

Deep Learning-Based Whole-Body Lesion Segmentation and Automated OMIS Computation on [^{68}Ga]Ga-DOTA-TOC PET/CT: Technical Feasibility of a SwinUNETR Pipeline for Pre-PRRT Bone Marrow Involvement Scoring

Romain Ferrara

`ferrara@ipc.unicancer.fr`

Department of Nuclear Medicine, Institut Paoli-Calmettes, ENETS Center of Excellence, IPC NET Center, Marseille, France <https://orcid.org/0009-0009-3227-7905>

Solène Malmon

Pixilib

Salim Kanoun

Pixilib

Joris Cocquebert

Department of Nuclear Medicine, Institut Paoli-Calmettes, ENETS Center of Excellence, IPC NET Center, Marseille, France

Thibaut Reichert

Department of Nuclear Medicine, Institut Paoli-Calmettes, ENETS Center of Excellence, IPC NET Center, Marseille, France

Adil Moussali

Department of Nuclear Medicine, Institut Paoli-Calmettes, ENETS Center of Excellence, IPC NET Center, Marseille, France

Nathalie Charrier

Department of Nuclear Medicine, Institut Paoli-Calmettes, ENETS Center of Excellence, IPC NET Center, Marseille, France

Sébastien Benzekry

Computational Pharmacology and Clinical Oncology (COMPO) Unit, Inria Sophia Antipolis-Méditerranée, Cancer Research Center of Marseille, Inserm UMR1068, CNRS UMR7258, Aix Marseille University UM105, 13385 Marseille

Daniel Ouk

Department of Nuclear Medicine, Institut Paoli-Calmettes, ENETS Center of Excellence, IPC NET Center, Marseille, France

Sandrine Oziel-Taieb

Department of Medical Oncology, Institut Paoli-Calmettes, ENETS Center of Excellence, IPC NET Center, Marseille, France

Research Article

Keywords: [68Ga]Ga-DOTA-TOC, [177Lu]Lu-DOTA-TATE, neuroendocrine tumour, SwinUNETR, OMIS, TMTV, deep learning, haematotoxicity, PRRT

Posted Date: July 3rd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10101940/v1>

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

[Read Full License](#)

Abstract

Background

Patients with neuroendocrine tumours undergoing [^{177}Lu]Lu-DOTA-TATE PRRT are at risk of haematological toxicity, and the Osteo-Medullary Invasion Score (OMIS) — a weighted bone-marrow involvement score based on Cristy's somatotope — has been reported to predict this risk but requires labour-intensive manual whole-body contouring. We assessed the technical feasibility of a SwinUNETR-based deep learning pipeline for automated whole-body [^{68}Ga]Ga-DOTA-TOC PET/CT lesion segmentation, total metabolic tumour volume (TMTV) quantification, and automated OMIS computation. A retrospective single-centre cohort of 115 patients (188 examinations; mean age 65.5 ± 11.7 years; 52.2% female) was split at the patient level into training ($n = 150$), validation ($n = 19$), and test ($n = 19$) sets. Three architectures were evaluated across four training configurations: Attention U-Net (scratch and FDG-pretrained), nnU-Net, and SwinUNETR. OMIS was computed automatically from predicted lesion masks and TotalSegmentator-derived skeletal anatomy.

Results

On the independent test set ($n = 19$), SwinUNETR achieved a Dice similarity coefficient (DSC) of 0.806 ± 0.123 , significantly outperforming the FDG-pretrained Attention U-Net (DSC 0.579 ± 0.189 ; Wilcoxon $p < 0.0001$) and comparable to nnU-Net (DSC 0.783 ± 0.181 ; $p = 0.44$), with strong biomarker agreement (TMTV $r = 0.997$, OMIS $r = 0.998$; both $p < 0.0001$); for reference, expert inter-observer agreement in this cohort was DSC 0.715. Both OMIS $\geq 30\%$ cases on the test set were correctly classified (2/2 true positives, 17/17 true negatives). Across all partitions including training data ($n = 188$), sensitivity was 90.2% and specificity 99.3%; the four false negatives all involved diffuse small-volume skeletal disease below the effective detection threshold. The skeletal-attribution component was externally validated on the public ENHANCE.PET 1.6k dataset (whole-skeleton Dice 0.913).

Conclusions

SwinUNETR demonstrates technical feasibility of automated whole-body lesion segmentation and OMIS computation on [^{68}Ga]Ga-DOTA-TOC PET/CT with performance approaching inter-expert agreement. Given the previously reported predictive value of OMIS for haematological toxicity after [^{177}Lu]Lu-DOTA-TATE PRRT, this pipeline offers a reproducible approach to automated OMIS quantification. These preliminary results support prospective multicentre validation with haematological outcome data before clinical deployment as a pre-therapeutic risk stratification tool.

Introduction

Neuroendocrine tumours (NETs) are a heterogeneous family of neoplasms arising from neuroendocrine cells, predominantly in the gastroenteropancreatic (GEP) tract and lungs. Their incidence has increased substantially over the past two decades, with current age-adjusted incidence rates of 6–7 per 100 000 person-years in North America and increasing trends across Europe [1, 2]. Although often indolent, metastatic NETs, present in 40–50% of patients at diagnosis, are associated with prolonged survival, frequently measured in years, which amplifies the clinical significance of treatment-related toxicity [1, 2].

[⁶⁸Ga]Ga-DOTA-TOC PET/CT exploits overexpression of somatostatin receptor subtype 2 (SSTR2) to deliver whole-body tumour staging with high sensitivity and specificity [3, 4]. Its theranostic counterpart, [¹⁷⁷Lu]Lu-DOTA-TATE (Lutathera®), delivers β^- radiation selectively to SSTR2-expressing lesions and has demonstrated superiority over high-dose octreotide in the landmark NETTER-1 trial [5]. With four intravenous cycles of 7.4 GBq every 8 weeks now constituting the standard protocol, PRRT has become a cornerstone of GEP-NET management [5, 6].

Despite its efficacy, PRRT carries a significant risk of haematological toxicity. Bone marrow receives inadvertent irradiation via direct uptake of [¹⁷⁷Lu]Lu-DOTA-TATE in somatostatin receptor-expressing skeletal metastases, via circulating unbound radiopharmaceutical, and via direct binding to SSTR2-expressing haematopoietic stem and progenitor cells, which have been shown to take up SSTR2 ligands at levels comparable to those of neuroendocrine tumour cells [7, 8]. Up to 10% of patients develop grade 3/4 haematological toxicity, with thrombocytopenia being the most frequent cause of dose reduction or treatment interruption [9, 10]. More critically, persistent thrombocytopenia (PT) after PRRT (defined as platelet count $< 100 \times 10^9/L$ two months after treatment completion) carries a 2% risk of myelodysplastic syndrome and 1% risk of acute myeloid leukaemia, malignancies almost invariably preceded by PT [10–13]. These risks are amplified in the era of PRRT escalation to earlier lines of therapy (NETTER-2 [6]), in which cumulative haematological exposure begins earlier in the disease course.

We previously introduced the Osteo-Medullary Invasion Score (OMIS) [12, 14], quantifying the proportion of haematopoietically active bone marrow invaded by NET metastases, by weighting lesion-positive skeletal regions according to their contribution to total-body haematopoiesis as mapped by the Cristy model [15]. Specifically, observable bone lesions are identified on somatostatin receptor imaging; each affected bone receives a weight proportional to its share of active marrow in a 40-year reference adult, and OMIS is the sum of weights over all affected bones, expressed as a percentage [14]. In the largest validation cohort to date ($n = 47$ patients treated with [¹⁷⁷Lu]Lu-DOTA-TATE), an OMIS $\geq 30\%$ predicted PT with an odds ratio of 124 ($p < 0.001$), and 100% of patients with OMIS $\geq 30\%$ developed PT versus 7.1% with OMIS $< 30\%$ [12]. However, the manual computation of OMIS requires expert whole-body lesion contouring, a time-consuming and highly operator-dependent process that is impractical in routine clinical workflow.

