

Multi-Time-Point Kidney Segmentation in Lutetium-177 SPECT/CT Imaging Using a 3D Deep Learning Approach for Patient-Specific Dosimetry

Jirachaya Chaijinda

Chulalongkorn university

Nattapon Kamboonsri

Chulalongkorn university

Kitiwat Khamwan

Chulalongkorn university

Manasavee Lohvithee

`manasavee.l@chula.ac.th`

Chulalongkorn university

Research Article

Keywords: Kidney segmentation, Lutetium-177, SPECT/CT imaging, 3D deep learning, Patient-specific dosimetry, Radiopharmaceutical therapy

Posted Date: October 6th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11150972/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Multi-Time-Point Kidney Segmentation in Lutetium-177 SPECT/CT Imaging Using a 3D Deep Learning Approach for Patient-Specific Dosimetry

Jirachaya Chaijinda,¹ Nattapon Kamboonsri,¹ Kitiwat Khamwan,² and Manasavee Lohvithee¹

¹Department of Nuclear Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok, Thailand

²Department of Radiology, Division of Nuclear Medicine,
Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand

Background: Multi-time-point kidney segmentation from Lutetium-177 (Lu-177) SPECT/CT is essential for patient-specific dosimetry for radiopharmaceutical therapy (RPT). However, expert segmentation remains time-consuming and subject to inter-observer variability, limiting its routine clinical use.

Methods: This study compared the performance of four 3D deep-learning models including U-Net, Attention U-Net, UNETR, and SwinUNETR, for automated kidney segmentation and assessed their impact on internal dosimetry. A retrospective study was conducted using SPECT/CT datasets from patients who received Lu-177 PSMA or Lu-177 DOTATATE therapy. Performance was assessed using Dice Similarity Coefficient (DSC) and 95th percentile Hausdorff Distance (HD95). The best-performing model was applied to internal dosimetry calculations using MIRDcalc software.

Results: SwinUNETR achieved the highest segmentation performance with DSC of 0.91 ± 0.07 and HD95 of 12.50 ± 14.82 mm. Dosimetry based on automated segmentation showed a mean absorbed dose difference of 10.28% relative to expert-based dosimetry. The DOTATATE group received mean kidney dose of 4.74 ± 1.70 Gy/cycle, while the PSMA group received 1.73 ± 1.89 Gy/cycle. Automated segmentation reduced processing time by 81–83% while maintaining accuracy comparable to expert segmentation.

Conclusion: 3D deep-learning model, particularly SwinUNETR, demonstrates clinical feasibility for automated segmentation in Lu-177 SPECT/CT dosimetry workflows. Moreover, it significantly reduces clinician workload while maintaining clinically acceptable dosimetric accuracy.

Keywords: Kidney segmentation, Lutetium-177, SPECT/CT imaging, 3D deep learning, Patient-specific dosimetry, Radiopharmaceutical therapy

1 Introduction

Radiopharmaceutical therapy (RPT) has become an important precision oncology approach in which radiolabelled compounds are used to deliver targeted ionizing radiation directly to tumor cells while minimizing radiation exposure to surrounding healthy tissues [1]. Lutetium-177 (Lu-177) labeled radiopharmaceuticals are widely used in radiopharmaceutical therapy (RPT) for progressive neuroendocrine tumors positive for somatostatin receptors (NETs) and metastatic castration-resistant prostate cancer [2, 3]. Despite the therapeutic efficacy of Lu-177 based, the kidneys represent a critical dose-limiting organ due to the renal clearance pathway of these agents. Accumulation of radiolabelled compounds in the renal parenchyma can lead to radiation nephropathy if cumulative absorbed doses exceed the thresholds, typically considered to be around 23 Gy in patients without risk factors [4, 5]. Given this dose-limiting relationship, accurate patient-specific dosimetry is essential for both optimizing therapeutic efficacy and ensuring renal safety across treatment cycles. The Medical Internal Radiation Dose (MIRD) formalism provides the established framework for calculating absorbed dose from time-integrated activity measurements, which in turn requires precise volumetric delineation of the kidneys at each imaging time point [6].

Current clinical practice relies predominantly on expert segmentation of the kidneys from Lu-177 SPECT/CT images acquired at multiple post-injection time points, typically 4, 24, 48, and 72 hours. This process is labour-intensive, requiring approximately 10 minutes per patient at each time point for an experienced nuclear medicine physician, and is subject to inter- and intra-observer variability that may contribute to dosimetric uncertainty [7, 8]. As patients may undergo multi-

ple SPECT/CT acquisitions across several treatment cycles, the associated manual workload can increase substantially, placing a significant burden on clinical staffs and limiting the routine implementation of personalized dosimetry. A validated automated segmentation pipeline could help addressing this limitation by standardizing kidney delineation, reducing observer variability and improving workflow efficiency.

Deep learning-based image segmentation has transformed medical image analysis, with convolutional neural network (CNN) architectures such as U-Net demonstrating strong performance across anatomical structures and imaging modalities [9]. More recently, Vision Transformer (ViT)-based architectures, including UNETR and SwinUNETR, have further advanced 3D volumetric segmentation by enabling long-range contextual learning through self-attention mechanisms [10, 11].

In recent years, many studies have investigated kidney segmentation in Lu-177 SPECT/CT. However, there is still no comprehensive study covering the entire process that combines state-of-the-art 3D deep learning methods for multi-time-point kidney segmentation with subsequent medical analyses extending conventional segmentation evaluation metrics. Spink et al. [7] performed manual kidney segmentation and compared single-time-point (STP) against multi-time-point (MTP) dosimetry methods for kidney dose estimation after Lu-177-DOTATATE. Nazari et al. [8] developed a 2D deep learning model for kidney and liver segmentation from CT and subsequently applied masks to SPECT image for activity quantification. Dewaraja et al. [12] used a 3D modified U-Net for single-time-point kidney segmentation from CT in the Lu-177-DOTATATE pipeline. Leube et al. [4] demonstrated that incorporating image processing improves

CT-based kidney segmentation, showing that a PET-derived coarse mask used to pre-segment the CT significantly outperformed CT-only segmentation in PSMA-PET/CT patients.

This study therefore applied 3D deep-learning models for multi-time-point kidney segmentation from Lu-177 SPECT/CT data. Each model was evaluated under two input conditions: CT only and combined SPECT/CT. The best-performing model was subsequently integrated into a complete dosimetry pipeline using the MIRDcalc software to evaluate kidney dose estimation in clinical pipeline.

Materials and Methods

This section describes the setup of dataset preparation, data preprocessing, deep learning-based methods for segmentation, experimental design, and clinical evaluation. The overall study workflow is illustrated in Fig. 1.

