Radiomics-Based Early Triage of Prostate Cancer: A Multicenter Study from the CHAIMELEON Project

Aikaterini Vraka[1,*], Manuel Marfil-Trujillo[1], Gloria Ribas-Despuig[1], Silvia Flor-Arnal[1], Leonor Cerdá-Alberich[1], Paula Jiménez-Gómez[1], Ana Jimenez-Pastor[2], Luís Martí-Bonmatí[1,3]

1 Biomedical Imaging Research Group-Grupo de Investigación Biomédica en Imagen (GIBI230), La Fe Health Research Institute, Valencia, Spain
2 Quibim SL, Valencia, Spain
3 Medical Imaging Department, La Fe University and Polytechnic Hospital, Valencia, Spain

* *aikaterini_vraka@iislafe.es*

Abstract

Prostate cancer (PCa) is the most commonly diagnosed malignancy in men worldwide. Accurate triage of patients based on tumor aggressiveness and staging is critical for selecting appropriate management pathways. While magnetic resonance imaging (MRI) has become a mainstay in PCa diagnosis, most predictive models rely on multiparametric imaging or invasive inputs, limiting generalizability in real-world clinical settings. This study aimed to develop and validate machine learning (ML) models using radiomic features extracted from T2-weighted MRI—alone and in combination with clinical variables—to predict ISUP grade (tumor aggressiveness), lymph node involvement (cN) and distant metastasis (cM). A retrospective multicenter cohort from three European sites in the Chaimeleon project was analyzed. Radiomic features were extracted from prostate zone segmentations and lesion masks, following standardized preprocessing and ComBat harmonization. Feature selection and model optimization were performed using nested cross-validation and Bayesian tuning. Hybrid models were trained using XGBoost and interpreted with SHAP values. The ISUP model achieved an AUC of 0.66, while the cN and cM models reached AUCs of 0.77 and 0.80, respectively. The best-performing models consistently combined prostate zone radiomics with clinical features such as PSA, PIRADSv2 and ISUP grade. SHAP analysis confirmed the importance of both clinical and texture-based radiomic features, with entropy and non-uniformity measures playing central roles in all tasks. Our results demonstrate the feasibility of using T2-weighted MRI and zonal radiomics for robust prediction of aggressiveness, nodal involvement and distant metastasis in PCa. This fully automated pipeline offers an interpretable, accessible and clinically translatable tool for first-line PCa triage, with potential integration into real-world diagnostic workflows.

Introduction

Prostate cancer (PCa) remains the most frequently diagnosed malignancy and the second leading cause of cancer-related death among men around the world [1]. Accurate risk stratification and staging are crucial for guiding treatment decisions, ranging from

active surveillance to radical interventions. Traditional clinical assessments, including prostate specific antigen (PSA) levels, digital rectal examinations and biopsies are often complemented by imaging modalities to enhance diagnosis accuracy [1–4].

Magnetic Resonance Imaging (MRI) has become a cornerstone in the non-invasive evaluation of PCa, with many studies showing clear benefits of biparametric (bpMRI) instead of multiparametric (mpMRI) MRI [5–8]. T2-weighted imaging (T2w) offers superior soft tissue contrast, facilitating detailed visualization of prostate anatomy and lesion localization. While diffusion-weighted imaging (DWI) and dynamic contrast-enhanced (DCE) imaging provide additional functional information, T2w sequences are widely available and consistently utilized across clinical settings [6].

Radiomics, the extraction of high-dimensional quantitative features from medical images, combined with machine learning (ML) techniques, has shown promise in capturing tumor heterogeneity and improving predictive modeling in oncology [9]. In PCa, radiomics have been applied to predict tumor aggressiveness, lymph node involvement and biochemical recurrence [10–15].

Building on these imaging capabilities, recent advances in artificial intelligence have enabled the development of predictive models that integrate radiomic features with clinical data to improve staging accuracy. It has been demonstrated that radiomics-based ML models using MRI show comparable accuracy to standard nomograms for lymph node invasion (LNI), while the use of computed tomography (CT) and positron emission tomography-CT (PET/CT) models can show high diagnostic and prognostic results [16, 17]. Other studies demonstrated that ML models integrating radiomics from T2w MRI with clinical and pathological data could predict biochemical recurrence with an area under the curve (AUC) of at least 0.84 [18, 19].

However, many existing studies have limitations, including reliance on single-center, retrospective datasets with enriched populations that may not reflect real-world disease prevalence [16, 20]. Additionally, while CT and PET/CT models are commonly used for staging, their sensitivity and specificity for detecting LNI are suboptimal, particularly in early-stage disease [21, 22]. As for distant metastasis, PET/CT still remains the reference in its detection, while recent studies report high performance using PET/MRI [23, 24]. Nevertheless, this technology shows significant limitations with respect to MRI, such as exposure to radiation, limited availability and high costs [25, 26].

In this study, we present a multi-institutional analysis involving real-world data from three European clinical sites of the Chaimeleon project. We developed and evaluated three ML models trained on radiomic features extracted from T2w MRI, guided by prostate zonal segmentation masks and combined with clinical variables. Besides the exclusive use of T2w, which are widely available and the baseline imaging modality for PCa diagnosis, we propose a single pre-processing and analysis pipeline yielding the assessment of three key biomarkers for a first-level PCa triage: aggressiveness, LNI and distant metastasis. In this way, our approach aims to provide generalizable and clinically translatable tools for the stratification and staging of PCa risk.

