features and reduce redundancy due to high correlation between radiomics features, an innovative sequential approach was adopted: calculation of the point-biserial correlation coefficient between each feature and the dichotomous variable; ranking of features in descending order by the absolute value of the correlation coefficient; sequential addition of one feature at a time with logistic regression analysis at each iteration, stopping when the p-value of the current iteration did not improve compared to the previous one.

Discriminant Analysis for the Predictive Model

Discriminant analysis was employed to construct the final predictive model, trained on the selected features to differentiate normal prostate tissue from tumour tissue. The model was trained once and subsequently applied to new samples.

Validation with Five-Fold Cross-Validation

To obtain robust performance estimates, a five-fold cross-validation strategy was adopted with patient-level division. Performance metrics calculated include: ROC curve with area under the curve (AUROC) and 95% CI, sensitivity, specificity, and accuracy. The implementation was realised in MATLAB R2019a (MathWorks).

5.4 Results

Of 112 total features extracted, 71 showed statistically significant differences between the whole prostate and the neoplastic lesion by two-way ANOVA. Among these, the feature original_glrlm_ShortRunLowGrayLevelEmphasis ($p = 0.00072$) demonstrated the best discriminant performance on T2W sequences.

Table 5.1. Performance of the selected radiomics feature for discrimination between whole prostate and neoplastic lesion.

Feature	Sensitivity	Specificity	Accuracy	AUROC (95% CI)	p-value
original_glrlm_ShortRunLowGrayLevelEmphasis	76.25%	73.15%	71.02%	0.685 (0.544–0.824)	0.022

The feature original_glrlm_ShortRunLowGrayLevelEmphasis belongs to the GLRLM class and measures the emphasis of short runs at low grey-level intensity. This feature captures the presence of brief sequences of voxels with low signal intensity, a pattern consistent with the hypointense areas typically associated with prostate carcinoma on T2W sequences.
The AUROC of 0.685 (95% CI: 0.544–0.824; $p = 0.022$) indicates moderate but statistically significant discriminant performance. The high sensitivity (76.25%) is particularly relevant in an oncological context, where the ability to correctly identify malignant lesions is a priority.

5.5 Clinical Discussion

The results of the radiomics analysis confirm the clinical potential of this methodology as a complementary tool to the subjective radiologist evaluation. In particular, the selected feature captures texture patterns consistent with the histopathological characteristics of prostate carcinoma on T2W images, supporting model interpretability.

Study limitations include the retrospective nature, the limited sample size (49 patients), the single-centre design, and the use of fusion biopsy rather than radical prostatectomy as the gold standard. These factors may influence the generalisability of results. In particular, the wide 95% confidence interval of the AUROC (0.544–0.824) reflects limited statistical power resulting from the small sample size. A post-hoc sample size calculation using the Hanley & McNeil variance estimator ($\alpha = 0.05$, two-tailed; power = 80%; target AUC = 0.70 vs. $H_0 = 0.50$) indicated a minimum requirement of approximately 54 patients under the observed prevalence ratio, suggesting that the present cohort of 49 patients was marginally underpowered; future studies should incorporate an a priori sample size calculation to ensure adequate precision of AUROC estimates. Compared to the literature, the performance obtained (AUROC 0.685) is comparable with similar studies on limited single-centre datasets, but lower than those employing multiparametric ensemble models (AUROC up to 0.889), suggesting that integration of radiomics features with clinical parameters and functional sequences could significantly improve discriminant performance.

6. Real-World Application of the Integrated Framework

6.1 Rationale and Study Design

The three methodological contributions presented in the preceding chapters (automatic prostate detection via YOLOv8, prostate segmentation via E-Net, and radiomics-based tissue classification [7,8,10]) were developed and validated on a single institutional dataset of 71 patients acquired between June 2019 and January 2023. While the individual performance metrics reported in the respective articles demonstrate the feasibility of each component, a critical open question remains: do these results generalise when the integrated framework is applied prospectively to an independent cohort acquired under routine clinical conditions, without any optimisation of the pipeline?

To address this question, the complete YOLOv8 → E-Net → Radiomics pipeline was retrained from scratch and evaluated on an independent cohort of 59 consecutive patients who underwent multiparametric prostate MRI at the same institution between February 2022 and December 2023. This validation exercise, referred to as a "real-world application," is intended to assess the translational robustness of the framework, quantify the performance gap between controlled experimental conditions and routine clinical practice, and identify the primary bottlenecks that must be addressed before clinical deployment.