Study Design and Patient Population

Dataset was collected from patients who received Lu-177 PSMA or Lu-177 DOTATATE therapy at King Chulalongkorn Memorial Hospital between January 2023 and March 2026. Inclusion criteria were: (1) adult patients (≥ 18 years) receiving Lu-177 PSMA or Lu-177 DOTATATE; and (2) availability of at least two SPECT/CT acquisitions per treatment cycle. Patients were excluded if they had fewer than two imaging time points or had renal pathology. A total of 10 patients (6 receiving Lu-177 PSMA and 4 receiving Lu-177 DOTATATE) with 29 SPECT/CT samples were included in the final analysis. Each patient contributed 2-3 usable time points within 4-72 hours post-injection, with the PSMA and DOTATATE groups contributing 12 and 17 samples, respectively.

SPECT/CT Imaging Protocol

All patients were imaged using a Veriton-CT SPECT/CT scanner. Whole-body SPECT/CT acquisitions were performed at multiple post-injection time points (4, 24, 48, and/or 72 h). Each acquisition consisted of six bed positions with acquisition times ranging from 3 to 6 minutes per bed position, resulting in a total scan duration of approximately 28 minutes. Data acquisition was centered on the 113 keV photopeak of Lu-177 using a $\pm 10\%$ energy window. Quantitative SPECT images were reconstructed using the ordered subset attenuation correction (OSAC) algorithm with 12 iterations 8 subsets, and CT images were acquired for attenuation correction and anatomical reference. All images were stored in DICOM format.

Ground Truth Segmentation

Expert segmentation (ground truth) was performed by an experienced nuclear medicine physician using MIM SurePlan™ MRT voxel-based dosimetry software (version 7.2, MIM Software Inc., Cleveland, OH). Bilateral kidneys were manually contoured slice-by-slice on the CT images at each time point. All segmentations were independently reviewed by a second expert to minimise inter-observer variability. Contours were exported in NIfTI format (.nii.gz) for model training and evaluation.

Data Preprocessing

SPECT and CT images were reoriented to the standard Right-Anterior-Superior (RAS) coordinate system. CT intensities were clipped to the range $[-175, 250]$ Hounsfield Units (HU) to encompass soft tissue and organ boundaries [13], then normalized to $[0, 1]$. SPECT images were normalized using 5th–95th percentile scaling to reduce the influence of outlier values [14], followed by rescaling to $[0, 1]$. For each time-point dataset, volumes were cropped using a bounding box derived from the ground truth kidney mask with a 10-voxel margin on all sides, then resized to a fixed spatial dimension before input to the networks. For combined CT and SPECT models, the two modalities were concatenated as a dual-channel input. Random horizontal flipping (probability 0.5) was applied as data augmentation during training. Finally, outputs were rescaled to the original resolution.

Deep Learning Architectures

Four 3D deep learning architectures were evaluated: (1) U-Net utilizes a symmetric encoder-decoder framework with skip connections to extract hierarchical features while preserving high-resolution spatial details for accurate anatomical reconstruction [9], (2) Attention U-Net extends the standard U-Net by integrating attention gates into skip connections to dynamically highlight target anatomical regions while suppressing irrelevant background noise [15], (3) UNETR combines a 12-layer Vision Transformer encoder with a 3D CNN decoder to model long-range spatial dependencies from non-overlapping $16 \times 16 \times 16$ voxel patches across the entire volume [10], and (4) SwinUNETR, a hierarchical Swin Transformer encoder paired with a residual CNN decoder, operating through four progressive downsampling stages using window-based multi-head self-attention (W-MSA) and shifted-window multi-head self-attention (SW-MSA) [11]. All models were implemented in Python using the MONAI [22] and PyTorch [23] libraries.

Model Training

All models were trained for 2,000 epochs using a combined Dice and cross-entropy loss function (DiceCE loss), optimized with Adam and a cosine annealing learning rate scheduler. A learning rate of 1×10^{-4} was used for all architectures. Gradient accumulation was used to stabilize training at a batch size of 1. Sliding window inference was applied during validation, with window sizes of 192^3 voxels for SwinUNETR and 256^3 voxels for the remaining architectures. Model weights were evaluated every 2 epochs, and the best-performing weights, defined as those achieving the highest validation DSC, were saved. All training was performed on four NVIDIA A100 GPUs. The complete training and validation pipeline for both the CT-only and combined SPECT/CT models is summarised in Fig. 2.

Experimental Design

Each architecture was evaluated under two input conditions: CT-only and combined SPECT/CT. Performance was assessed via 5-fold cross-validation with an approximate 60%/20%/20% training/validation/test split. Six quantitative metrics were computed: (1) Dice Similarity Coeffi-

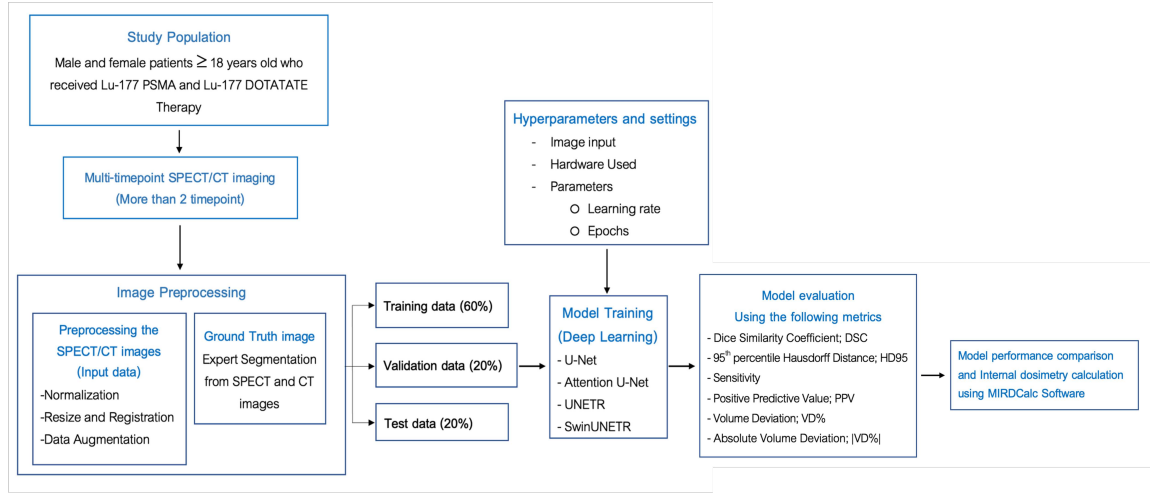


Fig. 1. Overall study workflow for deep learning-based kidney segmentation and internal dosimetry using multi-time-point SPECT/CT imaging.