Materials and Methods

Materials

Retrospective data from three European clinical sites from the Chaimeleon project were used [27]. These sites were Valencia (Spain), Porto (Portugal) and Angers (France). Due to the retrospective and de-identified use of the data, informed consent was waived. Originally, data from 534 patients with newly diagnosed PCa were recruited.

For each patient, an imaging study consisting of bpMRI of the prostate or pelvic area was included, accompanied by clinical information derived from the electronic case

report form (eCRF), including the age at diagnosis, total PSA, PIRADSv2 and the three variables that would be used as ground truth during the training of our three AI models: ISUP (after histopathological confirmation), LNI-clinically confirmed (cN) and distant metastasis after clinical confirmation (cM). Information on database size and prevalence as well as clinical information for each model can be seen in Table 1. Age appears in mean and standard deviation (sd), while PSA appears in median and interquartile range (IQR) after the application of Shapiro-Wilk test on both variables to check for normality [28]. The distribution per clinical center can be observed in Figure 1, where due to the lack of ground truth in some cases, the cN and cM models database appears reduced and the only clinical site is Valencia.

Table 1. Database size (N), prevalence and summary statistics for age and PSA per model. PSA is shown in median and IQR.

Model	N	Prevalence [%]	Age	PSA [ng/mL]
ISUP	534	36	67.8 ± 7.6	7.8 (5.8–11.4)
cN	350	10	68.6 ± 7.6	7.5 (5.7–11.4)
cM	256	7	68.9 ± 8.1	8.1 (5.8–13.4)

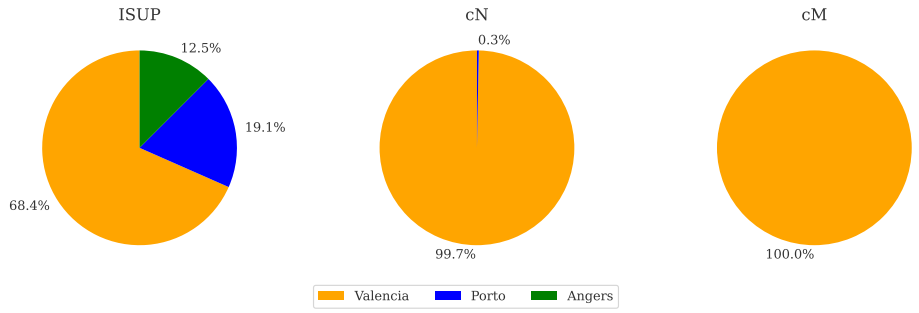

Figure 1. Center distribution per model. For the cN model, only one patient came from Angers, while the remaining patients came from Valencia. For the cM model, all data came from Valencia.

Methods

Axial T2w images were selected as the only imaging input. The pre-processing and main analysis pipeline was identical for all three developments (see Figure 2), with the only difference being found in the per case selected database (rows filtering) as well as the model training process. Development was conducted in python language using Python 3.11 version.

The first step of image pre-processing was the application of a N4 bias field correction algorithm for intensity normalization, followed by an anisotropic diffusion filter (ADF) to reduce noise while preserving prostate gland boundaries [29–31]. The last step of image pre-processing was the z-score normalization to account for differences in acquisition protocols [32]. An example of the image pre-processing result can be observed in Figure 3.

Segmentation masks of PCa lesions and prostate zones were acquired through QP-Prostate (Quibim, SL), as is shown in Figure 4. Radiomics were extracted using PyRadiomics with a variety of image filters. Clinical features were added to the radiomics signatures. For the prostate aggressiveness model, age, PSA and PIRADSv2 were added. For the cN and cM model, besides the above, ISUP information was

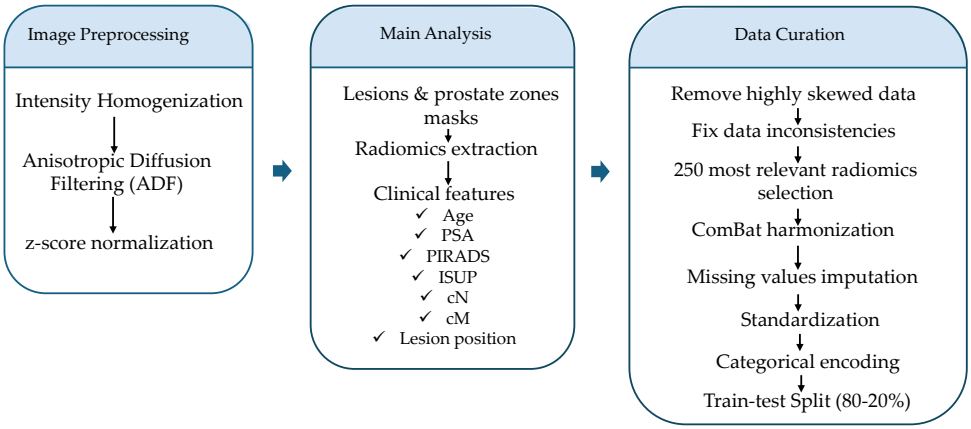

Figure 2. Image pre-processing, main analysis and data curation pipeline. The procedure was conducted once before the development of the three predictive models.