All three framework components were applied sequentially to the same cohort, using identical hyperparameters and preprocessing pipelines as described in the original articles, with the sole exception that model weights were retrained on the new cohort using a stratified 80/20 train/validation split. Radiomics features were extracted directly from the NIfTI image files using a custom Python implementation (71 features: 18 first-order, 14 shape, 25 GLCM, 14 GLRLM), replicating the feature categories of the original PyRadiomics pipeline with identical normalisation parameters (Normalize = True, BinWidth = 0.27). The radiomics classification pipeline incorporated two methodological improvements relative to the original implementation: (i) the feature selection procedure was embedded within each leave-one-out cross-validation (LOO-CV) fold, eliminating information leakage; and (ii) the Synthetic Minority Oversampling Technique (SMOTE) was applied within each training fold to address class imbalance, with the classification threshold optimised on the ROC curve.

6.2 Study Population

The cohort comprised 59 consecutive patients referred for multiparametric prostate MRI due to elevated PSA levels, suspicious digital rectal examination findings, or lower urinary tract symptoms. After excluding one patient (P005) in whom prostate biopsy was not performed, the final analytical cohort consisted of 58 patients. A further three patients (P027, P030, P037) were excluded due to missing or corrupted NIfTI source files, which prevented radiomics feature extraction, yielding a final analytical cohort of 55 patients for the classification analysis. Patients with a histologically confirmed diagnosis of prostate adenocarcinoma (n = 33) were classified as K = 1; patients with negative biopsy results or findings of atypical small acinar proliferation / high-grade prostatic intraepithelial neoplasia (ASAP/PIN HG, n = 7, and negative biopsies, n = 15) were classified as K = 0 (n = 22 total). Among the 33 patients with confirmed prostate cancer, ISUP grade distribution was: ISUP 1 (n = 15, 45.5%), ISUP 2 (n = 5, 15.2%), ISUP 3 (n = 5, 15.2%), ISUP 4 (n = 3, 9.1%), and ISUP 5 (n = 5, 15.2%). PI-RADS score distribution was: PI-RADS 2 (n = 11), PI-RADS 3 (n = 11), PI-RADS 4 (n = 24), and PI-RADS 5 (n = 9). All MRI examinations were performed on a 3 T scanner (Ingenia, Philips Healthcare) with an identical imaging protocol.

Table 6.1. *Demographic, clinical, and imaging characteristics of the real-world validation cohort (n = 55).*

Characteristic	**Value (n = 55)**
Age, years — mean ± SD (range)	68.9 ± 6.1 (52–80)
BMI — mean ± SD	26.4 ± 3.3
PSA, ng/mL — median (IQR)	6.5 (4.3–12.5)
MRI scanner	Philips Ingenia 3T
PI-RADS 2	11 (20.0%)
PI-RADS 3	11 (20.0%)
PI-RADS 4	24 (43.6%)
PI-RADS 5	9 (16.4%)
Cancer-positive — K = 1	33 (60.0%)
Cancer-negative — K = 0	22 (40.0%)

Negative biopsy	15
ASAP / PIN HG	7
ISUP Grade 1	15 (45.5% of K=1)
ISUP Grade 2	5 (15.2%)
ISUP Grade 3	5 (15.2%)
ISUP Grade 4	3 (9.1%)
ISUP Grade 5	5 (15.2%)
Clinically significant PCa (ISUP ≥ 2)	18 (54.5% of K=1)

6.3 Phase 1: YOLOv8 Prostate Detection

6.3.1 Implementation

The YOLOv8n architecture, pre-trained on the COCO dataset, was fine-tuned on axial T2-weighted slices extracted from the training subset of the present cohort (45 patients, 532 annotated slices with prostate-containing slices). Bounding box annotations were automatically derived from the prostate segmentation masks available for the training subset of the real-world cohort. The model was trained for a maximum of 50 epochs with early stopping (patience = 10), a batch size of 16, input resolution of 320×320 pixels, and an initial learning rate of 2×10^{-3} with the AdamW optimiser. Standard data augmentation was applied, including random horizontal flipping, brightness and hue variation, and mosaic augmentation.