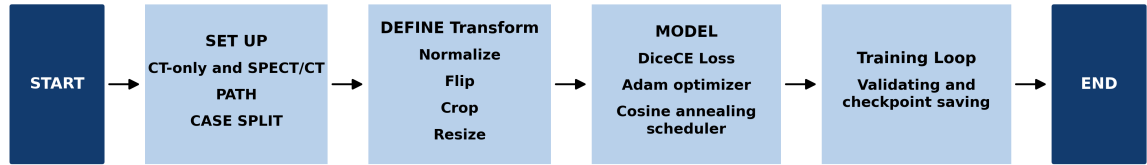


Fig. 2. Overview of the data preparation and model training workflow for CT-only and SPECT/CT kidney segmentation models.

cient (DSC); (2) 95th percentile Hausdorff Distance (HD95); (3) Sensitivity (True Positive Rate); (4) Positive Predictive Value (PPV); (5) Volume Deviation (VD%) and (6) Absolute Volume Deviation (|VD%|). Statistical comparisons between CT-only and SPECT/CT inputs were performed using the Wilcoxon signed-rank test, with significance set at $p < 0.05$.

Internal Dosimetry

The result from the best-performing model was further analysed to compute kidney-absorbed doses using MIRDCalc (v1.1-Genesis, SNMMI, New York) [16]. Automated segmentation masks were overlaid on quantitative SPECT images to extract kidney activity (MBq) at each time point. Time-Activity Curves (TAC) were fitted using monoexponential extrapolation to calculate Time-Integrated Activity Coefficients (TIAC, hours) for the left and right kidneys separately. Patient-specific phantom models were selected from the International Commission on Radiological Protection (ICRP) interpolation library based on patient gender and body weight [17]. Absorbed doses (Gy) were calculated for each kidney and compared with those derived from expert segmentation using Bland-Altman analysis.

Results

Segmentation Performance

The performance of kidney segmentation for all four architectures under both input conditions is presented in Table 1. SwinUNETR with CT-only input achieved the highest overall performance across the primary metrics (DSC: 0.91 ± 0.07 , Sensitivity: 0.93 ± 0.03 , HD95: 12.50 ± 14.82 mm, PPV: 0.90 ± 0.10). Attention U-Net recorded the lowest volume deviation (VD%) among all CT-only models ($1.96 \pm 14.01\%$), whereas the lowest absolute volume deviation (|VD%|) overall was achieved by SwinUNETR with SPECT/CT input (9.01 ± 15.54).

To complement the quantitative metrics summarised in Table 1, representative 3D visualizations of the segmentation outputs are presented in Fig. 3. Representative cases were selected by ranking all test-set patients according to their DSC with the SwinUNETR (CT input), from lowest to highest agreement with the expert (ground-truth) segmentation. Three cases spanning this range were selected for visualization: the case with the lowest DSC (largest discrepancy, DSC = 0.58 for CT input and DSC = 0.59 for SPECT/CT input), a case with an intermediate DSC (moderate discrepancy, DSC = 0.88 for CT input and DSC = 0.86 for SPECT/CT input), and the case with the highest DSC (best case, DSC = 0.96 for CT input and DSC = 0.97 for SPECT/CT input). These examples clearly illustrate that the case with the largest dis-

Table 1. Segmentation performance of the four 3D deep learning architectures under 5-fold cross-validation (mean $\pm$ SD).

Model	Input	DSC ($\uparrow$)	HD95 ($\downarrow$), mm	Sensitivity ($\uparrow$)	PPV ($\uparrow$)	VD, % ($\downarrow$)	IVDI, % ($\downarrow$)
U-Net	CT	0.86 $\pm$ 0.11	13.99 $\pm$ 10.32	0.89 $\pm$ 0.07	0.84 $\pm$ 0.15	12.47 $\pm$ 34.56	19.19 $\pm$ 33.24
	SPECT/CT	0.86 $\pm$ 0.06	17.01 $\pm$ 11.98	0.90 $\pm$ 0.05	0.85 $\pm$ 0.16	13.92 $\pm$ 40.12	21.51 $\pm$ 38.86
Attention U-Net	CT	0.87 $\pm$ 0.07	22.26 $\pm$ 16.67	0.87 $\pm$ 0.07	0.87 $\pm$ 0.10	1.96 $\pm$ 14.01*	9.39 $\pm$ 8.74
	SPECT/CT	0.83 $\pm$ 0.08	28.67 $\pm$ 21.50	0.87 $\pm$ 0.11	0.79 $\pm$ 0.17	15.25 $\pm$ 17.77	17.20 $\pm$ 19.84
UNETR	CT	0.87 $\pm$ 0.07	25.55 $\pm$ 19.73	0.91 $\pm$ 0.05	0.85 $\pm$ 0.10	9.21 $\pm$ 18.32	13.03 $\pm$ 15.68
	SPECT/CT	0.84 $\pm$ 0.08	39.10 $\pm$ 19.76	0.88 $\pm$ 0.06	0.81 $\pm$ 0.10	10.97 $\pm$ 20.78	17.50 $\pm$ 19.09
SwinUNETR	CT	0.91 $\pm$ 0.07*	12.50 $\pm$ 14.82*	0.93 $\pm$ 0.03*	0.90 $\pm$ 0.10*	6.75 $\pm$ 17.78	9.14 $\pm$ 16.10
	SPECT/CT	0.90 $\pm$ 0.07	13.60 $\pm$ 15.50	0.92 $\pm$ 0.04	0.90 $\pm$ 0.10	5.03 $\pm$ 17.45	9.01 $\pm$ 15.54*

*Best value in respective column. DSC = Dice Similarity Coefficient; HD95 = 95th percentile Hausdorff Distance; PPV = Positive Predictive Value; VD% = Volume Deviation; IVDI% = Absolute Volume Deviation.

crepancy exhibits substantial distortion of the resulting kidney structure, whereas the case with moderate discrepancy, despite closely resembling the ground truth in overall kidney morphology, still shows residual island noise. In contrast, the best-performing case exhibits neither issue. These qualitative differences were consistent regardless of whether CT or SPECT/CT was used as the input modality.

Statistical comparison using the Wilcoxon signed-rank test within SwinUNETR (Fig. 4) showed that DSC ($p = 0.1840$), HD95 ($p = 0.2250$), and PPV ($p = 0.4280$) did not differ significantly between CT-only and SPECT/CT inputs, indicating that the overall shape agreement and boundary accuracy of the segmentation were preserved regardless of input configuration. Notably, incorporating SPECT alongside CT significantly reduced volume deviation (VD%; $p = 0.0054$). This improvement in volumetric accuracy is particularly relevant for dosimetric applications, where directional bias in organ volume estimation can propagate into systematic errors in absorbed dose calculation across treatment cycles. The absolute magnitude of volumetric error (IVDI%) did not differ significantly ($p = 0.7766$), suggesting that this benefit stemmed specifically from correcting directional bias rather than a broad reduction in error magnitude. Sensitivity was significantly higher with CT-only input than with SPECT/CT input ($p = 0.0007$), consistent with the lower and more diffuse radiopharmaceutical accumulation typically observed at the kidney periphery compared with the renal cortex.