Deep learning has transformed segmentation of tumour burden on PET/CT [16, 17]. Transformer-based architectures, and particularly the Shifted-Window UNet TRansformer (SwinUNETR) [18, 19], built on the Swin Transformer [20], overcome the limited receptive field of convolutional networks by modelling long-

range spatial dependencies through hierarchical self-attention. This is especially relevant for whole-body [^{68}Ga]Ga-DOTA-TOC PET/CT, where tumour lesions may span the entire body in a heterogeneous, multi-focal pattern. Yazdani et al. demonstrated the superiority of SwinUNETR with self-supervised pre-training over nnU-Net on [^{68}Ga]Ga-PSMA-11 PET/CT (+ 7% Dice similarity coefficient (DSC), lesion DSC 0.68 vs 0.61) [21].

To situate the selected architecture within an interpretable comparison, four training configurations spanning three architectures were evaluated. A 3-D Attention U-Net trained from random initialisation served as a lower-bound convolutional baseline, quantifying the difficulty of learning whole-body segmentation from a limited labelled cohort without prior knowledge. The same architecture initialised with [^{18}F]FDG PET/CT pre-trained encoder weights tested whether transfer learning from a related whole-body PET task could compensate for data scarcity. SwinUNETR, trained from scratch, was the candidate architecture of interest, hypothesised to capture the globally distributed multi-focal lesion pattern of metastatic NETs through hierarchical self-attention. Finally, nnU-Net was included as a strong self-configuring convolutional reference, representing the current de facto standard for 3-D biomedical segmentation, to benchmark whether a transformer architecture offers an advantage over an optimised convolutional pipeline on this task.

We hypothesised that a SwinUNETR transformer architecture, by virtue of its ability to model long-range spatial dependencies through hierarchical self-attention, would outperform convolutional Attention U-Net architectures on the challenging task of whole-body [^{68}Ga]Ga-DOTA-TOC PET/CT segmentation. [^{18}F]FDG PET/CT pre-trained weights were available for the Attention U-Net from prior work [22]. The aims of this study were to: (i) develop a SwinUNETR model trained from scratch on whole-body [^{68}Ga]Ga-DOTA-TOC PET/CT and compare it against Attention U-Net with and without [^{18}F]FDG pre-training [22, 23] and nnU-Net [24] as a self-configuring reference; (ii) demonstrate the feasibility of automated OMIS computation by combining lesion segmentation with TotalSegmentator-derived skeletal compartment mapping, and assess its agreement with the manual reference at the clinically established $\geq 30\%$ threshold; and (iii) report the full suite of TMTV and bone-TMTV biomarker correlations on an independent test set, with all-partition values provided for technical consistency. The complete pipeline is illustrated in Fig. 1.

Materials and Methods

Study population and imaging protocol

This retrospective single-centre study was approved by the institutional ethics committee (approval number: PIXIRIV-IPC 2023-034; informed consent waiver). All consecutive patients with histologically confirmed NET who underwent [^{68}Ga]Ga-DOTA-TOC PET/CT at Institut Paoli-Calmettes, Marseille, between 8 March 2022 and 7 September 2024 were eligible. Inclusion criteria were: (1) histologically confirmed NET; (2) complete PET/CT acquisition from skull vertex to mid-thigh; (3) availability of full

DICOM images. Patients with incomplete acquisitions or inadequate image quality were excluded. A total of 115 patients contributing 188 examinations were included (Table 1).

[⁶⁸Ga]Ga-DOTA-TOC was administered intravenously at 153 ± 39 MBq. PET was acquired at a median of 66 minutes post-injection (mean 68 ± 11 min; range 48-103 min) on a Siemens Biograph Vision 450 (matrix 192 × 192; voxel size 4.07 × 4.07 × 4.07 mm³; axial FOV skull vertex to mid-thigh). Images were reconstructed using PSF+TOF iterative reconstruction (3 iterations, 5 subsets) with XYZ Gaussian post-filter (FWHM 4.0 mm). Attenuation correction was CT-based (HD CdV protocol; iMAR metal artefact reduction applied in 14/188 acquisitions). Scatter correction was model-based. PET data were converted to standardized uptake value (SUV) maps normalised to body weight. Low-dose CT was acquired in the same session for attenuation correction and anatomical co-registration.

Table 1. Patient and examination characteristics (n=115 patients; 188 examinations).

Characteristic	n (%) or mean $\pm$ SD
Patients (n)	115
Examinations (n)	188
Age, years	65.5 $\pm$ 11.7 (range 26-89)
Female sex	60 (52.2%)
Primary tumour site	
Small bowel	60 (52.2%)
Lung	17 (14.8%)
Pancreas	12 (10.4%)
Unknown primary	6 (5.2%)
Other	20 (17.4%)
Metastatic disease	
Any metastases	110 (95.7%)
Liver metastases	90 (81.8%)
Lymph node metastases	85 (77.3%)
Bone metastases	64 (58.2%)
Peritoneal metastases	22 (20.0%)
Dataset partition	
Training (examinations / patients)	150 / 85
Validation (examinations / patients)	19 / 11
Test (examinations / patients)	19 / 19

Reference standard: manual segmentation and OMIS

All 188 examinations were manually delineated by two nuclear medicine physicians with ≥ 5 years of [^{68}Ga]Ga-DOTA-TOC PET/CT reporting experience, using 3D Slicer 5.2. Lesions were defined as focal SSTR-avid foci exceeding local background activity that were confirmed pathological by nuclear

medicine consensus. A semi-automated threshold approach (40-50% of local SUVmax, adjusted manually per lesion) was applied to initialise contours, followed by voxel-by-voxel review in all three planes. Discrepancies were resolved by consensus.

Annotation quality and inter-observer variability. Inter-observer variability of the manual reference standard was assessed on n=38 examinations independently contoured by two nuclear medicine physicians (T.R. and D.O.). Mean DSC was 0.715 ± 0.169 (95% CI 0.663–0.769; median 0.718), with a TMTV Pearson correlation of $r=0.958$ (95% CI 0.919–0.978, $p<0.0001$) and a Bland–Altman bias of +13.7 mL (SD 35.6 mL; limits of agreement (LoA) –56.0 to +83.5 mL), indicating good inter-observer reproducibility of the reference segmentation.

OMIS was computed from manual whole-body lesion annotations on [^{68}Ga]Ga-DOTA-TOC PET/CT, using the Cristy bone-marrow somatotope [15], as previously described [12,14]. Briefly, TotalSegmentator [17] was applied to the CT component to delineate 62 individual skeletal structures: skull, sternum, sacrum, 25 vertebrae (C1-C7, T1-T12, L1-L5, S1), both humeri, both scapulae, both clavicles, both femora, both ilia, and 24 ribs (12 per side). Each structure was assigned a weight corresponding to its fractional contribution to active bone-marrow haematopoiesis in a 40-year reference adult per Cristy [15] (e.g., two invaded lumbar vertebrae yield an OMIS contribution of $\approx 5\%$, as the five lumbar vertebrae together represent 12.3% of active marrow, i.e. $\approx 2.5\%$ per vertebra [12]). A bone was scored as "invaded" if any manually delineated lesion voxel was present within that structure. OMIS was the sum of weights over all invaded bones, expressed as a percentage (range 0-100%). The clinically validated haematotoxicity risk threshold is $\text{OMIS} \geq 30\%$ [12].