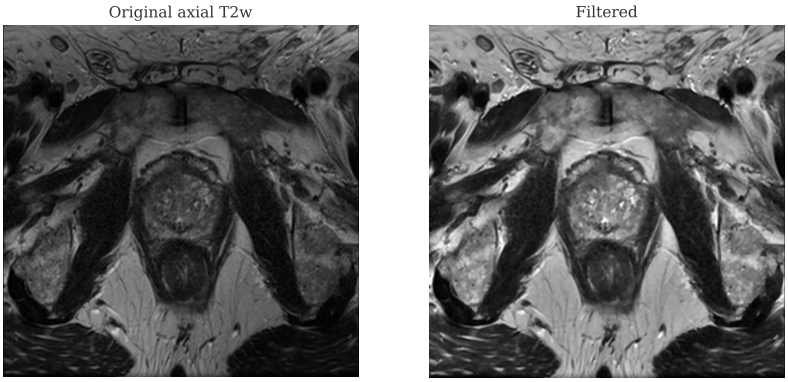

Figure 3. One slice of original (left) and pre-processed image (right) of a patient. Pre-processing aimed to smooth out signal intensity, preserve prostate boundaries and account for differences in acquisition protocols that were present due to the use of a multi-center database.

recruited. For all three models, lesion position was extracted as a binary value for peripheral (PZ), central (CZ), transitional (TZ) and seminal vesicles (SV) using the lesions masks and the information was added to the model input.

Clinical and radiomic features underwent a multi-step data curation process. Highly skewed features (skewness > 5.0) were removed and inconsistent or missing values were imputed using either the most frequent value (for categorical/boolean features) or K-nearest neighbors ($K = 3$) for numerical data. Boolean variables were binarized and filtered using the chi-squared test. Categorical batch-related variables (here: scanner vendor, magnetic field strength, center) were encoded and included in ComBat harmonization [33, 34]. All numerical features were standardized using z-score normalization.

Coarse feature selection was performed by applying a minimum variance threshold, removing highly correlated features (correlation > 0.8) and selecting the top 250 most

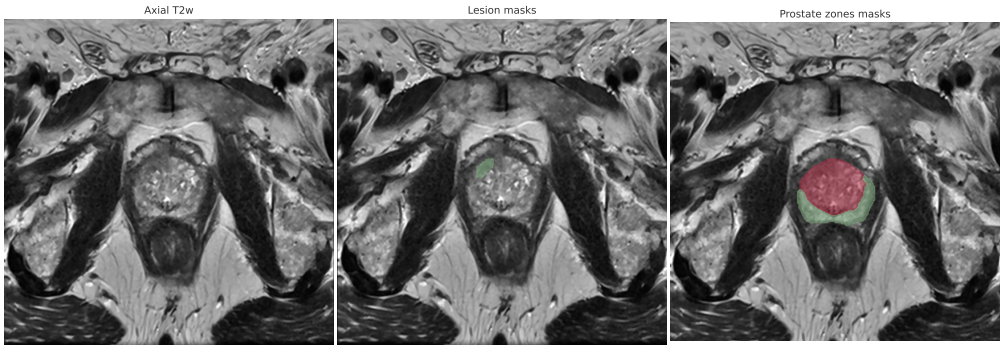

Figure 4. Example of a lesion and prostate zone mask for the same patient using QP-Prostate. Peripheral area is shown in green, while central and transitional area in red.

relevant features using Minimum Redundancy Maximum Relevance (mRmR) [35]. The curated dataset was then split into training and test sets (80/20) using stratified sampling by target class.

Model development was carried out, for each of the three cases, using a nested cross-validation framework with five outer folds and three inner folds to ensure unbiased evaluation. Within each inner loop, mRmR feature selection was applied to identify the most relevant features, with the number of selected features ranging from 3 to 30. Performing feature selection inside the inner loop prevents data leakage and ensures that the outer test set remains entirely independent, preserving the validity of performance estimates. To address class imbalance, the training data in each outer fold was balanced using a Synthetic Minority Oversampling Technique (SMOTE) and a range of *scale_pos_weight* values was also explored during hyperparameter tuning. Separate models were trained on radiomics-only data, clinical-only data and hybrid datasets combining both. Additionally, the contribution of lesion radiomics, prostate zones radiomics and their combination was investigated.

Model optimization was performed using Optuna, a Bayesian hyperparameter search framework, with the objective of maximizing AUC on the inner folds. The best-performing feature subset and hyperparameters were then used to train an XGBoost classifier on the full training split of each outer fold. Performance was evaluated on the held-out test split (20%) using AUC, accuracy, precision, recall, F1-score and Cohen's kappa. To enhance model interpretability, SHapley Additive exPlanations (SHAP) values were computed on the outer test predictions. SHAP values provide a unified framework for explaining individual predictions by estimating the contribution of each feature to the model's output, enabling case-level and global interpretability of the radiomic and clinical features involved [36].

Results

For PCa aggressiveness, a binary predictive model was trained as follows: ISUP grades ≤ 2 were considered low agrressive, while ISUP grades > 2 were considered highly aggressive. For cN and cM, the absence of LNI and metastasis, respectively, was considered as the positive class, while their presence was considered as the negative class.