Internal Dosimetry

A representative example of the dosimetric workflow applied to the automated segmentation output is shown in Fig. 5. The predicted kidney contour was overlaid onto the co-registered SPECT image to delineate the kidney volume of interest, from which kidney activity was extracted at each imaging time point. A time-activity curve was subsequently fitted for each patient, from which the time-integrated activity coefficient (TIAC) was derived by integration of the curve, as illustrated for a representative patient in Fig. 6.

The SwinUNETR model was applied to all eligible patients ($n = 10$) for dosimetry calculation. Bland-Altman analysis was performed to compare automated and expert segmentation-derived estimates across three parameters: time-integrated activity coefficient (TIAC; Fig. 7), absorbed dose (Fig. 8), and kidney volume (Fig. 9).

For TIAC (Fig. 7), the mean difference between automated and expert segmentation was -4.42% (95% Lim-

its of Agreement [LoA]: -105.05% to 96.20%) for CT input, widening to -26.34% (LoA: -132.42% to 79.73%) for SPECT/CT input, indicating a larger systematic underestimation and wider variability when SPECT was incorporated.

For absorbed dose (Fig. 8), the mean difference was 10.26% (LoA: -70.80% to 91.31%) for CT input and 15.31% (LoA: -64.23% to 94.86%) for SPECT/CT input, with both configurations showing comparably wide limits of agreement.

For kidney volume (Fig. 9), the mean difference was 14.73% (LoA: -34.29% to 63.75%) for CT input and 13.95% (LoA: -36.06% to 63.97%) for SPECT/CT input, showing narrower and more comparable limits of agreement between the two input configurations relative to TIAC and absorbed dose.

Mean absorbed doses for each radiopharmaceutical group are summarized in Table 2.

Table 2. Mean absorbed doses (Gy/cycle) by radiopharmaceutical group.

Group	Expert Seg. (Gy)	CT Input (Gy)	SPECT/CT (Gy)
DOTATATE	4.50 $\pm$ 1.15	4.74 $\pm$ 1.70	4.34 $\pm$ 1.57
PSMA	1.64 $\pm$ 1.53	1.73 $\pm$ 1.89	1.92 $\pm$ 1.81

Values expressed as mean $\pm$ SD per treatment cycle.

Segmentation Time

To evaluate the practical efficiency of the automated pipeline, segmentation time was compared between automated and expert segmentation. Automated segmentation time was measured as the model inference time alone, while expert segmentation time was recorded only for patients with more than one time point ($n = 7$); three patients were excluded from this comparison because only a single time point per case. Mean automated segmentation time was 1 minute 51 seconds (CT input) and 1 minute 37 seconds (SPECT/CT input) per patient per time point, compared with 9 minutes 37 seconds for expert segmentation (Table 3). This represents a time saving of approximately 81–83% relative to expert segmentation.

These findings indicate that automated segmentation can substantially reduce the workload associated with manual contouring, as the inference process operates automatically without requiring an expert. Manual contouring relies heavily on the expert's experience throughout the process and may be susceptible to error from fatigue, particularly when ana-

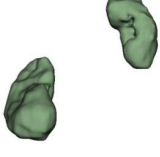
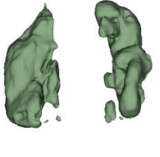
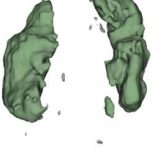
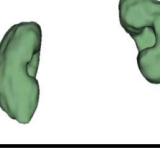
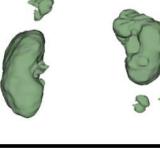
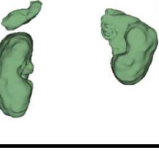
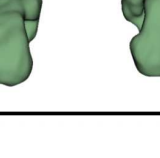
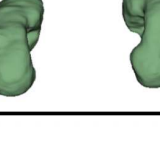
	Ground Truth (Reference)	CT Input	SPECT/CT Input
(a) Largest discrepancy		DSC = 0.58 	DSC = 0.59 
(b) Moderate discrepancy		DSC = 0.88 	DSC = 0.86 
(c) Best case		DSC = 0.96 	DSC = 0.97 

Fig. 3. Representative 3D visualizations of kidney segmentation outputs from SwinUNETR compared with the ground-truth (expert) segmentation. Columns from left to right: ground truth, CT input, SPECT/CT input. Rows from top to bottom: (a) case with the largest discrepancy, (b) case with moderate discrepancy, and (c) best case, with corresponding DSC values shown above each panel.

lyzing a large number of patients or multiple sequential images. In the context of ^{177}Lu -PSMA and ^{177}Lu -DOTATATE therapy, where each patient requires SPECT/CT acquisition at multiple time points per treatment cycle, this efficiency gain suggests that automated segmentation has the potential to improve the scalability of dosimetric workflows, accommodating a larger patient volume without compromising analytical quality.

Discussion

Input and Output

The input configurations evaluated in this study, CT-only (single-channel) and combined SPECT/CT (dual-channel), were designed to directly test whether functional information contributes to segmentation quality beyond what is achievable from anatomical information alone. In both configurations, the output was a binary segmentation mask for the left and right kidney at each individual imaging time point, which was subsequently used directly as the volume of interest for activity extraction and dosimetric calculation. In Leube et al. [4] work, a coarse PET-derived mask was first generated and used to explicitly pre-segment the CT before performing the fine segmentation. While the cascaded pre-segmentation strategy directly constrains the anatomical search space using functional information and yielded a significant improvement in DSC, the simple dual-channel concatenation approach used in this study requires the network to implicitly learn how to weight functional versus anatomical cues at every level of the architecture. This difference in strategy may explain why the addition of SPECT in the present study did not shown into a measurable improvement in overlap-based metrics (DSC, HD95, PPV), even though it produced a significant reduction in volumetric bias (VD%).

Table 3. Segmentation time comparison: automated vs. expert segmentation.