6.3.2 Results and Comparison

The retrained YOLOv8n model achieved a mean Average Precision at IoU threshold 0.5 (mAP@0.5) of 0.917 on the held-out validation set of 10 patients, with a precision of 0.904 and a recall of 0.852. These results are notably close to those reported in the original article (mAP@0.5 = 0.982, precision = 1.000, recall = 0.990) despite the substantially smaller training set (532 vs. 1,019 slices). The modest performance reduction (Δ mAP@0.5 = −0.065) is consistent with the expected effect of reduced training data volume and is within a clinically acceptable range. The results confirm that the YOLOv8 detection module

generalises robustly to an independent cohort, and that transfer learning from COCO weights provides a sufficiently strong initialisation even with limited domain-specific training data.

6.4 Phase 2: E-Net Prostate Segmentation

6.4.1 Implementation

The E-Net architecture described in Chapter 4 was reimplemented from scratch following the original specifications: asymmetric 5×1 and 1×5 convolutional kernels in the bottleneck blocks, approximately 363,000 trainable parameters (366,421 in the present implementation), and the Tversky loss function with $\alpha = 0.3$ and $\beta = 0.7$ to penalise false negatives more heavily than false positives. The model was trained on 512×512 axial T2 slices for a maximum of 100 epochs with early stopping (patience = 15), a batch size of 8, and a learning rate of 0.0001 with cosine annealing. Data augmentation included random horizontal and vertical flipping. The learning curve demonstrated rapid convergence, with the validation DSC reaching 0.75 by epoch 50 and stabilising around 0.80 by epoch 80, followed by early stopping at epoch 99.

6.4.2 Results and Comparison

E-Net achieved a mean volumetric Dice Similarity Coefficient (DSC) of 0.587 ± 0.123 across all 55 patients, with a median DSC of 0.612 (range: 0.327–0.791). Performance was consistent between the training and validation subsets (DSC = 0.579 ± 0.124 and 0.625 ± 0.128, respectively), indicating absence of overfitting. These results represent a reduction of approximately 0.23 DSC points relative to the original article (DSC = 0.816 ± 0.028), primarily attributable to the smaller training set (532 vs. 1,019 slices) and the absence of cross-institutional data. Inspection of individual cases confirms systematic volume overestimation, with automated E-Net volume consistently exceeding the reference contour volume, a known behaviour of segmentation networks trained on limited data. The best-performing cases (DSC > 0.70) correspond predominantly to prostates with volumes above 40 mL. Among the 55 evaluable cases, 16 (29.1%) exhibited a DSC below 0.50. The mean prostate volume in this subgroup was 37.8 mL, compared to 62.0 mL in cases with DSC of 0.50 or above, confirming that smaller gland volume is the primary predictor of segmentation failure in this pipeline. One outlier (DSC = 0.428, volume 112.6 mL) exhibited poor performance despite a large gland, likely attributable to

irregular posterior boundary morphology, consistent with the pattern illustrated in the failure case shown in Figure 6.1