Deep learning architectures and training configurations

Three architectures were evaluated across four training configurations: Attention U-Net (two configurations: scratch and FDG-pretrained), nnU-Net, and SwinUNETR. All architectures took dual-channel PET/CT input (in_channels=2, out_channels=1). Attention U-Net and SwinUNETR were implemented in MONAI 1.0 [25] and PyTorch 2.0; nnU-Net was trained using its native self-configuring pipeline (nnUNetv2). All models were trained on a single NVIDIA L40S 48 GB GPU. Attention U-Net and SwinUNETR shared the same data augmentation pipeline: random rotation ($\pm 30^\circ$), random flipping, and random intensity scaling.

Architecture 1, Attention U-Net:

No pre-training: 3-D Attention U-Net [23] (5 resolution levels; dropout 0.4) trained from random initialisation on the [^{68}Ga]Ga-DOTA-TOC dataset as a lower-bound baseline. Adam optimiser ($\text{lr}=1 \times 10^{-4}$, $\text{weight_decay}=1 \times 10^{-4}$), ReduceLROnPlateau scheduler (patience=10, factor=0.3), up to 200 epochs, Tversky loss [26] ($\alpha=0.3$, $\beta=0.7$).

FDG-pretrained: Identical architecture, encoder weights initialised from a model pre-trained on a [18F]FDG PET/CT whole-body segmentation dataset by Malmon et al. [22] where it was used for automated TMTV calculation in lymphomas. Fine-tuned on the [⁶⁸Ga]Ga-DOTA-TOC dataset with a lower learning rate (1×10^{-5}), ReduceLROnPlateau scheduler (patience=10, factor=0.3), up to 200 epochs, and Tversky loss ($\alpha=0.3$, $\beta=0.7$).

Architecture 2, nnU-Net: nnU-Net [24] with self-configuring 3D full-resolution configuration (PlainConvUNet; 6 stages; patch size 192×112×112 voxels; isotropic 1.65 mm spacing; ZScore normalisation for PET, CT normalisation for CT). SGD optimiser (initial lr=0.01, momentum=0.99, Nesterov), polynomial learning rate scheduler, deep supervision DSC+CE loss. Trained with the standard nnU-Net pipeline (fold_all) with default early stopping, on the same training set and patient-level split as the other architectures.

Architecture 3, SwinUNETR (selected model): 3-D SwinUNETR [18,19] with Swin Transformer encoder (4 stages; window size 7; patch size 2; feature size 24; depths (2,2,2,2); heads (3,6,12,24)) and CNN decoder with instance normalisation and skip connections. The model was trained from random weight initialisation, with no external pre-trained weights. Target volume: 128 × 128 × 256 voxels at isotropic 4.0 mm spacing.

Training details for SwinUNETR: Adam optimiser (lr= 1×10^{-4} , weight decay= 1×10^{-4}), ReduceLROnPlateau scheduler (patience=10, factor=0.3), 200 epochs, Tversky loss ($\alpha=0.3$, $\beta=0.7$).

Dataset partitioning

The 188 examinations from 115 patients were partitioned at the patient level into training (150 examinations, 85 patients), validation (19 examinations, 11 patients), and test (19 examinations, 19 patients) sets. Patient-level partitioning was enforced to prevent data leakage from longitudinal acquisitions of the same patient across subsets. Architecture selection was performed on the validation set; the test set was used exclusively for final performance assessment. Patient-level grouped k-fold cross-validation was considered but not implemented owing to the low prevalence of OMIS $\geq 30\%$ cases (41/188 examinations, 22%), which would yield unstable binary classification estimates per fold given the small absolute number of positive cases per partition.

Evaluation metrics and OMIS computation from predictions

Two distinct segmentation performance frameworks were applied:

(1) Voxel-level and lesion-level segmentation performance (n=19 test) : DSC [27], voxel sensitivity, precision (PPV), and lesion-level sensitivity, PPV, and F1-score comparing predicted masks to manual

reference across full segmentation masks. Prior to lesion-level metric computation, both predicted and ground-truth lesion masks were processed with a watershed transform to separate adjacent lesion instances. Lesion-level metrics required $\geq 10\%$ overlap for a predicted lesion to count as a true positive. Lesion-size categories: tiny (<1 mL), small (1-5 mL), medium (5-10 mL), large (>10 mL).

(2) OMIS biomarker evaluation (n=19 test, n=188 all examinations) : Pearson correlation and Bland–Altman analysis between automated and manual TMTV, bone TMTV (TMTV_bone), and OMIS continuous scores. For binary OMIS classification at the $\geq 30\%$ threshold, sensitivity, specificity, accuracy, PPV, and NPV were computed.

Prior to all volumetric computations, predicted and ground-truth lesion masks were thresholded at 30% of the local lesion SUVmax per lesion, retaining only metabolically active voxels above this threshold; total TMTV was defined as the volume of voxels passing this threshold. Automated OMIS was computed identically to the manual reference: TotalSegmentator-derived bone masks were intersected with the thresholded lesion masks, and Cristy weights were applied per bone as described above. Bone TMTV was the overlap between thresholded predicted lesions and TotalSegmentator-derived skeletal structures.

External validation of skeletal segmentation

Because OMIS depends on the correct assignment of lesions to skeletal compartments, the reliability of the TotalSegmentator-derived bone masks was evaluated independently of the PET lesion-segmentation model. As no manual skeletal reference was available in the present cohort, TotalSegmentator – run with the identical configuration used in the OMIS pipeline – was applied to an independent subset of 30 whole-body PET/CT studies drawn from the publicly available ENHANCE.PET 1.6k dataset [34], which provides expert-verified CT-derived segmentations of skeletal structures generated by an independent pipeline. Per-compartment and whole-skeleton Dice and average symmetric surface distance (ASSD) were computed between the TotalSegmentator output and the reference standard. The cranium and the S1 segment, which lack a one-to-one correspondence in the ENHANCE.PET label scheme, were excluded from this comparison.

Statistical analysis

Continuous data are reported as mean $\pm$ standard deviation or median (range). Architecture comparison on the test set was performed with the Wilcoxon signed-rank test (two-tailed; significance $\alpha=0.05$). Pearson correlation coefficients are reported with 95% confidence intervals; Bland–Altman plots illustrate systematic bias and limits of agreement (LoA) for TMTV and OMIS. All analyses were performed in Python 3.10 (SciPy 1.10, NumPy 1.24) and R 4.3.

Results

Patient and examination characteristics

The cohort comprised 115 patients (65.5 ± 11.7 years; 52.2% female) contributing 188 [^{68}Ga]Ga-DOTA-TOC PET/CT examinations (Table 1). The most frequent primary site was small bowel ($n = 60$; 52.2%), followed by lung ($n = 17$; 14.8%) and pancreas ($n = 12$; 10.4%). Metastatic disease was present in 110/115 patients (95.7%), with liver (81.8%), lymph node (77.3%), and bone (58.2%) involvement predominating. The median number of examinations per patient was 1 (range 1–5).

Architecture comparison

On the test set, the non-pretrained Attention U-Net yielded a mean DSC of 0.37, the lowest performance among the four configurations on the [^{68}Ga]Ga-DOTA-TOC dataset. FDG-based encoder pre-training improved the DSC to 0.579 ± 0.189 . nnU-Net achieved a mean DSC of 0.783 ± 0.181 (median 0.850; TMTV $r = 0.995$; TMTV MAE 26.6%, median 7.2%). SwinUNETR, trained from random initialisation, achieved the highest DSC (Table 2).

