Anatomy-Informed Deep Learning for Dual-Tracer PET/CT Lesion Detection

DeepPSMA – Grand Challenge 2025

Arthur Dujardin Paul Blanc-Durand

August 20, 2025

Abstract

We present our submission to the DEEP-PSMA Grand Challenge, focusing on automated lesion detection in PSMA and FDG PET/CT imaging. Our approach leverages ensemble methods with nnU-Net architectures, incorporating organ segmentation masks and signed distance functions (SDF) to improve lesion detection accuracy. We explore multiple preprocessing strategies, including binary organ masks and distance-based representations, demonstrating the potential of anatomical context for reducing false positives and improving clinical robustness. Our postprocessing pipeline introduces tracer-specific refinements, particularly a PSMA-guided correction of FDG lesions, which reflects clinical knowledge of tracer complementarity in prostate cancer imaging.

1 Introduction

Automated lesion detection in PET/CT represents a critical step in improving efficiency and reproducibility in nuclear medicine. Traditionally, lesion delineation has been performed manually by physicians, a process that is labor-intensive, time-consuming, and prone to variability across readers. With the growing role of PET tracers such as FDG and PSMA, the clinical demand for robust AI solutions is stronger than ever.

PSMA PET/CT has rapidly gained traction for prostate cancer detection and staging, offering higher specificity for prostate-derived lesions. FDG PET/CT, by contrast, is widely used across oncology because it reflects glucose metabolism, but suffers from high levels of non- specific uptake in physiological organs. When both tracers are used in tandem, physicians can obtain a complementary view: PSMA identifies prostate-related lesions, while FDG helps capture aggressive disease phenotypes and dedifferentiated tumors.

However, analyzing both tracers simultaneously presents a non-trivial computational problem. Lesions may appear in one tracer but not the other, and physiological uptake differs substantially. An automated algorithm must therefore adapt to tracer-specific uptake patterns while still ensuring consistency across the dual-tracer framework.

The DEEP-PSMA challenge provides a unique opportunity to evaluate methods in this context. Our submission introduces an **ensemble-based nnU-Net** [1] pipeline that integrates anatomical priors through organ masks and distance transforms, explores **signed distance functions** for better organ-lesion boundaries, and implements **cross-tracer postprocessing** where PSMA is used to refine FDG predictions.

2 Methodology

2.1 Ensemble architecture

We adopt the nnU-Net framework because of its adaptive design and proven track record in medical image segmentation. Instead of relying on a single model, we build an ensemble of nnU-Net variants trained on

distinct dataset configurations and preprocessing strategies. This ensemble strategy improves robustness and accounts for the inherent heterogeneity in PET/CT data.

Our ensemble includes:

1. **Standard nnUNet** - baseline model with default hyperparameters.
2. **nnUNetResEncUNetL** - large residual encoding variant with increased representational capacity. representational capacity.
3. **nnUNetResEncUNetM** - medium-sized residual encoding variant offering a trade-off between accuracy and computational cost.

2.2 Dataset configurations

To maximize performance, each variant is trained on one of four dataset configurations:

1. **Baseline** (801/802) - standard PET/CT preprocessing without organ context.
2. **Organ-enhanced** (901/902) - integration of TotalSegmentator masks collapsed into eight simplified classes: spleen, kidney, liver, lung, brain, bladder, thyroid, and osteomedullary structures (bones).
3. **Corrected** (911/912) - datasets with manual corrections of annotations combined with organ context.
4. **SDF-enhanced** (921/922) - binary organ masks converted into signed distance functions with 10 mm sigmoid smoothing, encoding proximity information.

This multi-dataset training ensures that each model specializes in a slightly different view of the same problem, while the ensemble integrates their strengths.

2.3 Preprocessing pipeline

PET images are normalized by **SUV thresholds**, a clinically meaningful scale that ensures comparability across patients. For CT, images are resampled to PET resolution, clipped to the 5th-95th percentiles, and standardized by mean/variance normalization.

For organ-enhanced configurations, **TotalSegmentator outputs 104 labels**, which we map into eight clinically relevant categories. This reduced label space both simplifies the task and encodes key organs where physiological uptake is problematic.