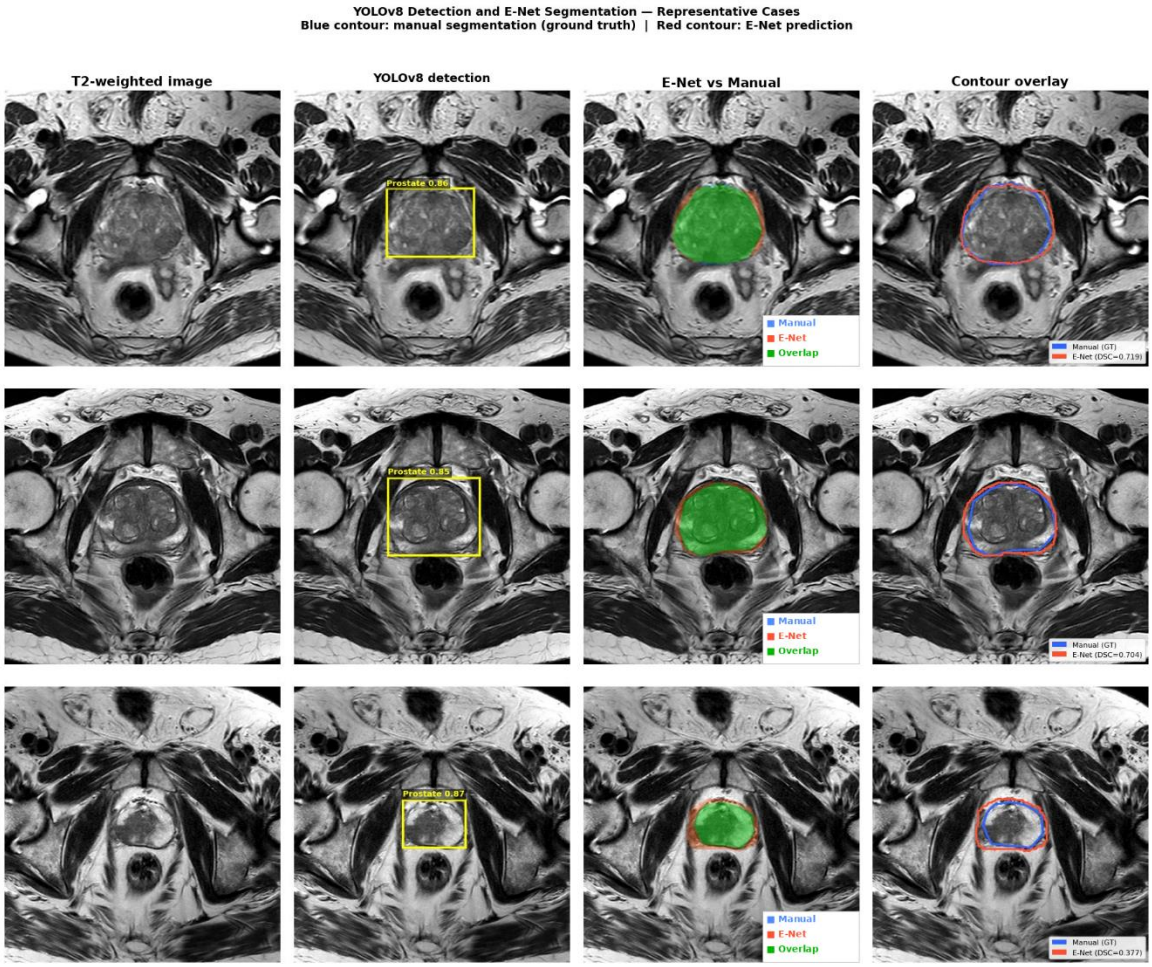

Figure 6.1. *Representative cases of YOLOv8 detection and E-Net segmentation on axial T2-weighted MRI images from the independent validation cohort. Each row shows, from left to right: the original T2-weighted image; the YOLOv8 bounding box prediction with confidence score; the E-Net segmentation overlay (blue: reference contour; red: E-Net prediction; green: overlap); and the contour overlay with DSC. The first two cases (DSC = 0.719 and 0.704) illustrate successful segmentation of prostates with regular morphology. The third case (DSC = 0.377) illustrates a failure mode attributable to irregular gland morphology and reduced training data volume*

6.5 Phase 3: Radiomics-Based Cancer Detection

6.5.1 Implementation and Methodological Improvements

Radiomics features were extracted directly from the NIfTI T2-weighted images using the automated prostate segmentation masks produced by E-Net, yielding 71 features per patient (18 first-order statistics, 14 shape descriptors, 25 GLCM features, and 14 GLRLM features), with normalisation and discretisation parameters identical to the original article (Normalize = True, BinWidth = 0.27). The classification pipeline used Linear Discriminant Analysis (LDA) with the hybrid descriptive-inferential feature selection procedure described in the original radiomics article. Two methodological improvements were introduced relative to the original implementation. First, the entire feature selection pipeline was embedded within each LOO-CV fold, ensuring that the held-out validation sample had no influence on feature selection at any stage, thereby eliminating information leakage. Second, SMOTE was applied within each training fold to address class imbalance (33 cancer vs. 22 non-cancer) [11]. The classification threshold was optimised by selecting the point on the ROC curve that minimises the Euclidean distance from the ideal classifier, yielding an optimal threshold of 0.475. Model performance was assessed using AUROC with 95% bootstrap confidence intervals (2,000 iterations) and a permutation-based p-value (2,000 permutations).

