2026 JULY 19–22 VANCOUVER, BC
JOINT AAPM | COMP MEETING

A Method for Prostate Cancer Metastases Classification In ^{18}F-PSMA PET/CT Images Using a Machine Learning Stacking Ensemble Model

I. Stamouli[1], I. Gatos[1], P. Katsakiori[1], S. Tsantis[1], N. Papathanasiou[1], L. Zoglopitou[2], T. Kalathas[2], V. Chatzipavlidou[2], D.N. Mihailidis[3], A. Makridou[2], G.C. Kagadis[1]

[1]University of Patras, Rion, Greece; [2] 'Theagenio' Cancer Hospital of Thessaloniki, Thessaloniki, Greece; [3]University of Pennsylvania, Philadelphia, USA

ABSTRACT

Purpose: To develop a computer-aided diagnosis (CAD) system for the classification of benign versus malignant findings on 18F-PSMA PET/CT imaging. It was applied in patients with biochemical recurrent (BCr) prostate cancer (PCa), using post-therapy follow-up (FU) imaging information (tumor regression or lesion regression in size indicated malignancy) as the Gold Standard.

Methods: A dataset of 44 patients with BCr PCa who underwent 18F-PSMA PET/CT imaging was analyzed. Suspicious findings were categorized based on post-therapy imaging FU. Findings that did not respond to therapy and were interpreted as non-specific by the physician were labeled as benign, whereas findings that demonstrated treatment response were labeled as malignant. A total of 285 findings were included (Benign: 123, Malignant: 162), with data augmentation applied on the benign class to mitigate class imbalance. For each finding, the maximum intensity point was selected, and 1D intensity profiles along the x, y, and z axes were extracted in the axial plane, directly used as inputs to a multi-layer perceptron (MLP) classifier. These intensity profiles were also used to extract statistical and Gaussian-fit features to train a random forest (RF) classifier. The final model includes a stacking ensemble architecture combining MLP and RF base models with logistic regression serving as the meta-classifier. Forward feature selection was used, and the model was trained with 10-fold cross validation.

Results: Our model achieved a mean accuracy of 92.6±5.3%, sensitivity of 94.4±6.2%, specificity of 90.3±10.7%, and Area Under the Receiver Operating Characteristic Curve (AUROC) of 97.4±3.6%. Feature selection analysis showed that the top 15 features ranked by importance were predominantly profile-derived intensity features in contrast to Gaussian fit–based features.

Conclusion: Using post-therapy imaging FU, this work demonstrates that lesion-level intensity profile–based features can differentiate benign from malignant findings on 18F-PSMA PET/CT images, highlighting the importance of FU validated CAD approaches.

CONTACT:
George C. Kagadis, MSc, PhD, FAAPM
University of Patras, Rion, Greece
kagadis@upatras.gr / gkagad@gmail.com

INTRODUCTION

Prostate cancer (PCa) is the most frequent malignancy among men[1]. Prostate-specific membrane antigen (PSMA) is overexpressed in prostate cancer cells. When labeled with ^{18}F for PET/CT imaging, it enables the visualization of metastases in recurrent PCa cases[2]. However, a major clinical challenge associated with ^{18}F-PSMA PET/CT imaging is the frequent occurrence of false-positive findings, particularly within the ribs and pelvic bones.[3] These false positive findings must be better understood to prevent clinical over-staging and subsequent inappropriate treatment. However, establishing a reliable ground truth is highly challenging because histopathologic confirmation is usually unavailable. To address this challenge, this study introduces a computer-aided diagnosis (CAD) system trained on the baseline images with follow-up (FU) information as a gold standard. Our decision-support tool aims to assist physicians by providing a more reliable lesion classification than single-time-point visual interpretation by the nuclear medicine physician, ultimately reducing the misclassification of non-specific, benign uptake.

METHODS AND MATERIALS

Our dataset consists of:
- 44 patients with biochemical recurrent PCa who underwent ^{18}F-PSMA PET/CT imaging
- 285 suspicious findings
- Two classes:
 - 123 benign findings
 - 162 malignant

Findings demonstrating treatment response in FU imaging, such as tumor regression or a visible reduction in size, were labeled as malignant, whereas findings that failed to respond to therapy and were interpreted as benign or non-specific by the nuclear medicine physician were labeled as benign. For the machine learning workflow, the maximum intensity pixel of each finding was identified to extract 1D intensity profiles along the x, y, and z axes within the axial plane.

Two categories of features were extracted from these profiles:
- Profile-based features
- Gaussian-based features

Profile-based features included:
- morphology metrics (profile length, max intensity position, min/max intensity values),
- distribution statistics (mean intensity, standard deviation (Std), skewness, and kurtosis),
- cumulative signal intensity via the Area Under the Curve (AUC), and
- first-derivative spatial transitions (mean, standard deviation, min, and max gradient magnitudes).
- cross-axis averages and inter-axis ratios of maximum intensity and gradient

Gaussian-based features were extracted by fitting a parametric Gaussian curve to each profile.