Patient	Automated Segmentation (min)		Expert Seg. (min)
	CT Input	SPECT/CT Input	
1	01:00	01:37	10:13
2	01:00	01:37	10:09
4	02:27	01:39	09:07
6	02:27	01:35	09:45
7	02:27	01:35	09:58
9	01:00	01:37	09:55
10	01:00	01:37	08:13
Mean	01:51	01:37	09:37

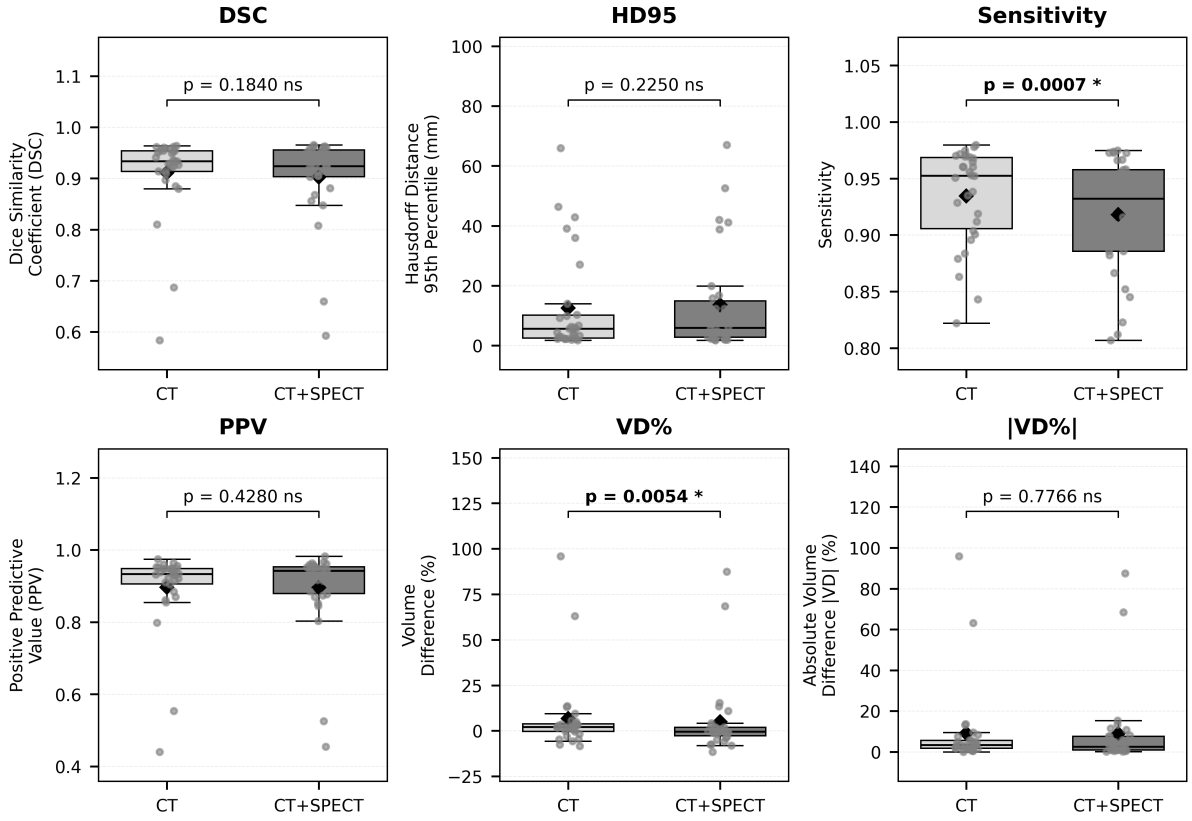


Fig. 4. Performance comparison between CT-only and SPECT/CT inputs for SwinUNETR across six evaluation metrics (DSC, HD95, Sensitivity, PPV, VD% and |VD%|). Statistical significance was assessed using the Wilcoxon signed-rank test (* $p < 0.05$: significant; ns: not significant).

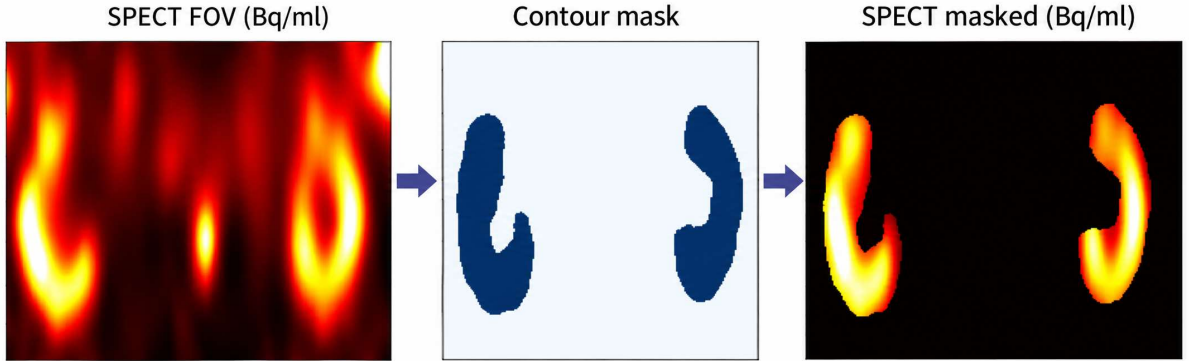


Fig. 5. Representative example of automated kidney contour delineation on SPECT imaging. (Left) SPECT field of view; (middle) predicted kidney contour mask; (right) SPECT image masked by the predicted contour to extract kidney activity.

Dataset

The dataset comprised 29 SPECT/CT acquisitions from 10 patients, with 2–3 usable time points per patient spanning 4–72 hours post-injection, and included both ^{177}Lu -PSMA and ^{177}Lu -DOTATATE cohorts. Although modest in absolute size compared with CT-only segmentation studies that leverage large public datasets [8], or single time-point

PET/CT cohorts of over 100 patients [4], this dataset is, to our knowledge, among the few to provide expert-verified, per-time-point kidney contours across multiple post-therapy SPECT/CT acquisitions for ^{177}Lu -labelled therapies. Combining two pharmacologically distinct radiopharmaceuticals into the dataset to enhance the generalizability of kidney anatomical variations.