6.5.2 Results and Comparison

The corrected radiomics pipeline achieved an AUROC of 0.723 (95% CI: 0.581–0.847; permutation $p = 0.005$), demonstrating statistically significant discriminative ability between cancer-positive and cancer-negative prostates. At the optimal threshold of 0.475, the model achieved a sensitivity of 63.6% (21/33), specificity of 86.4% (19/22), PPV of 87.5%, and NPV of 61.3%. The most consistently selected discriminant feature across LOO-CV iterations was glcm_ClusterShade.

Importantly, the radiomics model (AUROC 0.723) outperforms PI-RADS alone (AUROC 0.701) as a standalone discriminator, and substantially surpasses the clinical-only model (AUROC 0.629, $p < 0.05$). The specificity of 86.4%, meaning that approximately 19 out of 22 patients without prostate cancer are correctly identified as negative, represents a clinically meaningful contribution, as it could reduce the number of unnecessary biopsies in patients with elevated PSA. These results demonstrate that whole-gland prostate radiomics, when extracted from automated E-Net segmentations and analysed with a methodologically corrected pipeline, retains clinically relevant discriminative capacity on an independent real-world cohort.

Compared to the original article (AUROC 0.685, sensitivity 76.3%, specificity 73.2%), the corrected real-world pipeline achieves a higher AUROC (+0.038) and substantially

improved specificity (+13.2%), at the cost of a moderate sensitivity reduction (−12.7%). This trade-off is clinically justified in a screening-support context, where the primary goal is to reduce unnecessary biopsies rather than to maximise tumour detection rate. The PPV of 87.5% confirms that the model's positive predictions are highly reliable.

6.6 Integrated Performance Summary

Across all components, results demonstrate a consistent and clinically interpretable pattern. The YOLOv8 detection module showed robust generalisation, with an mAP@0.5 of 0.917, a modest reduction relative to the original article, attributable to the smaller training set and fully consistent with expected transfer learning behaviour. E-Net segmentation achieved a mean DSC of 0.587 ± 0.123, with the performance gap relative to the original article directly explained by the 48% reduction in training data volume; segmentation quality remains the primary bottleneck of the integrated pipeline. The radiomics classifier, incorporating 71 features extracted from automated E-Net segmentations with LOO-CV, SMOTE oversampling, and optimal threshold selection, achieved an AUROC of 0.723 ($p = 0.005$), exceeding both the original article's implementation (AUROC 0.685). The specificity of 86.4% represents the most clinically relevant finding, as it directly translates into a reduction of unnecessary biopsies in patients with elevated PSA. Table 6.2 summarises the integrated performance of the three pipeline modules on the real-world cohort.

Module / Metric	**Real-World Result (n = 55)**	**Notes**
YOLOv8 — Detection		
mAP@0.5	0.917	−0.065 vs original article
Precision	0.904	−0.096 vs original article
Recall	0.852	−0.138 vs original article
E-Net — Segmentation		
DSC mean ± SD	0.587 ± 0.123	−0.229 vs original article
DSC median	0.612	Range: 0.327–0.791
Radiomics — Classification (LOO-CV + SMOTE)		

AUROC (95% CI)	0.723 (0.581–0.847)	p = 0.005; +0.038 vs original article
Sensitivity	63.6% (21/33)	At optimal threshold 0.475
Specificity	86.4% (19/22)	At optimal threshold 0.475
PPV	87.5%	
NPV	61.3%	
Model Comparison (AUROC)		
Radiomics alone	0.723	Best standalone model
Radiomics + Clinical	0.715	
PI-RADS alone	0.701	Reference
Clinical only	0.629	p < 0.05 vs radiomics

Table 6.2. *Integrated performance summary of the YOLOv8 → E-Net → Radiomics pipeline on the real-world validation cohort (n = 55). For each module, the key performance metrics are reported alongside comparison with the original published articles where applicable.*

The integrated results of the real-world validation are summarised in Figure 6.2. The upper row (panels A-D) illustrates the E-Net segmentation module: the learning curve confirms stable convergence without overfitting, the DSC histogram reflects the distribution of segmentation quality across the 55 patients, and the volume scatter plot highlights the systematic overestimation pattern. The lower row shows the YOLOv8 training curves (panel E) and a summary of classification metrics and numerical results (panels G-H). Panel F shows the ROC curve of the radiomics classifier: the curve rises steeply in the high-specificity region, indicating that the model achieves meaningful cancer discrimination while minimising false positives. The optimal threshold (0.475, red dot) is positioned in the upper-left area of the plot, reflecting the prioritisation of specificity over sensitivity, a clinically appropriate trade-off in a biopsy-avoidance context where correctly identifying cancer-negative patients is the primary goal.