Table 2
Architecture comparison on the independent test set ($n = 19$).

Architecture	DSC	TMTV MAE%
Attention U-Net (scratch)	0.37	-
Attention U-Net (FDG-pretrained)	0.579 ± 0.189	33.1%*
nnU-Net (reference)	0.783 ± 0.181	26.6%*
SwinUNETR (selected)	0.806 ± 0.123	21.7%*

* Mean absolute percentage error ($n = 19$). Wilcoxon test SwinUNETR vs. Attention U-Net (FDG-pretrained): $p < 0.0001$; SwinUNETR vs. nnU-Net: $p = 0.44$ (ns). TMTV MAE not computed for the non-pretrained Attention U-Net.

SwinUNETR segmentation performance (test set, $n = 19$)

On the independent test set, SwinUNETR achieved a mean DSC of 0.806 ± 0.123 (median 0.804; range 0.524–0.990), voxel sensitivity 0.807 ± 0.115 , and PPV 0.831 ± 0.177 (Table 3). Lesion-level sensitivity was 0.666 ± 0.184 , PPV 0.745 ± 0.249 , and F1 0.652 ± 0.158 . A size-stratified failure analysis confirmed a strong relationship between lesion volume and detectability: sensitivity was 0.451 for tiny lesions (< 1 mL, $n = 699$ across 16 test patients), 0.799 for small (1–5 mL, $n = 272$), 0.986 for medium (5–10 mL, $n = 56$), and 1.000 for large (> 10 mL, $n = 69$). Detection failures were therefore almost entirely confined to the tiny-lesion category: missed lesions were predominantly sub-centimetre lymph node, peritoneal, or osseous deposits at or below the effective spatial resolution of the system. False-positive detections were uncommon and arose mainly in regions of high physiological uptake (liver, spleen, kidneys, and bowel), consistent with the competing-signal challenge characteristic of [^{68}Ga]Ga-DOTA-TOC PET. Representative success and failure cases are shown in Fig. 5.

At the patient level, segmentation quality was not driven by overall tumour burden: per-patient DSC was uncorrelated with reference TMTV (Pearson $r = 0.14$, $p = 0.57$). Instead, performance variability decreased as tumour burden increased, with DSC rising from 0.789 ± 0.189 in the lowest TMTV tercile (5–90 mL) to 0.801 ± 0.082 (110–221 mL) and 0.832 ± 0.056 in the highest tercile (314–2132 mL). The two lowest-DSC cases (0.524 and 0.559) were both small-burden examinations (22 mL and 5 mL of reference TMTV), in which a small absolute number of mis-segmented voxels produced a large relative penalty; volumetric error was correspondingly unstable for sub-30 mL cases (median TMTV error 36%, up to 86%) while remaining low and stable for larger burdens. The network was therefore most reliable for the clinically significant disease burdens that drive OMIS, and least reliable for very low-burden examinations – a regime in which the absolute skeletal volume involved, and hence the impact on the OMIS score, is intrinsically limited.

Table 3

Comprehensive SwinUNETR performance metrics on the independent test set ($n = 19$ examinations). All values mean $\pm$ SD unless stated.

Metric	Value	N / definition	95% CI
DSC	0.806 ± 0.123 (median 0.804)	$n = 19$; manual mask vs predicted mask	[0.751–0.857]
Voxel sensitivity	0.807 ± 0.115	$n = 19$	[0.756–0.858]
PPV (precision)	0.831 ± 0.177	$n = 19$	[0.749–0.904]
Lesion sensitivity	0.666 ± 0.184	$n = 19$; lesion-level TP/(TP + FN)	[0.584–0.747]
Lesion PPV	0.745 ± 0.249	$n = 19$; lesion-level TP/(TP + FP)	[0.628–0.852]
Lesion F1	0.652 ± 0.158	$n = 19$	[0.581–0.719]
Sensitivity by lesion size			
Tiny < 1 mL ($n = 699$ total lesions)	0.451	$n = 16$ patients with tiny lesions	-
Small 1–5 mL ($n = 272$)	0.799	$n = 18$ patients	-
Medium 5–10 mL ($n = 56$)	0.986	$n = 14$ patients	-
Large > 10 mL ($n = 69$)	1.000	$n = 16$ patients	-
TMTV quantification			
TMTV MAE	21.7% (median 15.2%)	$n = 19$; $ pred - GT /GT \times 100$	-
Pearson r	0.997 ($p < 0.0001$)	$n = 19$	[0.993–0.999]

Bootstrap 95% confidence intervals (5 000 resamples, n = 19).

The 95th-percentile Hausdorff distance (HD95) was not reported as it is not interpretable in the whole-body multi-lesion setting, where inter-lesion distances dominate over contour precision; lesion-level sensitivity, PPV and F1 were used instead.

TMTV and OMIS quantification

On the independent test set (n = 19), automated biomarkers showed strong agreement with manual references (Table 4). OMIS Pearson r = 0.998 (R²=0.996; p < 0.0001), Bland–Altman bias + 0.05 percentage points (pp) (SD 1.71 pp; LoA – 3.3 to + 3.4 pp), MAE 0.79 pp. TMTV Pearson r = 0.997 (p < 0.0001), bias – 32.4 mL (SD 77.0 mL; LoA – 183.3 to + 118.4 mL).

For completeness, correlations across all 188 examinations (train + val+test) were: TMTV r = 0.952 (R²=0.906), Bone TMTV r = 0.882 (R²=0.779), OMIS r = 0.956 (R²=0.914); Bland–Altman for TMTV (n = 188): bias – 36.4 mL (SD 124.6 mL; LoA – 280.7 to + 207.9 mL), and for OMIS (n = 188): bias – 2.1 pp (SD 8.5 pp; LoA – 18.8 to + 14.6 pp). Bland–Altman plot for TMTV is shown in Fig. 2; the OMIS scatter plot (automated vs. manual) is shown in Fig. 3; and the Bland–Altman plot for OMIS is shown in Fig. 4.

On the independent ENHANCE.PET subset (n = 30), TotalSegmentator reproduced the expert-verified skeletal reference with high accuracy and low variability: whole-skeleton Dice was 0.913 ± 0.007 (mean ASSD 0.68 mm) and the red-marrow-bearing compartment most relevant to OMIS (spine, pelvis, ribs) reached Dice 0.899 ± 0.007 (ASSD 0.61 mm). Per-compartment Dice ranged from 0.816 ± 0.018 for the ribs – the thinnest structures, where volumetric overlap is intrinsically penalised despite sub-millimetre-to-millimetric surface agreement – to 0.967 ± 0.008 for the femur, with spine 0.918 ± 0.010, pelvis 0.922 ± 0.006, and proximal upper limb 0.907 ± 0.013. The narrow standard deviations indicate consistent skeletal labelling across studies, supporting the robustness of the anatomical compartment-assignment layer on PET/CT-derived low-dose CT.

Table 4

Pearson correlations between automated (SwinUNETR) and manual biomarkers. Primary analysis on the independent test set (n = 19 examinations); all-dataset values (n = 188, train + val+test) provided for completeness.