Gaussian-based features included:
- fundamental curve parameters (amplitude (α) mean spatial offset (μ) standard deviation (σ) sample fit lengths, and the coefficient of determination (R^2))
- Gaussian maximum slope,
- Full Width at Half Maximum (FWHM),
- Cross-axis and three-dimensional descriptors (3D Gaussian spread volume, curve peak sharpness, cross-axis max slope ratios, and absolute 3D gradient magnitude)

A two-stage stacking ensemble architecture was utilized (Figure 1). Out-of-fold probability predictions from the base models were used as inputs to a Logistic Regression meta-learner for final classification.

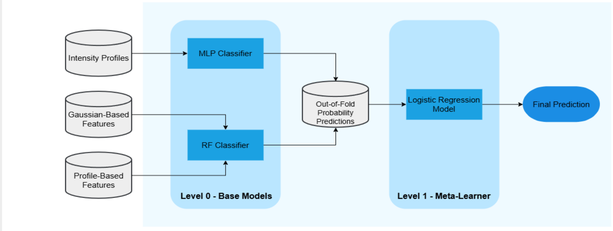


Figure 1. Stacking ensemble architecture.

RESULTS

Performance (mean ± Std, 10-fold cross-validation):
- Accuracy: 92.6 ± 5.3%,
- Sensitivity: 94.4 ± 6.2%,
- Specificity: 90.3 ± 10.7%,
- AUROC: 97.4 ± 3.6% (Figure 2)

As shown in Figure 3:
- Most frequently selected feature: Mean Gradient (X)
- Highest importance score: Gradient Std (X)

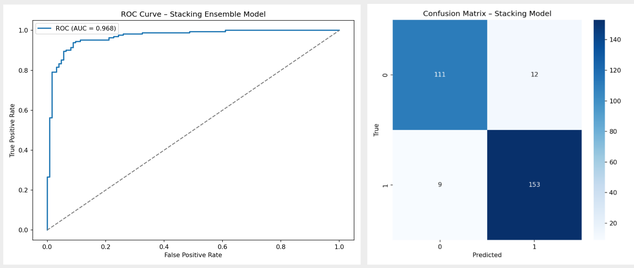


Figure 2. ROC curve (left) and confusion matrix (right) for the stacking ensemble classifier.

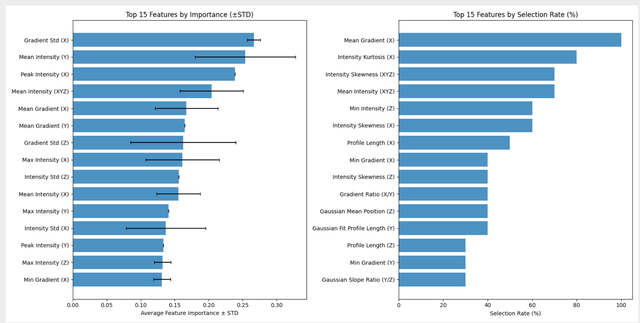


Figure 3. Feature importance and selection rate across 10-fold cross-validation for the stacking ensemble classifier.

DISCUSSION

- Profile-based intensity features and gradient descriptors were the strongest predictors, consistently outperforming Gaussian-based features in feature importance and selection frequency (Figure 3).
- Localized, asymmetric spatial changes contain the important information needed to identify true malignancy.
- Consistent with previous studies[4,5], high classification performance was achieved. However, unlike prior AI studies, post-therapy imaging follow-up rather than expert visual interpretation was used as the objective reference standard.

CONCLUSIONS

Take-home messages:

1. A novel stacking ensemble framework successfully differentiated malignant from benign ^{18}F-PSMA PET/CT findings, achieving 92.6% accuracy and 97.4% AUROC.
2. Post-therapy imaging follow-up provided an objective reference standard for malignancy, overcoming the lack of histopathological confirmation in biochemical recurrence.
3. Localized 1D intensity profiles and spatial gradient features carry significantly higher diagnostic value than Gaussian-based features.
4. The proposed framework serves as an objective decision-support tool capable of reducing diagnostic uncertainty, preventing patient over-staging, and optimizing therapy paths.

REFERENCES

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. CA Cancer J Clin. 2022;72(1):7-33. doi:10.3322/caac.21708.
2. Lawhn-Heath C, Salavati A, Behr SC, et al. Prostate-specific Membrane Antigen PET in Prostate Cancer. Radiology. 2021;299(2):248-260. doi:10.1148/radiol.2021202771
3. Orevi M, Ben-Haim S, Abourbeh G, et al. False Positive Findings of [18F]PSMA-1007 PET/CT in Patients After Radical Prostatectomy with Undetectable Serum PSA Levels. Front Surg. 2022;9. doi:10.3389/fsurg.2022.943760.
4. Erle A, Moazemi S, Lutje S, Essler M, Schultz T, Bundschuh RA. Evaluating a machine learning tool for the classification of pathological uptake in whole-body PSMA-PET/CT scans. Tomography. 2021;7(3):301-312. doi:10.3390/tomography7030026.
5. Li Y, Imami MR, Zhao L, et al. An automated deep learning-based framework for uptake segmentation and classification on PSMA PET/CT imaging of patients with prostate cancer. J Imaging Inform Med. 2024;37(5):2206-2215. doi:10.1007/s10278-024-01046-6.

© Poster Template by Genigraphics® 1.800.790.4001 www.genigraphics.com