Final models specifications are shown in Table 2. For all classification tasks, hybrid models with radiomics from prostate zones showed the highest performance [37]. For the prostate aggressiveness classification, six features were selected, three radiomics and

three clinical. For the cN model, 11 features in total were selected, four of which were clinical and the remaining radiomics. For the cM model, 13 features were selected, with three being clinical and the remaining 10 radiomic signatures.

Table 2. Optimal configuration of the best-performing model per classification task.

Model	Type	Radiomics	Feature number	N estimators	Max depth	Learning rate
ISUP	hybrid	prostate zones	6	58	3	0.040
cN	hybrid	prostate zones	11	74	9	0.025
cM	hybrid	prostate zones	13	233	5	0.170

Table 3. Selected features per classification task. Radiomic feature types are abbreviated as: GLCM: Gray-Level Co-occurrence Matrix; GLSZM: Gray-Level Size Zone Matrix; GLDM: Gray-Level Dependence Matrix; GLRLM: Gray-Level Run Length Matrix.

PCa aggressiveness	cN	cM
First-order: Median	GLDM: Non-uniformity	GLCM: Correlation (LHH)
GLSZM: Zone Entropy	First-order: 90th Percentile	GLDM: Dependence Entropy
First-order: Skewness	GLSZM: Non-uniformity	First-order: Median
PZ (Peripheral Zone)	GLRLM: Short Run Emphasis	GLDM: Variance
TZ (Transition Zone)	GLSZM: Emphasis	First-order: Skewness
PIRADSv2	GLSZM: Zone Entropy	GLRLM: Entropy
—	GLCM: Inverse Difference	First-order: Minimum
—	PIRADSv2	GLCM: Difference Entropy
—	ISUP	GLSZM: Zone Entropy
—	PSA	GLCM: Correlation (LLL)
—	PZ (Peripheral Zone)	ISUP
—	—	PIRADSv2
—	—	TZ (Transition Zone)

The selected features per task are shown in Table 3. It can be observed that texture-based radiomics, such as entropy and non-uniformity as well as first-order statistics play a prominent role in all three classification tasks. PIRADSv2 and zonal location (PZ, TZ) appear consistently regardless of the task, confirming the importance of anatomical and expert-derived clinical variables. The distant metastasis (cM) model was notably more based on high-order texture characteristics, possibly reflecting increased heterogeneity in advanced disease.

Figure 5 shows the receiver-operating characteristics (ROC) curves at the holdout test set for each classification task. The performance metrics are then summarized in Table 4. The model trained to predict prostate cancer aggressiveness (ISUP) achieved an AUC of 0.66, a balanced accuracy of 0.64 and a sensitivity of 0.61 on the test set. The ROC curve reflects modest separation between aggressive (ISUP > 2) and non-aggressive cases. While the overall performance is moderate, the model was able to identify a substantial proportion of clinically significant tumors despite the presence of high class imbalance, possibly due to the oversampling and scaling operations during training.

The cN model achieved an AUC of 0.77, a balanced accuracy of 0.76 and a sensitivity of 0.74. Despite the low prevalence of positive LNI (10%), the model maintained strong sensitivity, which is particularly relevant from a clinical standpoint. Early and accurate detection of LNI is crucial for treatment planning and the model's

ability to identify true positives without overfitting to the dominant class reflects its robustness. The best overall performance was observed in the cM model, with an AUC of 0.80, balanced accuracy of 0.79 and sensitivity of 0.78. These values indicate strong discriminative power despite a highly imbalanced dataset (7% cases with distant metastasis). The high sensitivity is of particular clinical interest, as early detection of distant metastasis significantly influences prognosis and therapeutic decision-making. The balanced accuracy also suggests the model retained specificity while effectively detecting metastatic cases.

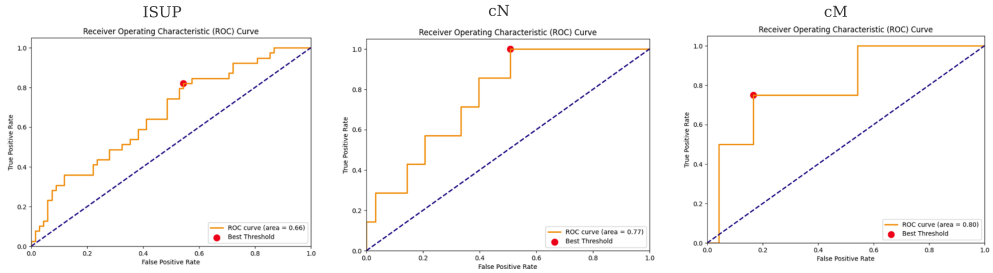

Figure 5. ROC curve for the test set of each predictive task.

Table 4. Performance matrix for the test set of each predictive task.

	AUC	Accuracy (balanced)	F1	Precision	Sensitivity	Specificity	Recall
ISUP	0.664	0.626	0.597	0.597	0.487	0.706	0.596
cN	0.769	0.675	0.565	0.566	0.714	0.635	0.675
cM	0.802	0.698	0.644	0.621	0.500	0.896	0.698

The confusion matrices with the true positive and true negative outcomes on the test set of each classification task can be seen in Figure 6. The ISUP model achieved a relatively balanced performance between the "Low" and "High" aggressiveness groups, though there was some degree of misclassification between the two classes. For the cN and cM models, the matrices highlighted a tendency toward higher sensitivity for the cN0 and cM0 classes, with fewer false negatives. However, specificity was lower, with some overprediction of nodal and metastatic involvement.