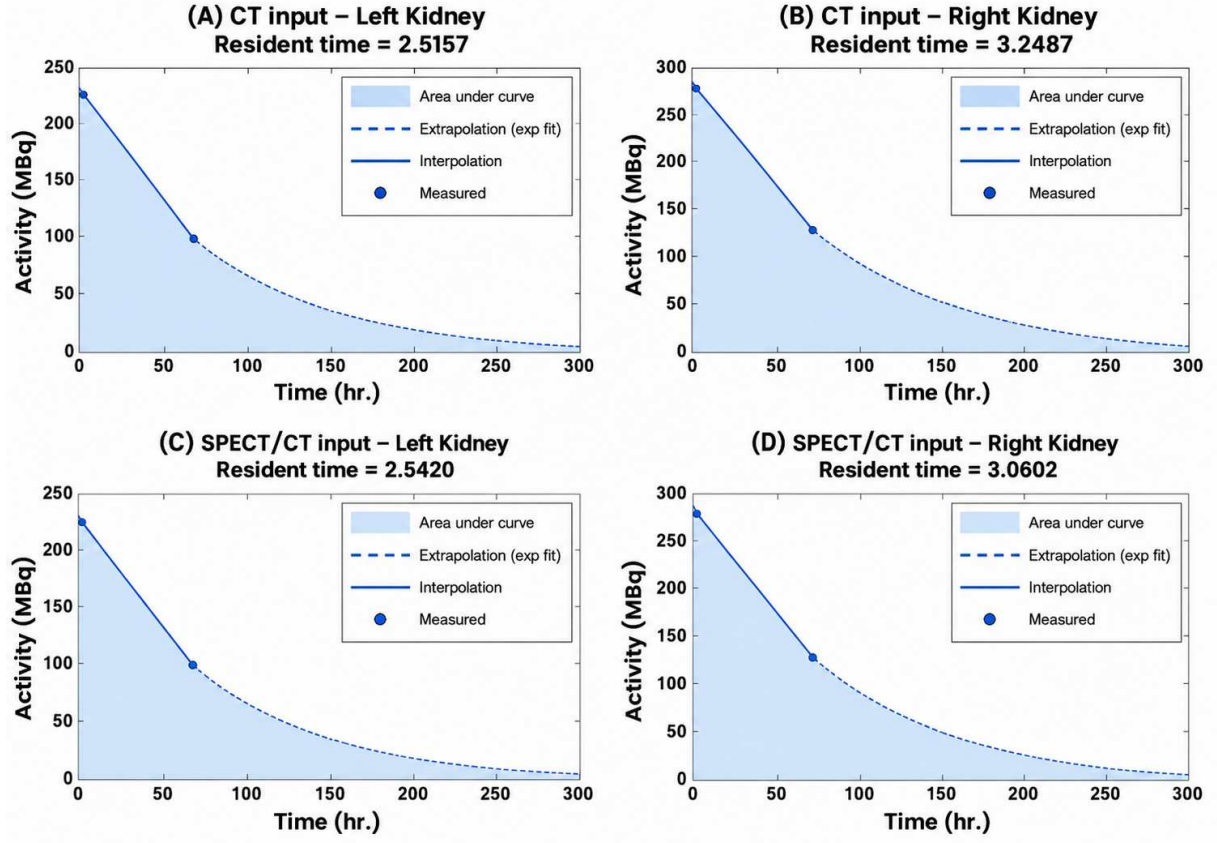


Fig. 6. Representative time-activity curves used to derive the time-integrated activity coefficient (TIAC), shown as resident time, for (A) left and (B) right kidney using CT-only input, and (C) left and (D) right kidney using SPECT/CT input.

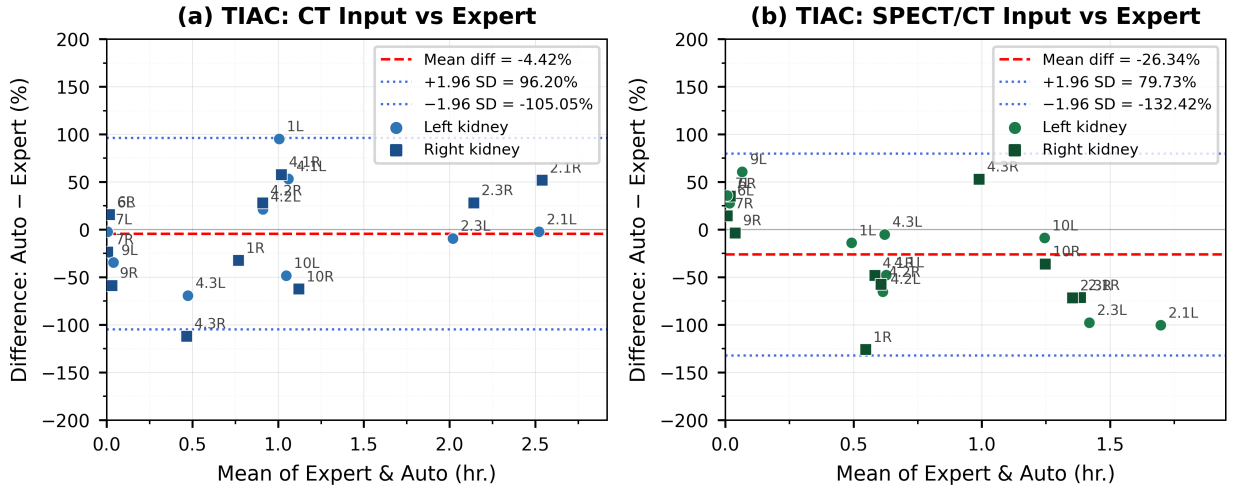


Fig. 7. Bland-Altman plots comparing automated and expert-derived time-integrated activity coefficient (TIAC) for (a) CT-only and (b) SPECT/CT inputs.

Model Architecture

In this study, the state-of-the-art 3D deep learning architectures were used, which is different from previous work [4, 8, 12]. Four 3D deep learning architectures were com-

pared: U-Net, representing a standard convolutional encoder-decoder baseline [9]; Attention U-Net, which augments skip connections with attention gates to suppress irrelevant background features [15]; UNETR, which replaces the convolutional encoder with a Vision Transformer operating on non-

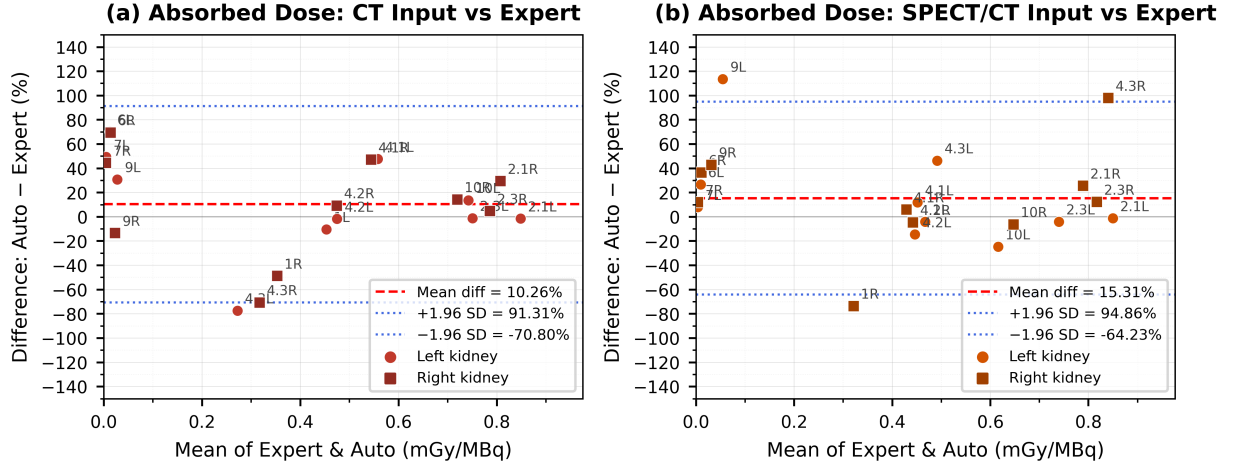


Fig. 8. Bland-Altman plots comparing automated and expert-derived absorbed dose for (a) CT-only and (b) SPECT/CT inputs.