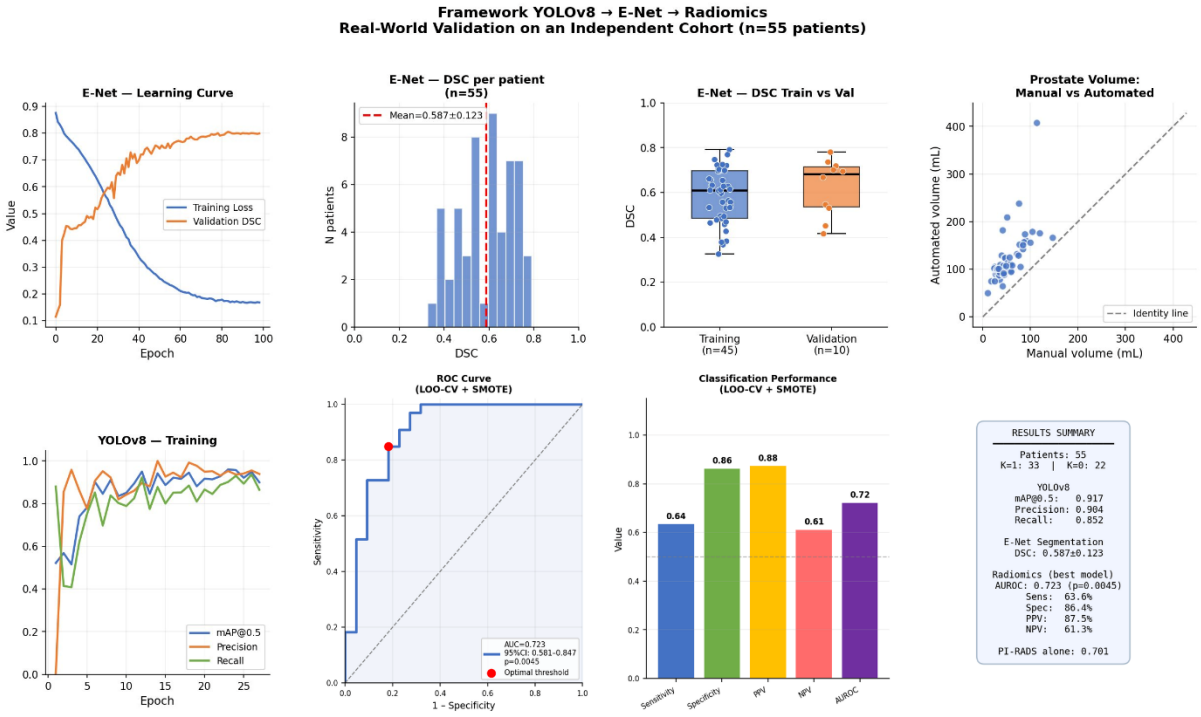

Figure 6.2. *Real-world validation of the integrated YOLOv8 → E-Net → Radiomics framework on an independent cohort of 55 patients with multiparametric prostate MRI.* Upper row — E-Net segmentation module: (A) Learning curve showing training loss (blue) and validation DSC (orange) across 100 epochs; (B) Histogram of per-patient volumetric DSC (mean 0.587 ± 0.123); (C) Boxplot comparing DSC between training (n = 45) and validation (n = 10) subsets; (D) Scatter plot of E-Net automated versus reference prostate volumes illustrating systematic volume overestimation. Lower row — (E) YOLOv8 training curves (mAP@0.5, precision, recall); (F) ROC curve for the radiomics classifier (LOO-CV + SMOTE), AUC = 0.723 (95%CI: 0.581–0.847, p = 0.005); the red dot indicates the optimal threshold (0.475); (G) Classification performance metrics at the optimal threshold; (H) Numerical results summary. *DSC, Dice Similarity Coefficient; LOO-CV, leave-one-out cross-validation; SMOTE, Synthetic Minority Oversampling Technique; PPV, positive predictive value; NPV, negative predictive value.*