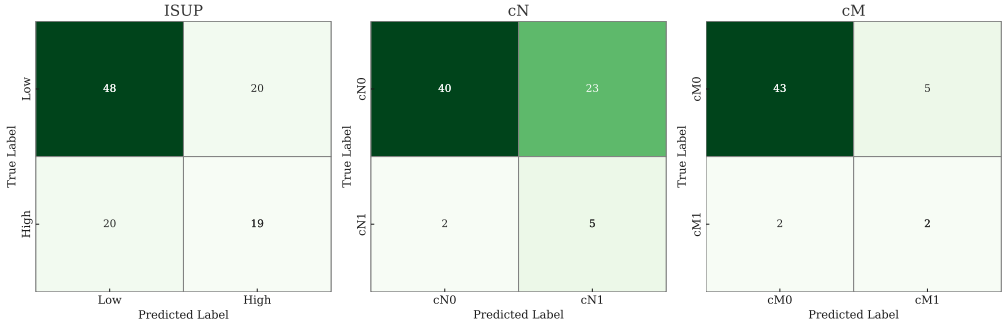

Figure 6. Confusion matrices for the classification of ISUP grade group (Low vs High aggressiveness), clinical LNI (cN0-no invasion vs cN1-invasion) and clinical metastatic status (cM0-no metastasis vs cM1-metastasis). True positives (TP) and true negatives (TN) are highlighted in green. The matrices show the number of correctly and incorrectly classified cases.

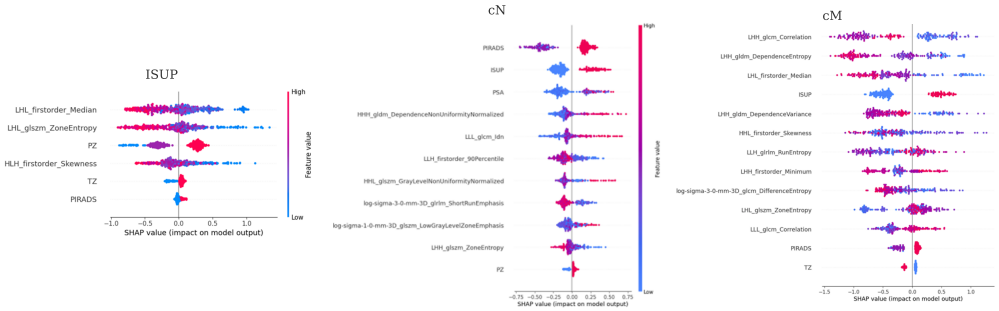

Figure 7. SHAP summary plots illustrating the impact of individual features on the model output for each classification task. Each point represents a patient case, colored by the value of the corresponding feature (red = high, blue = low). Features are ordered by overall importance, with SHAP values indicating the direction and magnitude of each feature's contribution to the model's decision.

SHAP summary plots (Figure 7) were used to visualize the contribution of individual features to model predictions for prostate aggressiveness, cN and cM status. Across all models, both clinical variables and radiomic features from various wavelet decompositions demonstrated relevance, though the patterns of importance varied by outcome.

For ISUP grade classification, the most influential features included $LHL_firstorder_Median$, $LHL_glszm_ZoneEntropy$ and $HLH_firstorder_Skewness$, suggesting that intensity and texture heterogeneity within tumor regions are associated with aggressiveness. Notably, the PZ and TZ location indicators also contributed, indicating zonal variation in predictive characteristics.

In the cN model, PIRADSv2, ISUP and PSA were among the top contributors, reinforcing their central clinical relevance. In addition, texture-based radiomic features showed consistent importance, indicating that more subtle structural tissue changes may reflect nodal involvement.

For cM status, radiomic features dominated the top rankings. Features such as $LHH_glcm_Correlation$, $LHH_gldm_DependenceEntropy$ and $LHL_firstorder_Median$ were strong contributors, while ISUP and PIRADSv2 had lower influence. This suggests that the model may be capturing latent imaging signatures of metastatic potential not directly represented in standard clinical scores.

Across all tasks, SHAP values highlighted not only the magnitude of feature impact but also the directionality: for example, higher values of certain entropy-related features were generally associated with increased predicted risk (positive SHAP values). The color gradients (blue to red) further revealed interactions between feature magnitude and effect, offering deeper insights into feature behavior across patient subgroups.

These results reinforce the utility of combining radiomics with clinical parameters to enhance model explainability and clinical insight, potentially guiding personalized risk stratification strategies.

Discussion

In this study, we present a robust, interpretable and unified ML framework for the first-level triage of prostate cancer aggressiveness, LNI and distant metastasis using only T2w MRI and routinely available clinical data. Our results demonstrate competitive performance when compared to more complex state-of-the-art (SOTA) models that typically require multiparametric imaging, biopsy-derived data or high-cost PET/MRI.