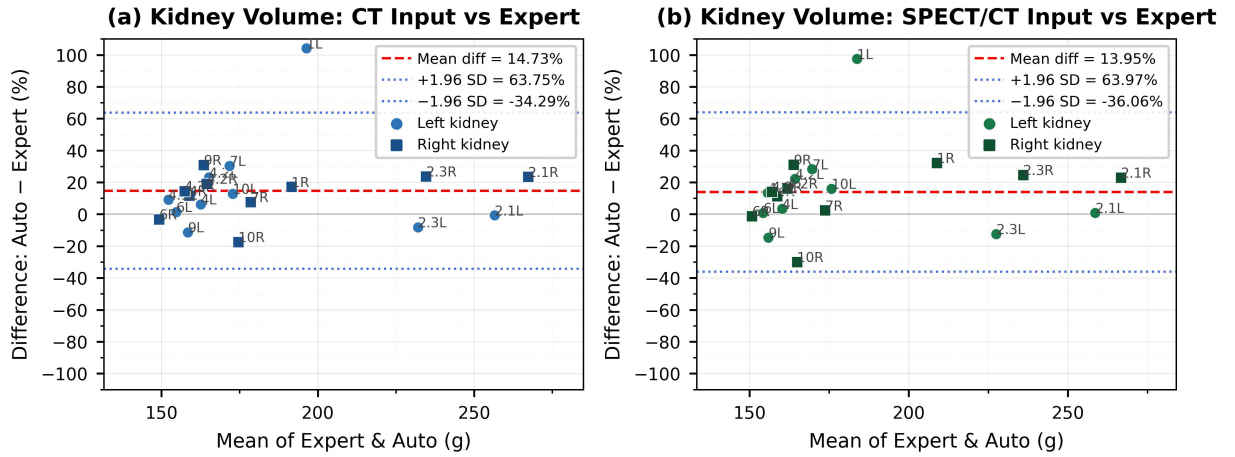


Fig. 9. Bland-Altman plots comparing automated and expert-derived kidney volume for (a) CT-only and (b) SPECT/CT inputs.

overlapping 3D patches [10]; and SwinUNETR, which uses a hierarchical, shifted-window transformer encoder to capture both local detail and long-range spatial context at multiple resolutions [11]. The models selected in this study were architecturally designed for volumetric segmentation and focuses on real-world clinical data to ensure their suitability for kidney segmentation.

Evaluation Metric and Internal Dosimetry

Model performance was assessed using a two-tiered evaluation framework. At the segmentation level, five-fold cross-validation was combined with six complementary metrics capturing overlap (DSC), boundary accuracy (HD95), detection sensitivity, positive predictive value, and volumetric agreement (VD% and IVD%), allowing overlap-based and volume-based performance to be assessed independently rather than assuming they are interchangeable. Across these kidney segmentation experiments, SwinUNETR achieved comparably strong performance regardless of whether CT-only or SPECT/CT data were used as input. At the dosimetric

level, the best-performing model was integrated into a complete internal dosimetry pipeline using MIRDcalc [16, 17], in which kidney activity extracted from the automated masks was fitted to a monoexponential time-activity curve to derive the time-integrated activity coefficient, and absorbed doses were calculated using gender- and weight-matched ICRP phantom models. Automated and expert-derived estimates of TIAC, absorbed dose, and kidney volume were then compared using Bland-Altman analysis.

DSC did not differ significantly between CT-only and SPECT/CT input, yet the corresponding volumetric and dosimetric estimates diverged between the two inputs. This finding indicates that segmentation overlap metrics such as DSC cannot reliably predict how accurately a model will perform in dosimetric calculations; validation of segmentation models intended for dosimetry must therefore also assess clinically relevant endpoints such as absorbed dose, in addition to segmentation accuracy[5].

Strengths and Limitations

A key finding of this study is that SwinUNETR-based kidney segmentation achieves high performance regardless of whether CT or SPECT/CT is used as input, while also significantly reducing the time required for segmentation. However, because MIRDcalc dosimetry calculations directly incorporate organ mass as a parameter, accurate volume estimation is at least as important as overlap-based metrics for clinical dosimetry [5]. The SPECT channel appears to improve the model's ability to delimit the functional kidney volume, defined as the region actively handling the radiopharmaceutical, rather than merely the anatomical boundaries visible on CT. This observation supports the use of SPECT/CT input in dosimetry-oriented workflows, particularly where precise volume estimation is prioritised.

Absorbed dose values in the DOTATATE group (4.74 ± 1.70 Gy/cycle) are consistent with published ranges from large multi-center studies: Sandström et al. reported 4.1–4.4 Gy/cycle in 500 patients [18], and Guerriero et al. reported a mean kidney absorbed dose of approximately 6 ± 2 Gy/cycle in a 25-patient cohort [19]. Absorbed doses in the PSMA group (1.73 ± 1.89 Gy/cycle) were lower and more variable, consistent with inter-patient pharmacokinetic variability reported in the PSMA literature [20]. The mean difference in absorbed dose between automated and expert segmentation was +10.26%, with Limits of Agreement ranging from –70.80% to +91.31%. Although this range is wider than ideal for individual patient-level dose reporting, it should be interpreted in the context of the limited sample size ($n = 10$) and the inherent variability of SPECT-based quantification, both of which are known to widen Bland-Altman limits even under conditions of good average agreement.

The 81–83% reduction in segmentation time achieved by the automated pipeline has direct clinical implications. In a standard Lu-177 dosimetry protocol requiring at least three SPECT/CT acquisitions per cycle, automated segmentation reduces the per-patient expert segmentation burden from approximately 29 minutes to approximately 5 minutes of inference time. This scalability is essential for institutions aiming to implement routine personalized dosimetry as part of RPT management.