Biomarker	N examinations	Pearson r	R ²	p-value
TMTV	19 (test) / 188 (all)	0.997 (test, n = 19) / 0.952 (all, n = 188)	0.994 (test) / 0.906 (all)	p < 0.0001
Bone TMTV	19 (test) / 188 (all)	0.989 (test, n = 19) / 0.882 (all, n = 188)	0.978 (test) / 0.779 (all)	p < 0.0001
OMIS	19 (test) / 188 (all)	0.998 (test, n = 19) / 0.956 (all, n = 188)	0.996 (test) / 0.914 (all)	p < 0.0001

OMIS-based haematotoxicity risk classification

For binary classification at OMIS $\geq 30\%$, the threshold previously validated as predictive of persistent thrombocytopenia (OR = 124 [12]), on the independent test set (n = 19) both OMIS $\geq 30\%$ cases were correctly classified (2/2 true positives, 17/17 true negatives; sensitivity 100%, specificity 100%; Table 5). Given the small number of positive cases on the test set, classification performance is additionally reported across all partitions including training data (n = 188): sensitivity 90.2% (37/41 true positives), specificity 99.3% (146/147 true negatives), accuracy 97.3%, PPV 97.4%, NPV 97.3% (Table 5). The four false negatives all involved diffuse small-volume skeletal disease below the effective detection threshold.

Table 5
Confusion matrix for binary OMIS classification at the $\geq 30\%$ threshold. Primary analysis: independent test set (n = 19 examinations). Secondary analysis across all partitions including training data (n = 188 examinations) provided for completeness.

	Primary - Test set (n = 19)			Secondary - All partitions (n = 188)		
	Pred < 30%	Pred $\geq 30\%$	Total	Pred < 30%	Pred $\geq 30\%$	Total
True < 30%	TN = 17	FP = 0	17	TN = 146	FP = 1	147
True $\geq 30\%$	FN = 0	TP = 2	2	FN = 4	TP = 37	41
Total	17	2	19	150	38	188

Primary (test, n = 19): Sens. 100% (2/2), Spec. 100% (17/17), PPV 100%, NPV 100%. **Secondary (all partitions, n = 188):** Sens. 90.2%, Spec. 99.3%, PPV 97.4%, NPV 97.3%, Accuracy 97.3%.

All four false negatives involved extensive small-volume skeletal disease (whether as numerous discrete sub-centimetre lesions or diffuse bone marrow infiltration) distributed across multiple bone segments (Fig. 5). The single false positive was a borderline case (true OMIS 29.0%, predicted 34.0%).

Discussion

This study demonstrates that automated whole-body lesion segmentation on [⁶⁸Ga]Ga-DOTA-TOC PET/CT using SwinUNETR is feasible and yields OMIS estimates in close agreement with the manual reference, supporting the technical feasibility of automated OMIS computation before [¹⁷⁷Lu]Lu-DOTA-TATE PRRT. The pipeline achieved DSC 0.806 ± 0.123 and OMIS $r = 0.998$ on the independent test set; both OMIS $\geq 30\%$ cases were correctly classified on the test set (2/2 true positives, 17/17 true negatives), though the very small number of positive cases precludes definitive conclusions on sensitivity from this set alone. The central contribution of this work is therefore not the segmentation architecture per se, but the demonstration that the full OMIS computation – which couples whole-body lesion segmentation with skeletal compartment mapping and Cristy bone-marrow weighting – can be automated end-to-end, replacing a labour-intensive manual procedure with a reproducible pipeline.

SwinUNETR substantially outperformed both Attention U-Net architectures. The non-pretrained Attention U-Net (DSC 0.37) highlighted the inadequacy of training complex volumetric networks from scratch on limited labelled data in a rare-disease setting. FDG pre-training improved performance (DSC 0.579), confirming that cross-tracer transfer learning provides useful prior representations for whole-body PET geometry. However, the performance gain was insufficient: the domain gap between [^{18}F]FDG and [^{68}Ga]Ga-DOTA-TOC, in particular the high physiological uptake in liver, spleen, and kidneys that creates strong competing signals absent in FDG imaging, limits transferability of FDG-derived encoder representations. SwinUNETR, trained from scratch, overcame this limitation through its shifted-window self-attention mechanism, which models long-range spatial dependencies across the entire body volume, capturing the globally distributed multi-focal lesion pattern characteristic of metastatic NETs. The margin of improvement (DSC 0.806 vs. 0.579; Wilcoxon $p < 0.0001$) suggests that architectural expressivity can compensate for the absence of pre-training when the model is sufficiently expressive. The asymmetric Tversky loss ($\alpha = 0.3$, $\beta = 0.7$) was chosen to reflect the clinical cost asymmetry of OMIS-based risk stratification: a false negative (missed lesion, OMIS underestimation) risks misclassifying a high-risk patient as low-risk before PRRT, whereas a false positive leads to conservative management (the less dangerous error). Penalising false negatives more heavily thus aligns the training objective with the downstream clinical task.

nnU-Net, included as a self-configuring convolutional reference, achieved comparable overall performance (DSC 0.783 ± 0.181 ; Wilcoxon vs. SwinUNETR $p = 0.44$). The two architectures are therefore broadly equivalent from a pure segmentation standpoint. SwinUNETR was nonetheless selected as the deployed model for two reasons. First, it showed greater consistency across challenging cases (DSC range 0.524–0.990 vs. 0.33–0.97 for nnU-Net), which is clinically important given the heterogeneity of NET disease burden. Second, in our local clinical research environment (Gaeloflow PACS), nnU-Net's self-configuring pipeline – which generates preprocessing configurations and model weights tightly coupled to training-time data statistics – presented greater complexity for export of a standalone DICOM-compatible inference container compared with MONAI Deploy-based architectures [28]. Compared with published benchmarks, our DSC of 0.806 is competitive: prior studies on NET somatostatin receptor PET segmentation report DSC of 0.64 on [^{68}Ga]Ga-DOTATATE [29] and 0.80 on [^{64}Cu]Cu-DOTATATE [30], while a model trained on pheochromocytoma and paraganglioma (PPGL) achieved DSC 0.60 when applied to a small NET cohort [31] whereas Ciarmiello et al. [32] reported DSC 0.879 with nnU-Net on [^{18}F]FDG PET/CT across three tumour types and a large dataset ($n = 1334$), and Yazdani et al. [21] DSC 0.68 on PSMA PET/CT with SwinUNETR using self-supervised pre-training on 652 unlabelled images.

The integration of PET lesion segmentation with CT-based skeletal anatomy from TotalSegmentator enables automatic labelling of each lesion to its anatomical site, opening the way to bone-specific dosimetry and organ-level disease tracking over serial examinations. These capabilities are beyond the reach of volume-only segmentation methods.

The feasibility of fully automated OMIS computation is the key clinical contribution of this work. OMIS quantification has historically been performed manually, limiting its deployment in routine PRRT practice.

On the independent test set, automated classification at OMIS $\geq 30\%$ agreed with the manual reference in all 19 cases, although the small number of positive cases ($n = 2$) precludes robust estimation of sensitivity; across all partitions including training data ($n = 188$, reported here only as a consistency check), sensitivity was 90.2% and specificity 99.3%. These preliminary results suggest that automated OMIS computation could, after prospective validation, support pre-PRRT haematotoxicity risk assessment. The four false-negative cases all involved extensive small-volume skeletal disease below the network's effective detection threshold at 4 mm isotropic resolution, consistent with the known limitation of somatostatin receptor PET for small lesions [3], and clinicians should apply additional scrutiny in patients presenting with diffuse low-grade bone marrow signal on pre-PRRT DOTA-TOC PET/CT.