Several recent studies have reported strong predictive performance using mpMRI or PET/MRI combined with radiomic ML models. For example, Bourbonne et al. developed a multimodal MRI-based model to predict LNI, achieving a C-index of 0.89 by integrating radiomics with clinical variables [12]. Similarly, Ayyıldız et al. trained ML models on biparametric MRI for distinguishing clinically significant PCa, reaching AUCs up to 0.86 [38]. While these values exceed our results, they rely on broader imaging modalities or curated datasets with enriched populations. Interestingly, when T2w-only images are used, an AUC of 0.64 is obtained [38].

Our pipeline, by contrast, utilizes T2w MRI alone, which is widely available and routinely acquired in prostate cancer diagnosis. This design offers significant advantages in terms of clinical scalability, lower cost and ease of integration into existing workflows, especially in settings where diffusion or contrast-enhanced sequences are unavailable. Our findings are further supported by recent work from Yi et al., who demonstrated that T2w-only meta-learning models can achieve comparable accuracy to mpMRI-based methods [39].

When comparing task-specific performance, our results for cN and cM (AUCs: 0.77 and 0.80) are particularly noteworthy given the real-world, imbalanced, multi-institutional data used for training. Recently, a study achieved an AUC of 0.90 for the prediction of distant metastasis using a semi-automated deep learning framework which integrated bpMRI, radiomics and clinical data [40]. Nevertheless, the reported prevalence ($30 - 43\%$) was substantially higher than real-world rates, potentially inflating predictive performance and limiting generalizability [41]. Other studies using advanced PET/MRI pipelines report AUCs in the range of 0.90–0.94 for predicting biochemical recurrence or overall risk [42] but at the cost of accessibility and patient exposure to ionizing radiation. Our model maintains strong performance without such trade-offs.

In terms of ISUP classification, although the AUC of 0.66 is modest, it is consistent with prior findings on PIRADSv2 category 3 lesions, where radiomic models achieve AUCs around 0.64–0.68 using T2w and ADC data [43]. These results highlight the intrinsic difficulty of aggressiveness classification using imaging alone but also reinforce the value of hybrid clinical-radiomic approaches.

A key strength of our approach is its methodological rigor, including the use of nested cross-validation (CV) for model development and evaluation. This strategy avoids data leakage by ensuring feature selection and hyperparameter tuning are performed within inner CV loops, while outer folds remain strictly independent for unbiased performance assessment. Such design choices increase the reliability of reported metrics and enhance model generalizability across clinical settings.

While our AUCs for ISUP (0.66), cN (0.77) and cM (0.80) may appear modest compared to some mpMRI or PET-based models, it is important to emphasize that our approach relies solely on T2w MRI and routinely available clinical data. This design prioritizes generalizability and real-world feasibility and the achieved sensitivity and balanced accuracy demonstrate clinical utility for first-line staging.

Finally, our use of SHAP-based explainability shows that radiomic features such as entropy, non-uniformity and median intensity, combined with anatomical zone indicators and PIRADSv2 scores, consistently drive predictions across all tasks. This transparency supports clinical adoption and patient-specific risk discussions.

1 Conclusions

This work introduces a fully automated, interpretable and resource-conscious pipeline for first-level triage in prostate cancer based on T2w MRI, radiomic features and clinical variables. The pipeline addressed three clinically relevant classification tasks: tumor

aggressiveness (ISUP), LNI (cN) and distant metastasis (cM). Despite relying solely on T2w images, the models demonstrated competitive performance, particularly in the detection of metastatic disease, while maintaining high sensitivity — a key requirement for clinical screening and staging. The integration of prostate zone–specific radiomics, ComBat harmonization across centers and SHAP-based explainability contributed to both model robustness and interpretability.

Acknowledgments

CHAIMELEON has been funded by Horizon 2020 project (RIA, topic DT-TDS-05-2020-AI for Health Imaging; call SC1-FA-DTS-2019-1, under Grant Agreement No. 952172).

References

1. R. L. Siegel, T. B. Kratzer, A. N. Giaquinto, H. Sung, and A. Jemal, "Cancer statistics, 2025," *CA: A Cancer Journal for Clinicians*, vol. 75, no. 1, pp. 10–45, 2025.

2. J. Morote, Borque-Fernando, L. M. Esteban, A. Celma, M. Campistol, *et al.*, "Investigating efficient risk-stratified pathways for the early detection of clinically significant prostate cancer," *Journal of Personalized Medicine*, vol. 14, no. 2, 2024.

3. J. Zhu, W. Xiaohui, X. Yuting, Z. Qingshui, X. Xueyi, H. Zhiyang, and C. Shaohao, "Prospective analysis of the diagnostic accuracy of digital rectal examination and magnetic resonance imaging for t staging of prostate cancer," *Journal of Cancer Research and Therapeutics*, vol. 19, no. 4, pp. 1024–1030, 2023.

4. Y. Hou, K. Jiang, L. Wang, R. Zhi, M. Bao, *et al.*, "Biopsy-free ai-aided precision mri assessment in prediction of prostate cancer biochemical recurrence," *British Journal of Cancer*, vol. 129, no. 10, pp. 1625–1633, 2023.