Several limitations should be acknowledged. First, the dataset was small ($n = 10$ patients, 29 datasets) and originated from a single institution, increasing the risk of overfitting and potentially limiting generalizability to other imaging protocols and equipment. Second, absorbed dose was calculated using the MIRD schema via MIRDcalc, which assumes uniform dose distribution within the kidney; voxel-level dosimetry accounting for non-uniform activity and density has shown close agreement with organ-level MIRD calculation (median difference of 2.4% in kidney dose rate [21]). Third, pooling patients from two pharmacologically distinct groups (PSMA and DOTATATE) increased training data diversity but also introduced heterogeneity in renal pharmacokinetics and dose distribution. Future work should include multi-center validation with larger cohorts, prospective evaluation within a real-time dosimetry workflow, and integration of voxel-level dosimetry to better characterise intra-renal dose heterogeneity.

Conclusion

This study demonstrates that SwinUNETR with CT-only input achieves clinically acceptable automated kidney segmentation ($DSC = 0.91 \pm 0.07$) from multi-time-point Lu-177 SPECT/CT, enabling a mean absorbed dose estimation accuracy of +10.28% relative to expert segmentation and an 81–83% reduction in segmentation time. The incorporation of SPECT data as a second input channel significantly improves volumetric accuracy, which directly supports more precise dosimetry. These results support the clinical integration of 3D deep learning as a practical tool for patient-specific dosimetry in Lu-177 radiopharmaceutical therapy.

Declarations

Ethics approval and consent to participate: This study was approved by the Institutional Review Board of the Faculty of Medicine, Chulalongkorn University (IRB No. 0051/69). Patient consent was waived due to the retrospective nature of the study.

Availability of data and materials: The datasets used in this study are not publicly available due to patient privacy restrictions.

Authors' contributions: JC was responsible for data curation, model application, analysis, and writing of the original draft. ML contributed to conceptualization, supervision, and review and editing of the manuscript. KK contributed to conceptualization, clinical supervision, and review and editing of the manuscript. All authors read and approved the final manuscript.

Acknowledgements: The authors thank the nuclear medicine staff at King Chulalongkorn Memorial Hospital for assistance with data acquisition. This project is supported by 111th Anniversary Engineering Research Catalyst Fund Towards U Top 100, Faculty of Engineering, Chulalongkorn University.

References

- [1] Sgouros, G., L. Bodei, M.R. McDevitt, and J.R. Nedrow, *Radiopharmaceutical therapy in cancer: clinical advances and challenges*. Nature Reviews Drug Discovery, 2020. **19**(9): p. 589-608.
- [2] Strosberg, J., et al., *Phase 3 Trial of ¹⁷⁷Lu-Dotatate for Midgut Neuroendocrine Tumors*. N Engl J Med, 2017. **376**(2): p. 125-135.
- [3] George, S.C. and E.J.J. Samuel, *Developments in ¹⁷⁷Lu-based radiopharmaceutical therapy and dosimetry*. Front Chem, 2023. **11**: p. 1218670.
- [4] Leube, J., et al., *PSMA-PET improves deep learning-based automated CT kidney segmentation*. Z Med Phys, 2024. **34**(2): p. 231-241.
- [5] Wessels, B.W., et al., *MIRD Pamphlet No. 20: The Effect of Model Assumptions on Kidney Dosimetry and Response—Implications for Radionuclide Therapy*. Journal of Nuclear Medicine, 2008. **49**(11): p. 1884-1899.
- [6] Zanzonico, P., *The MIRD Schema for Radiopharmaceutical Dosimetry: A Review*. Journal of Nuclear Medicine Technology, 2024: p. jnmt.123.265668.
- [7] Spink, S., et al., *Estimation of kidney doses from*

- [¹⁷⁷Lu]Lu-DOTA-TATE PRRT using single time point post-treatment SPECT/CT. *EJNMMI Phys*, 2024. **11**(1): p. 68.
- [8] Nazari, M., et al., *Automated and robust organ segmentation for 3D-based internal dose calculation*. *EJNMMI Res*, 2021. **11**(1): p. 53.
- [9] Ronneberger, O., P. Fischer, and T. Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. 2015. Cham: Springer International Publishing.
- [10] Hatamizadeh, A., et al. *Unetr: Transformers for 3d medical image segmentation*. in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. 2022.
- [11] Hatamizadeh, A., et al. *Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images*. in *International MICCAI brainlesion workshop*. 2021. Springer.
- [12] Dewaraja, Y.K., et al., *A Pipeline for Automated Voxel Dosimetry: Application in Patients with Multi-SPECT/CT Imaging After ¹⁷⁷Lu-Peptide Receptor Radionuclide Therapy*. *J Nucl Med*, 2022. **63**(11): p. 1665-1672.
- [13] Ji, Y., et al., *A deep learning model for CT-based kidney volume determination in dogs and normal reference definition*. *Front Vet Sci*, 2022. **9**: p. 1011804.
- [14] Cardoso, M.J., et al., *Monai: An open-source framework for deep learning in healthcare*. arXiv preprint arXiv:2211.02701, 2022.
- [15] Oktay, O., et al., *Attention u-net: Learning where to look for the pancreas*. arXiv preprint arXiv:1804.03999, 2018.
- [16] Committee, M., *MIRDcalc User Manual v1.1-Genesis*. 2021: New York.
- [17] Kesner, A.L., et al., *MIRD Pamphlet No. 28, Part 1: MIRDcalc-A Software Tool for Medical Internal Radiation Dosimetry*. *J Nucl Med*, 2023. **64**(7): p. 1117-1124.
- [18] Sandström, M., et al., *Kidney dosimetry during ¹⁷⁷Lu-DOTATATE therapy in patients with neuroendocrine tumors: aspects on calculation and tolerance*. *Acta Oncol*, 2018. **57**(4): p. 516-521.
- [19] Guerriero, F., et al., *Kidney dosimetry in ¹⁷⁷Lu and ⁹⁰Y peptide receptor radionuclide therapy: influence of image timing, time-activity integration method, and risk factors*. *Biomed Res Int*, 2013. **2013**: p. 935351.
- [20] Topal, E., et al., *Impact of extended [¹⁷⁷Lu] Lu-PSMA-617 therapy on absorbed kidney dose and CKD-EPI values: how long can therapy be safely continued?* *Eur J Nucl Med Mol Imaging*, 2025. **52**(9): p. 3135-3144.
- [21] Dieudonné, A., et al., *Towards fully automatized [¹⁷⁷Lu]Lu-PSMA personalized dosimetry based on 360° CZT whole-body SPECT/CT: a proof-of-concept*. *EJNMMI Physics*, 2025. **12**(1): p. 25.
- [22] *MONAI: Medical Open Network for AI 2023*
- [23] *Pytorch: An imperative style, high-performance deep learning library 2023*