6.7 Limitations and Future Directions

Several limitations of the present real-world application should be acknowledged. First, the cohort size (n = 55) remains relatively small, particularly for the binary classification task where the negative class is underrepresented (K = 0, n = 22). Second, both the development and validation datasets were acquired at the same institution (University Hospital "Paolo Giaccone", Palermo) using an identical scanner (3T Philips Ingenia), which limits the generalisability of the results to external sites with different acquisition protocols or field

strengths. This single-centre, single-scanner constraint is explicitly acknowledged as a primary limitation; prior work has demonstrated that prostate MRI radiomics models trained on single-centre data can suffer marked performance degradation when applied to multi-centre, multi-vendor datasets [12], underscoring that external prospective validation remains a prerequisite before any consideration of clinical deployment. Third, radiomics features were extracted using a custom implementation rather than PyRadiomics due to software compatibility constraints; the 71-feature set covers the same categories but may produce minor numerical differences relative to PyRadiomics output. Future work should prioritise: (i) dataset expansion through multi-centre collaboration; (ii) integration of multiparametric features (ADC map, DCE) to further strengthen discriminative power; (iii) application of explainability tools such as SHAP [13] and Grad-CAM [14] to improve clinical interpretability; and (iv) prospective external validation on data from independent institutions.

6.8 Conclusions

This chapter presents the first integrated real-world validation of the YOLOv8 → E-Net → Radiomics framework on an independent cohort of 55 patients with multiparametric prostate MRI. The detection module demonstrated robust generalisation (mAP@0.5 = 0.917). The segmentation module achieved adequate performance (DSC = 0.587 ± 0.123), with segmentation quality identified as the primary bottleneck. The radiomics classifier, incorporating 71 features extracted from NIfTI images, SMOTE oversampling, and optimal threshold selection, achieved a statistically significant AUROC of 0.723 ($p = 0.005$) with sensitivity of 63.6% and specificity of 86.4%, outperforming the original article's implementation (AUROC 0.685). These results provide a transparent and methodologically rigorous characterisation of the framework's real-world behaviour, and support the potential clinical utility of whole-gland prostate radiomics as a non-invasive complement to PSA and PI-RADS in the pre-biopsy workup of patients with suspected prostate cancer.

7. Conclusions

This doctoral thesis has developed, validated, and integrated an automated framework based on Artificial Intelligence and Radiomics for the analysis of multiparametric prostate MRI images. The framework is articulated in three sequential and complementary phases, each of which has produced scientifically relevant and clinically applicable results.

The YOLOv8 model demonstrated excellent capability for automatic detection of the prostate gland on T2W images, with mAP@0.5 of 0.982 and inference time of 0.05 s per image, confirming the clinical feasibility of the real-time system. Automatic segmentation with E-Net outperformed U-Net in both accuracy (DSC 81.6% vs 77.5%) and computational efficiency (362K vs 5.4M parameters; 1.88 s vs 4.66 s GPU inference), making it the ideal candidate for integration into clinical workflows. Radiomics analysis identified the feature original_glrlm_ShortRunLowGrayLevelEmphasis as the most discriminant parameter between normal and neoplastic prostate tissue on T2W sequences, with an AUROC of 0.685 and sensitivity of 76.25%.

The real-world validation performed on an independent cohort of 55 patients, acquired with a different scanner (3T Philips Ingenia) and during a distinct time period, confirmed the generalisability of the integrated framework. YOLOv8 maintained robust detection performance (mAP@0.5 = 0.917). E-Net achieved adequate segmentation (DSC = 0.587 ± 0.123), with training data volume identified as the primary determinant of performance. The methodologically corrected radiomics pipeline achieved an AUROC of 0.723 ($p = 0.005$), outperforming the original article's implementation (AUROC 0.685), with substantially improved specificity (86.4%) that could meaningfully reduce unnecessary biopsies.

Taken together, these results demonstrate that an integrated AI approach is technically feasible and clinically promising for supporting the diagnosis of prostate cancer. The proposed framework has the potential to standardise and accelerate the diagnostic workflow, reduce inter-observer variability, and provide quantitative imaging biomarkers for risk stratification and personalised therapeutic planning.

The work undertaken lays the groundwork for future research oriented towards multicentre validation, multimodal integration, and the development of explainable and clinically reliable AI systems, contributing to the paradigm of precision medicine in prostate cancer.