Beyond risk stratification, automated OMIS computation could facilitate prospective patient selection for PRRT by making the score accessible at centres without dedicated nuclear medicine physics support. While the present results are preliminary and require prospective validation on a larger number of OMIS-positive cases, an automated, reproducible score could in principle prompt earlier consideration of prophylactic dose reduction, enhanced haematological monitoring, or multidisciplinary discussion before treatment initiation. Beyond OMIS, the pipeline also provides splenic volume and spleen length from TotalSegmentator-derived CT segmentation, which are candidate predictors of haematological toxicity during PRRT [12, 33] and will be included in future multivariate modelling.

Several limitations must be acknowledged. The study is retrospective and single-centre, performed on a single scanner model (Siemens Biograph Vision 450), which limits generalisability across acquisition protocols and reconstruction parameters. The independent test set comprises 19 examinations, a sample size inherent to rare-disease AI development but insufficient for stable estimation of binary classification metrics; in particular, the perfect OMIS classification on the test set rests on only two OMIS-positive and seventeen OMIS-negative cases, and should be interpreted as an encouraging preliminary result rather than robust evidence of clinical risk-stratification performance. The all-partition analysis ($n = 188$) includes training data and is therefore reported only as a technical consistency check, not as independent validation. The pipeline relies on TotalSegmentator for skeletal segmentation; while its performance was not benchmarked on the present low-dose, non-contrast PET/CT cohort directly (no manual skeletal reference being available), it was externally validated on an independent subset of the ENHANCE.PET 1.6k dataset, where it reproduced expert-verified skeletal anatomy with a whole-skeleton Dice of 0.913 and a red-marrow-compartment Dice of 0.899; nonetheless, that dataset uses [^{18}F]FDG rather than [^{68}Ga]Ga-DOTA-TOC, and although the CT component is tracer-independent, a dedicated validation on DOTA-TOC PET/CT-derived CT would further strengthen confidence in the anatomical compartment-assignment layer.

Inter-observer variability of the manual reference standard was quantified on $n = 38$ examinations (see Methods), yielding a mean DSC of 0.715 ± 0.169 , which contextualises the algorithm DSC of 0.806 against a human-level ceiling of approximately 0.72.

External prospective validation correlating automated OMIS with observed haematotoxicity grades after [¹⁷⁷Lu]Lu-DOTA-TATE PRRT is planned.

Conclusion

This study demonstrates the technical feasibility of automated whole-body [⁶⁸Ga]Ga-DOTA-TOC PET/CT lesion segmentation and OMIS computation using SwinUNETR, with performance approaching inter-expert agreement on an independent test set. These preliminary results support the potential of automated OMIS as a reproducible pre-PRRT biomarker of bone marrow involvement. Prospective multicentre validation with haematological outcome data is required before clinical deployment as a haematotoxicity risk stratification tool.

Abbreviations

CT: computed tomography

DSC: Dice similarity coefficient

LoA: limits of agreement

FDG: fluorodeoxyglucose

GEP: gastroenteropancreatic

GT: ground truth

MAE: mean absolute percentage error

NET: neuroendocrine tumour

NPV: negative predictive value

OMIS: Osteo-Medullary Invasion Score

PET: positron emission tomography

pp : percentage points

PPV: positive predictive value

PRRT: peptide receptor radionuclide therapy

SSTR: somatostatin receptor

SwinUNETR: Shifted-Window UNet TRansformer

TMTV: total metabolic tumour volume

TMTV_bone: bone-compartment TMTV

Declarations

Ethics approval and consent to participate : Retrospective study approved by the Institut Paoli-Calmettes ethics committee (PIXIRIV-IPC 2023-034). Informed consent waiver granted.

Consent for publication: Not applicable. This manuscript contains no individually identifiable patient data.

Competing interests: S.M. and S.K. are employees of Pixilib SAS (Toulouse, France). All other authors declare no competing interests.

Funding: No financial funding was received for this study. Pixilib SAS (Toulouse, France) contributed in-kind through engineering time.

Availability of data and materials: Anonymised performance metrics are available from the corresponding author upon request. Patient-level imaging data cannot be shared owing to RGPD/GDPR constraints. Model weights will be released upon acceptance.

Authors' contributions: R.F.: study conception, pipeline development, deep learning implementation, statistical analysis, and manuscript writing. S.M. and S.K.: code review, pseudonymisation solution, and Attention U-Net architecture. J.C.: clinical and biological data extraction and statistical analysis. T.R.: manual lesion contouring and inter-observer variability assessment. A.M.: clinical and biological data extraction. N.C.: critical revision of the manuscript. S.B.: critical revision of the manuscript. D.O.: manual lesion contouring, inter-observer variability assessment, intellectual contribution to model design, and critical revision of the manuscript. S.O.-T.: study initiation, patient referral, medical review, and clinical validation. All authors read and approved the final manuscript.

Acknowledgements: The authors thank the nuclear medicine technologists and clinical research staff of the Institut Paoli-Calmettes for their support in data collection, and the ENHANCE.PET initiative for making their dataset publicly available.

References

1. Yao JC, Hassan M, Phan A, Dagohoy C, Leary C, Mares JE, et al. One hundred years after “carcinoid”: epidemiology of and prognostic factors for neuroendocrine tumors in 35,825 cases in the United States. *J Clin Oncol*. 2008;26:3063–72. <https://doi.org/10.1200/JCO.2007.15.4377>

2. Dasari A, Shen C, Halperin D, Zhao B, Zhou S, Xu Y, et al. Trends in the Incidence, Prevalence, and Survival Outcomes in Patients With Neuroendocrine Tumors in the United States. *JAMA Oncol.* 2017;3:1335–42. <https://doi.org/10.1001/jamaoncol.2017.0589>
3. Bozkurt MF, Virgolini I, Balogova S. Guideline for PET/CT imaging of neuroendocrine neoplasms with 68Ga-DOTA-conjugated somatostatin receptor targeting peptides and 18F-DOPA. *Eur J Nucl Med Mol Imaging.* 2017;44:1588–601.
4. Hope TA, Bergsland EK, Bozkurt MF. Appropriate use criteria for somatostatin receptor PET imaging in neuroendocrine tumors. *J Nucl Med.* 2018;59:66–74.
5. Strosberg J, El-Haddad G, Wolin E, Hendifar A, Yao J, Chasen B, et al. Phase 3 Trial of 177Lu-Dotatate for Midgut Neuroendocrine Tumors. *New England Journal of Medicine.* 2017;376:125–35. <https://doi.org/10.1056/NEJMoa1607427>
6. Singh S, Halperin D, Myrehaug S, Herrmann K, Pavel M, Kunz PL, et al. [177Lu]Lu-DOTA-TATE plus long-acting octreotide versus high-dose long-acting octreotide for the treatment of newly diagnosed, advanced grade 2–3, well-differentiated, gastroenteropancreatic neuroendocrine tumours (NETTER-2): an open-label, randomised, phase 3 study. *The Lancet. Elsevier;* 2024;403:2807–17. [https://doi.org/10.1016/S0140-6736\(24\)00701-3](https://doi.org/10.1016/S0140-6736(24)00701-3)
7. Oomen SP, Hennik PB, Antonissen C. Somatostatin is a selective chemoattractant for primitive (CD34(+)) hematopoietic progenitor cells. *Exp Hematol.* 2002;30:116–25. [https://doi.org/10.1016/s0301-472x\(01\)00772-x](https://doi.org/10.1016/s0301-472x(01)00772-x)
8. Nguyen N, Min Y, Rivière J. Limitations of the radiotheranostic concept in neuroendocrine tumors due to lineage-dependent somatostatin receptor expression on hematopoietic stem and progenitor cells. *Theranostics.* 2025;15:6497–515. <https://doi.org/10.7150/thno.113354>
9. Brabander T, van der Zwan WA, Teunissen JJM, Kam BLR, Feelders RA, de Herder WW, et al. Long-Term Efficacy, Survival, and Safety of [177Lu-DOTA0,Tyr3]octreotate in Patients with Gastroenteropancreatic and Bronchial Neuroendocrine Tumors. *Clin Cancer Res.* 2017;23:4617–24. <https://doi.org/10.1158/1078-0432.CCR-16-2743>
10. Bergsma H, Konijnenberg MW, Kam BLR, Teunissen JJM, Kooij PP, de Herder WW, et al. Subacute haematotoxicity after PRRT with 177Lu-DOTA-octreotate: prognostic factors, incidence and course. *Eur J Nucl Med Mol Imaging.* 2016;43:453–63. <https://doi.org/10.1007/s00259-015-3193-4>
11. Bergsma H, van Lom K, Raaijmakers MHGP, Konijnenberg M, Kam BLR, Teunissen JJM, et al. Persistent Hematologic Dysfunction after Peptide Receptor Radionuclide Therapy with 177Lu-DOTATATE: Incidence, Course, and Predicting Factors in Patients with Gastroenteropancreatic Neuroendocrine Tumors. *J Nucl Med.* 2018;59:452–8. <https://doi.org/10.2967/jnumed.117.189712>
12. Ferrara R, Zemmour C, Reichert T, Ouk D, Niccoli P, Maniry-Quellier J, et al. Predictive factors for persistent thrombocytopenia after peptide receptor radioligand therapy in enteropancreatic neuroendocrine tumors. *Front Endocrinol [Internet].* 2025 [cited 2025 July 15];16. <https://doi.org/10.3389/fendo.2025.1568243>