5. N. Huebner, S. Korn, I. Resch, Grubmüller, T. Gross, R. Gale, *et al.*, "Visibility of significant prostate cancer on multiparametric magnetic resonance imaging (mri)—do we still need contrast media?," *European Radiology*, vol. 31, pp. 3754–3764, 2020.

6. 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*, vol. 53, no. 1, pp. 283–291, 2021.

7. C. Charlotte, S. Montagne, S. Bourrelier, M. Roupret, E. Barret, F. Rozet, E. Comperat, *et al.*, "Prostate cancer local staging using biparametric mri: assessment and comparison with multiparametric mri," *European Journal of Radiology*, vol. 132, p. 109350, 2020.

8. P. Palumbo, R. Manetta, A. Izzo, F. Bruno, F. Arrigoni, M. De Filippo, *et al.*, "Biparametric (bp) and multiparametric (mp) magnetic resonance imaging (mri) approach to prostate cancer disease: a narrative review of current debate on dynamic contrast enhancement," *Gland Surgery*, vol. 9, no. 6, 2020.

9. P. Lambin, E. Rios-Velazquez, R. Leijenaar, *et al.*, "Radiomics: Extracting more information from medical images using advanced feature analysis," *European Journal of Cancer*, vol. 48, no. 4, pp. 441–446, 2020.

10. W. I. Luining, D. E. Oprea-Lager, A. N. Vis, R. J. A. van Moorselaar, *et al.*, "Optimization and validation of 18f-dcfpyl pet radiomics-based machine learning models in intermediate- to high-risk primary prostate cancer," *PLOS ONE*, vol. 18, pp. 1–15, 11 2023.

11. X. Shu, Y. Liu, X. Qiao, G. Ai, L. Liu, *et al.*, "Radiomic-based machine learning model for the accurate prediction of prostate cancer risk stratification," *British Journal of Radiology*, vol. 96, p. 20220238, 01 2023.

12. V. Bourbonne, V. Jaouen, T. A. Nguyen, V. Tissot, L. Doucet, *et al.*, "Development of a radiomic-based model predicting lymph node involvement in prostate cancer patients," *Cancers*, vol. 13, no. 22, 2021.

13. M. C. F. Cysouw, B. H. E. Jansen, T. van de Brug, D. E. Oprea-Lager, E. Pfaehler, *et al.*, "Machine learning-based analysis of [18f]dcfpyl pet radiomics for risk stratification in primary prostate cancer," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 48, no. 2, pp. 340–349, 2021.

14. Y.-D. Wang, C.-P. Huang, Y.-R. Yang, H.-C. Wu, Y.-J. Hsu, *et al.*, "Machine learning and radiomics of bone scintigraphy: Their role in predicting recurrence of localized or locally advanced prostate cancer," *Diagnostics*, vol. 13, no. 21, 2023.

15. X. Qiao, X. Gu, Y. Liu, X. Shu, G. Ai, S. Qian, *et al.*, "Mri radiomics-based machine learning models for ki67 expression and gleason grade group prediction in prostate cancer," *Cancers*, vol. 15, no. 18, 2023.

16. E. Faiella, F. Vaccarino, R. Ragone, G. D'Amone, V. Cirimele, *et al.*, "Can machine learning models detect and predict lymph node involvement in prostate cancer? a comprehensive systematic review," *Journal of Clinical Medicine*, vol. 12, no. 22, 2023.

17. D. Santucci, R. Ragone, E. Vergantino, F. Vaccarino, F. Esperto, *et al.*, "Comparison between three radiomics models and clinical nomograms for prediction of lymph node involvement in pca patients combining clinical and radiomic features," *Cancers*, vol. 16, no. 15, 2024.

18. S. R. Duenweg, S. A. Bobholz, M. J. Barrett, A. K. Lowman, A. Winiarz, *et al.*, "T2-weighted mri radiomic features predict prostate cancer presence and eventual biochemical recurrence," *Cancers*, vol. 15, no. 18, 2023.

19. N. P. Nanekaran, T. H. Felefly, N. Schieda, S. C. Morgan, R. Mittal, and E. Ukwatta, "Prediction of prostate cancer recurrence after radiotherapy using a fused machine learning approach: utilizing radiomics from pretreatment t2w mri images with clinical and pathological information," *Biomedical Physics Engineering Express*, vol. 10, p. 065035, oct 2024.

20. H. Ueki, T. Terakawa, T. Hara, M. Uemura, Y. Okamura, *et al.*, "Utility of machine learning models to predict lymph node metastasis of japanese localized prostate cancer," *Cancers*, vol. 16, no. 23, 2024.

21. J. Zhou, R. Wu, W. Wang, Y. Zhao, and L. X., "^{68}ga-psma pet/ct for the evaluation of metastasis in patients with prostate cancer: A systematic review and meta-analysis," *Hell J Nucl Med*, vol. 25, no. 3, pp. 297–311, 2022.

22. A. Heidenreich, M. Schmautz, D. A. Pfister, and F. Hartmann, "68 ga-psma pet/ct for preoperative lymph nosde staging of high-risk prostate cancer.," *Journal of Clinical Oncology*, vol. 40, no. 6_suppl, pp. 38–38, 2022. PMID:.