References

[1] M.S. Litwin, H.-J. Tan, The Diagnosis and Treatment of Prostate Cancer: A Review, JAMA 317 (2017) 2532. https://doi.org/10.1001/jama.2017.7248.

[2] 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, 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 76 (2019) 340–351. https://doi.org/10.1016/j.eururo.2019.02.033.

[3] G. Cutaia, G. La Tona, A. Comelli, F. Vernuccio, F. Agnello, C. Gagliardo, L. Salvaggio, N. Quartuccio, L. Sturiale, A. Stefano, M. Calamia, G. Arnone, M. Midiri, G. Salvaggio, Radiomics and Prostate MRI: Current Role and Future Applications, J. Imaging 7 (2021) 34. https://doi.org/10.3390/jimaging7020034.

[4] M. Hussain, YOLOv1 to v8: Unveiling Each Variant–A Comprehensive Review of YOLO, IEEE Access 12 (2024) 42816–42833. https://doi.org/10.1109/ACCESS.2024.3378568.

[5] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, in: N. Navab, J. Hornegger, W.M. Wells, A.F. Frangi (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Springer International Publishing, Cham, 2015: pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.

[6] A. Paszke, A. Chaurasia, S. Kim, E. Culurciello, ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation, (2016). https://doi.org/10.48550/ARXIV.1606.02147.

[7] A. Rasool, M. Ali, L. Salvaggio, V. Benfante, A. Comelli, Automated Prostate Gland Detection in MRI Using YOLOv8: A Deep Learning Approach for Early Diagnosis, in: E. Rodolà, F. Galasso, I. Masi (Eds.), Image Analysis and Processing - ICIAP 2025 Workshops, Springer Nature Switzerland, Cham, 2026: pp. 243–254. https://doi.org/10.1007/978-3-032-11317-7_21.

[8] G. Centineo, L. Salvaggio, S.I. Hussain, M. Ali, K. Meiburger, F. Molinari, A. Comelli, Prostate MRI Segmentation: A Comparative Analysis of U-Net and E-Net Deep Learning Models, in: E. Rodolà, F. Galasso, I. Masi (Eds.), Image Analysis and Processing - ICIAP 2025 Workshops, Springer Nature Switzerland, Cham, 2026: pp. 299–309. https://doi.org/10.1007/978-3-032-11317-7_26.

[9] S.S.M. Salehi, D. Erdogmus, A. Gholipour, Tversky Loss Function for Image Segmentation Using 3D Fully Convolutional Deep Networks, in: Q. Wang, Y. Shi, H.-I. Suk, K. Suzuki (Eds.), Machine Learning in Medical Imaging, Springer International Publishing, Cham, 2017: pp. 379–387. https://doi.org/10.1007/978-3-319-67389-9_44.

[10] M. Ali, V. Benfante, G. Cutaia, L. Salvaggio, S. Rubino, M. Portoghese, M. Ferraro, R. Corso, G. Piraino, T. Ingrassia, G. Tulone, N. Pavan, D. Di Raimondo, A. Tuttolomondo, A. Simonato, G. Salvaggio, Prostate Cancer Detection: Performance of Radiomics Analysis in Multiparametric MRI, in: G.L. Foresti, A. Fusiello, E. Hancock (Eds.), Image Analysis and Processing - ICIAP 2023 Workshops, Springer Nature Switzerland, Cham, 2024: pp. 83–92. https://doi.org/10.1007/978-3-031-51026-7_8.

[11] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: Synthetic Minority Over-sampling Technique, Jair 16 (2002) 321–357. https://doi.org/10.1613/jair.953.

[12] J. Bleker, D. Yakar, B. Van Noort, D. Rouw, I.J. De Jong, R.A.J.O. Dierckx, T.C. Kwee, H. Huisman, Single-center versus multi-center biparametric MRI radiomics approach for clinically significant peripheral zone prostate cancer, Insights Imaging 12 (2021) 150. https://doi.org/10.1186/s13244-021-01099-y.

[13] S. Lundberg, S.-I. Lee, A Unified Approach to Interpreting Model Predictions, (2017). https://doi.org/10.48550/arXiv.1705.07874.

[14] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization, in: 2017 IEEE International Conference on Computer Vision (ICCV), IEEE, Venice, 2017: pp. 618–626. https://doi.org/10.1109/ICCV.2017.74.