13. Vigne J, Chrétien B, Bignon A-L, Bouhier-Leporrier K, Dolladille C. [177Lu]Lu-DOTATATE peptide receptor radionuclide therapy–associated myeloid neoplasms: insights from the WHO pharmacovigilance database. *Eur J Nucl Med Mol Imaging*. 2022;49:3332–3. <https://doi.org/10.1007/s00259-022-05833-6>
14. Ferrara R, Darreon J, Giraud R. Investigation of risk factors for 177Lu-DOTATATE-induced thrombocytopenia in routine follow-up. *Médecine Nucléaire*. 2020;44:40–52.
15. Cristy M. Active bone marrow distribution as a function of age in humans. *Physics in Medicine and Biology*. 1981;26:389–400. <https://doi.org/10.1088/0031-9155/26/3/003>
16. Zhao Y, Gafita A, Vollnberg B. Deep neural network for automatic characterization of lesions on 68Ga-PSMA-11 PET/CT. *Eur J Nucl Med Mol Imaging*. 2020;47:603–13.
17. Wasserthal J, Breit H-C, Meyer MT. TotalSegmentator: robust segmentation of 104 anatomic structures in CT images. *Radiology AI*. 2023;5.
18. Hatamizadeh A, Nath V, Tang Y, Yang D, Roth HR, Xu D. Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images. In: Crimi A, Bakas S, editors. *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries* [Internet]. Cham: Springer International Publishing; 2022 [cited 2026 Apr 8]. p. 272–84. https://doi.org/10.1007/978-3-031-08999-2_22
19. Tang Y, Yang D, Li W. Self-supervised pre-training of Swin Transformers for 3D medical image analysis. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. p. 20730–40.
20. Liu Z, Lin Y, Cao Y. Swin Transformer: hierarchical vision transformer using shifted windows. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. p. 10012–22.
21. Yazdani E, Karamzadeh-Ziarati N, Cheshmi SS. Automated segmentation of lesions and organs at risk on [68Ga]Ga-PSMA-11 PET/CT images using self-supervised learning with Swin UNETR. *Cancer Imaging*. 2024;24.
22. Malmon S, Elsensohn MH, Thieblemont C. Prognostic impact of metabolic tumor volume using the SUV4.0 segmentation threshold in 1,960 lymphoma patients from prospective LYSA trials. *Eur J Nucl Med Mol Imaging*. 2025;52:3180–9.
23. Oktay O, Schlemper J, Le Folgoc L, Lee M, Heinrich M, Misawa K, et al. Attention U-Net: learning where to look for the pancreas. *Proceedings of the 1st Conference on Medical Imaging with Deep Learning (MIDL 2018)*. 2018.
24. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. 2021;18:203–11.
25. Cardoso MJ, Li W, Brown R, Ma N, Kerfoot E, Wang Y, et al. MONAI: An open-source framework for deep learning in healthcare [Internet]. *arXiv*; 2022 [cited 2026 Apr 9]. <https://doi.org/10.48550/arXiv.2211.02701>
26. Salehi SSM, Erdogmus D, Gholipour A. Tversky Loss Function for Image Segmentation Using 3D Fully Convolutional Deep Networks. In: Wang Q, Shi Y, Suk H-I, Suzuki K, editors. *Machine Learning in*

- Medical Imaging [Internet]. Cham: Springer International Publishing; 2017 [cited 2026 Apr 8]. p. 379–87. https://doi.org/10.1007/978-3-319-67389-9_44
27. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Med Imaging*. 2015;15:29. <https://doi.org/10.1186/s12880-015-0068-x>
 28. Zhang L, LaBelle W, Unberath M, Chen H, Hu J, Li G, et al. A vendor-agnostic, PACS integrated, and DICOM-compatible software-server pipeline for testing segmentation algorithms within the clinical radiology workflow. *Front Med*. 2023;10:1241570. <https://doi.org/10.3389/fmed.2023.1241570>
 29. Santilli A, Panyam P, Autz A, Wray R, Philip J, Elnajjar P. Automated full body tumor segmentation in DOTATATE PET/CT for neuroendocrine cancer patients. *Int J Comput Assist Radiol Surg*. 2023;18:2083–90. <https://doi.org/10.1007/s11548-023-02968-1>
 30. Carlsen EA, Johnbeck CB, Loft M. A convolutional neural network for total tumor segmentation in [64Cu]Cu-DOTATATE PET/CT of patients with neuroendocrine neoplasms. *EJNMMI Res*. 2022;12.
 31. Haque F, Carrasquillo JA, Turkbey EB, Mena E, Lindenberg L, Eclarinal PC, et al. An automated pheochromocytoma and paraganglioma lesion segmentation AI-model at whole-body 68Ga-DOTATATE PET/CT. *EJNMMI Res*. 2024;14:103. <https://doi.org/10.1186/s13550-024-01168-5>
 32. Ciarmiello A, Yosifov N, Masciale D. Global and regional accuracy of deep learning-based tumor segmentation from whole-body [18F]FDG PET/CT images. *EJNMMI Res*. 2026;16. <https://doi.org/10.1186/s13550-025-01333-4>
 33. Steinhelfer L, Ilhan H, Todica A. Spleen volume reduction is a reliable and independent biomarker for long-term risk of leukopenia development in peptide receptor radionuclide therapy. *J Nucl Med*. 2024;65:1244–9.
 34. Ferrara D, Pires M, Gutschmayer S, Yu J, Rausch I, Beyer T, et al. Sharing a whole-/total-body [18F]FDG-PET/CT dataset with CT-derived segmentations: an ENHANCE.PET initiative. *Sci Data*. 2026. <https://doi.org/10.1038/s41597-026-07218-y>