The **SDF preprocessing** represents a novel contribution: instead of using binary organ masks, we compute signed distance maps, where positive values denote voxels inside the organ and negative values outside. After applying sigmoid smoothing, we obtain soft organ boundaries. This strategy provides the network with nuanced spatial context and helps distinguish lesions located at or near organ borders

3 Training and Inference

We train all models using **5-fold cross-validation** to maximize data utilization. Folds are chosen to balance patient characteristics and imaging distributions. Training uses the nnU-Net default trainer for most datasets, while complex configurations such as SDF-enhanced models employ **extended 250-epoch schedules** for stable convergence.

Optimization is performed with **SGD + momentum**, and the loss combines **Dice and Cross- Entropy**, ensuring both overlap maximization and voxel-wise stability. Mixed-precision training on CUDA devices accelerates convergence without sacrificing accuracy.

Inference proceeds sequentially: PSMA predictions are generated first, followed by FDG predictions. This ordering allows PSMA masks to guide FDG postprocessing, reflecting the clinical reality that PSMA is usually more sensitive for differentiated prostate cancer.

4 Postprocessing Pipeline

4.1 Standard morphological refinements

The baseline postprocessing provided by the organizers includes morphological expansion and contraction with a 7 mm radius. We adopt this approach but extend it by excluding specific organs from expansion, including spleen, kidneys, liver, bladder, and brain. This exclusion prevents propagation of lesions into physiologically active regions.

4.2 Tracer-specific thresholds

Thresholding values are tailored to each tracer. For PSMA, balanced thresholds (0.5/0.5) are used to maximize sensitivity. For FDG, asymmetric thresholds (0.25/0.75) are chosen, acknowledging FDG's higher metabolic variability and the risk of false positives.

4.3 Cross-tracer refinement

Our most important innovation is **cross-tracer refinement of FDG lesions using PSMA**. The rationale is that prostate adenocarcinoma lesions are usually visible in PSMA, while FDG positivity alone may correspond to inflammation or non-specific uptake.

The procedure is:

1. Map each FDG lesion to its anatomical organ class using TotalSegmentator labels.
2. Extract features: **SUVmax**, **log(volume)**, and **presence of a corresponding PSMA lesion**.
3. Train a logistic classifier on the training dataset.
4. Retain FDG lesions only if the classifier score exceeds 0.5, corresponding to 96% sensitivity and 99% specificity.

This rule reduces false positives while preserving clinically meaningful FDG lesions that are PSMA-negative, such as dedifferentiated disease.

5 Results & Discussion

5.1 Impact of anatomical context

Introducing organ masks, whether binary or SDF-enhanced, consistently reduced false positives. The model learns to disregard high uptake in organs like bladder or liver, which otherwise generate many false detections. The SDF variant, in particular, improved delineation at organ boundaries, though at the cost of longer training times.

5.2 Value of ensemble learning

The ensemble strategy significantly outperformed single models. Baseline nnU-Nets excelled at large, high-contrast lesions, while organ-enhanced and SDF models better handled small or ambiguous cases. By averaging predictions, the ensemble benefited from both robustness and sensitivity.

5.3 Cross-tracer refinement

The logistic refinement step eliminated many spurious FDG lesions, particularly in the bowel and bone marrow, two regions prone to false uptake. Importantly, this refinement did not erase clinically relevant lesions, preserving sensitivity. This step demonstrates the potential of **incorporating clinical priors directly into AI postprocessing.**

5.4 Limitations

- Training and inference were resource-intensive, especially for SDF datasets.
- Some FDG-positive but PSMA-negative lesions may have been incorrectly discarded if their feature values were atypical.
- We did not yet explore uncertainty estimation, which could further guide clinical adoption by flagging doubtful predictions.

5.5 Conclusion

Our submission to the DEEP-PSMA Grand Challenge demonstrates the effectiveness of combining **nnU-Net ensembles**, **anatomical priors**, and **cross-tracer refinements** for PET/CT lesion detection. By incorporating organ segmentation masks and signed distance representations, the models achieve more clinically trustworthy segmentation, reducing false positives in physiologically active organs.

The cross-tracer refinement step, where PSMA predictions refine FDG outputs, exemplifies how domain knowledge can be fused with machine learning to mirror real-world diagnostic reasoning.

Future work will extend this pipeline with uncertainty quantification, adaptive thresholding, and prospective multi-center validation. Ultimately, we aim to provide nuclear medicine physicians with a tool that is not only accurate but also interpretable, efficient, and ready for integration into clinical workflows

6 References

[1] Fabian Isensee et al. “nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation”. In: *CoRR* abs/1809.10486 (2018). arXiv: 1809.10486. URL: http://arxiv.org/abs/1809.10486.