23. J. Van Damme, B. Tombal, L. Collette, S. Van Nieuwenhove, V. Pasoglou, *et al.*, "Comparison of 68ga-prostate specific membrane antigen (psma) positron emission tomography computed tomography (pet-ct) and whole-body magnetic resonance imaging (wb-mri) with diffusion sequences (dwi) in the staging of advanced prostate cancer," *Cancers*, vol. 13, no. 21, 2021.

24. S. A. Mirshahvalad, A. Basso Dias, C. Ortega, J. A. Abreu Gomez, S. Krishna, *et al.*, "[18f]f-dcfpyl pet/mri radiomics for intraprostatic prostate cancer detection and metastases prediction using whole-gland segmentation," *British Journal of Radiology*, p. tqaf014, 01 2025.

25. J. Nikitas, E. C. Lam, K. M. Booker, W. P. Fendler, M. Eiber, *et al.*, "Update from the psma-drt trial: A randomized phase iii trial of psma-pet prior to definitive radiation therapy for unfavorable intermediate-risk or high-risk prostate cancer.," *Journal of Clinical Oncology*, vol. 42, no. 4_suppl, pp. 291–291, 2024. PMID:.

26. I. Tsechelidis and A. Vrachimis, "Psma pet in imaging prostate cancer," *Frontiers in Oncology*, vol. Volume 12 - 2022, 2022.

27. L. M. Bonmatí, A. Miguel, A. Suárez, M. Aznar, J. P. Beregi, *et al.*, "Chaimeleon project: Creation of a pan-european repository of health imaging data for the development of ai-powered cancer management tools," *Frontiers in Oncology*, vol. Volume 12 - 2022, 2022.

28. S. Shapiro and M. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, pp. 591–611, 1965.

29. N. J. Tustison, B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich, and J. C. Gee, "N4itk: Improved n3 bias correction," *IEEE Transactions on Medical Imaging*, vol. 29, no. 6, pp. 1310–1320, 2010.

30. P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, 1990.

31. M. Fernandez-Paton, L. Cerda-Alberich, C. Sangüesa Nebot, B. Martinez de las Heras, D. Veiga Canuto, A. Cañete Nieto, and L. Marti-Bonmati, "Mr denoising increases radiomic biomarker precision and reproducibility in oncologic imaging," *Journal of Digital Imaging*, vol. 34, pp. 1134–1145, 2021.

32. P. Lambin, R. Leijenar, D. T., J. Peerlings, E. de Jong, J. van Timmeren, *et al.*, "Radiomics: the bridge between medical imaging and personalized medicine," *nature reviews clinical oncology*, vol. 14, no. 12, pp. 749–762, 2017.

33. W. E. Johnson, C. Li, and A. Rabinovic, "Adjusting batch effects in microarray expression data using empirical bayes methods," *Biostatistics*, vol. 8, pp. 118–127, 04 2006.

34. J.-P. Fortin, N. Cullen, Y. I. Sheline, W. D. Taylor, I. Aselcioglu, P. A. Cook, P. Adams, C. Cooper, M. Fava, P. J. McGrath, M. McInnis, M. L. Phillips, M. H. Trivedi, M. M. Weissman, and R. T. Shinohara, "Harmonization of cortical thickness measurements across scanners and sites," *NeuroImage*, vol. 167, pp. 104–120, 2018.

35. H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.

36. S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.

37. A. Rodrigues, J. Santinha, B. Galvão, C. Matos, *et al.*, "Prediction of prostate cancer disease aggressiveness using bi-parametric mri radiomics," *Cancers*, vol. 13, no. 23, 2021.

38. H. Ayyıldız, O. Ince, E. Korkut, *et al.*, "Machine learning models for discriminating clinically significant from clinically insignificant prostate cancer using bi-parametric magnetic resonance imaging," *Diagnostic and Interventional Radiology*, 2024.

39. W. Yi, Y. Wang, N. Thorley, *et al.*, "T2-only prostate cancer prediction by meta-learning from bi-parametric mr imaging," *arXiv preprint arXiv:2411.07416*, 2024.

40. S. Xinyang, S. Tianci, H. Xiangyu, Z. Shuang, W. Yangyang, D. Mengying, X. Tonghui, Z. Jingran, and Y. Feng, "A semi-automatic deep learning model based on biparametric mri scanning strategy to predict bone metastases in newly diagnosed prostate cancer patients," *Frontiers in Oncology*, vol. Volume 14 - 2024, 2024.

41. J. Cullen, Y. Chen, H.-C. Kuo, K. R. Rice, I. L. Rosner, and J. Forsberg, "Modeling time from bone metastasis to death in a racially diverse cohort of military health care beneficiaries.," *Journal of Clinical Oncology*, vol. 37, no. 7_suppl, pp. 246–246, 2019. PMID:.

42. L. Papp, C. Spielvogel, B. Grubmüller, *et al.*, "Supervised machine learning enables non-invasive lesion characterization in primary prostate cancer with [68ga]ga-psma-11 pet/mri," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 48, pp. 1795–1805, 2020.

43. C. Lim, J. Abreu-Gomez, R. Thornhill, *et al.*, "Utility of machine learning of apparent diffusion coefficient (adc) and t2-weighted (t2w) radiomic features in pi-rads version 2.1 category 3 lesions to predict prostate cancer diagnosis," *Abdominal Radiology*, pp. 1–12, 2021.