Figures

Automated OMIS Workflow using [⁶⁸Ga]Ga-DOTA-TOC PET/CT

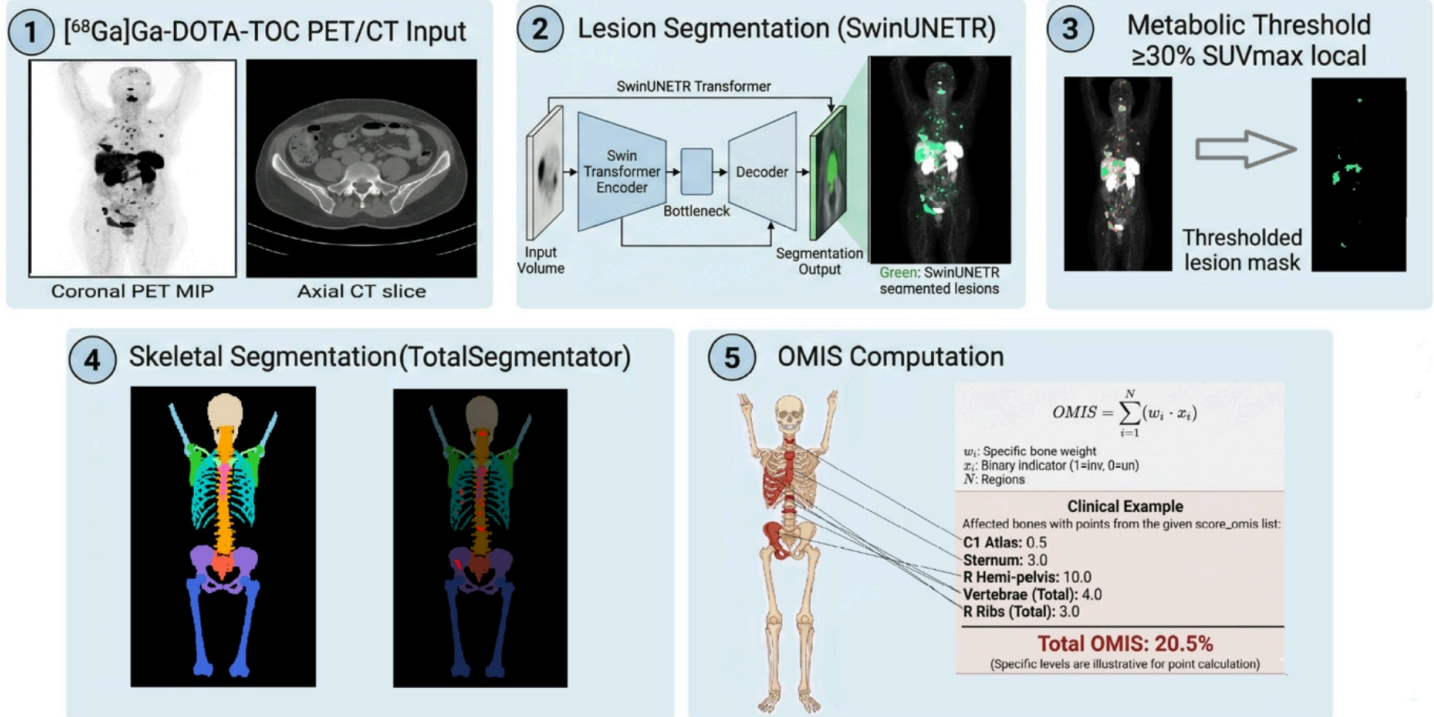


Figure 1

Automated OMIS computation pipeline. Step 1: [⁶⁸Ga]Ga-DOTA-TOC PET/CT input. Step 2: whole-body lesion segmentation by SwinUNETR. Step 3: per-lesion 30% SUVmax thresholding. Step 4: skeletal segmentation by TotalSegmentator. Step 5: OMIS computation using Cristy bone-marrow weights.

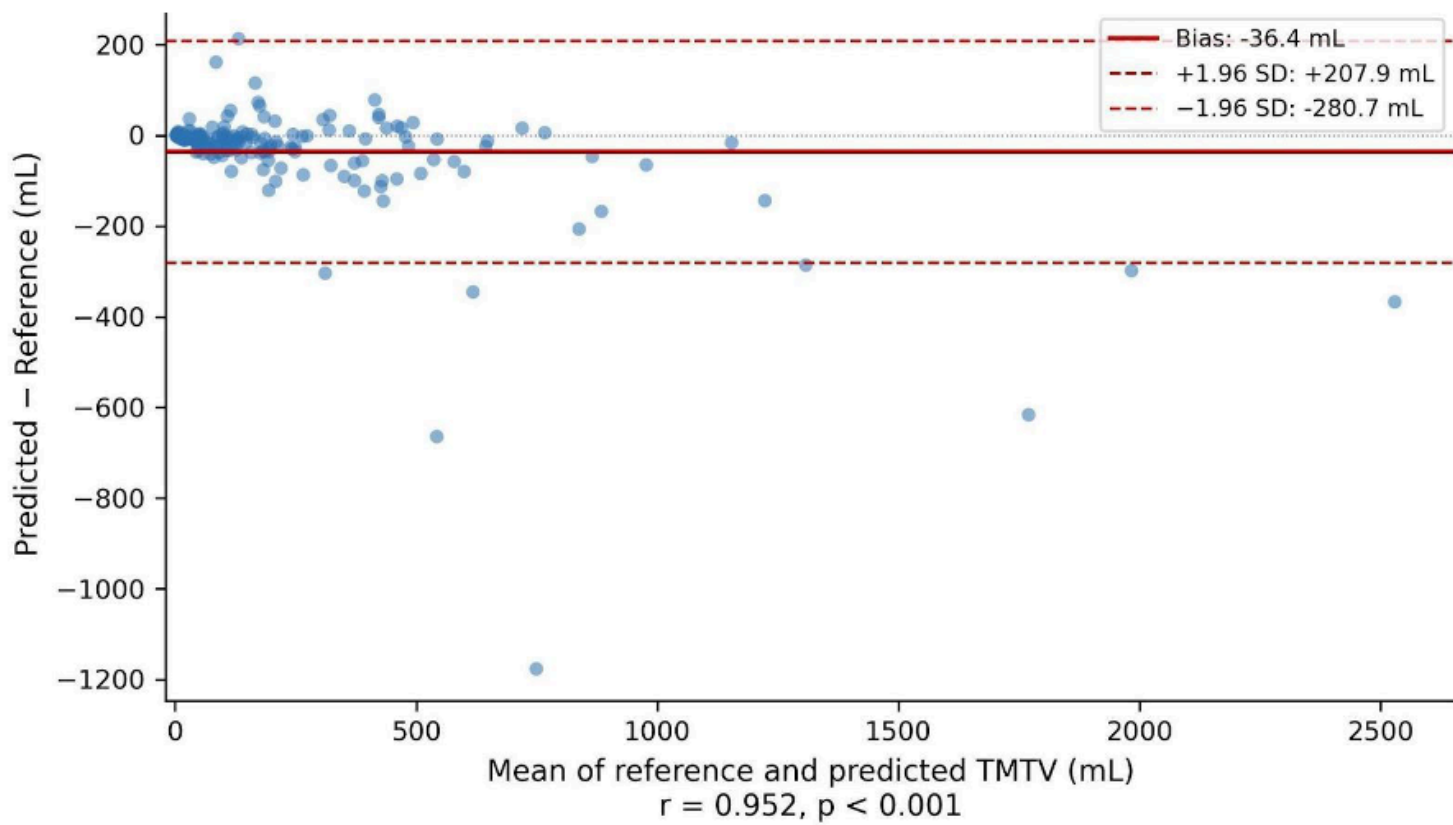


Figure 2

Bland-Altman analysis of TMTV between automated and manual segmentation on the independent test set (n=19 examinations).

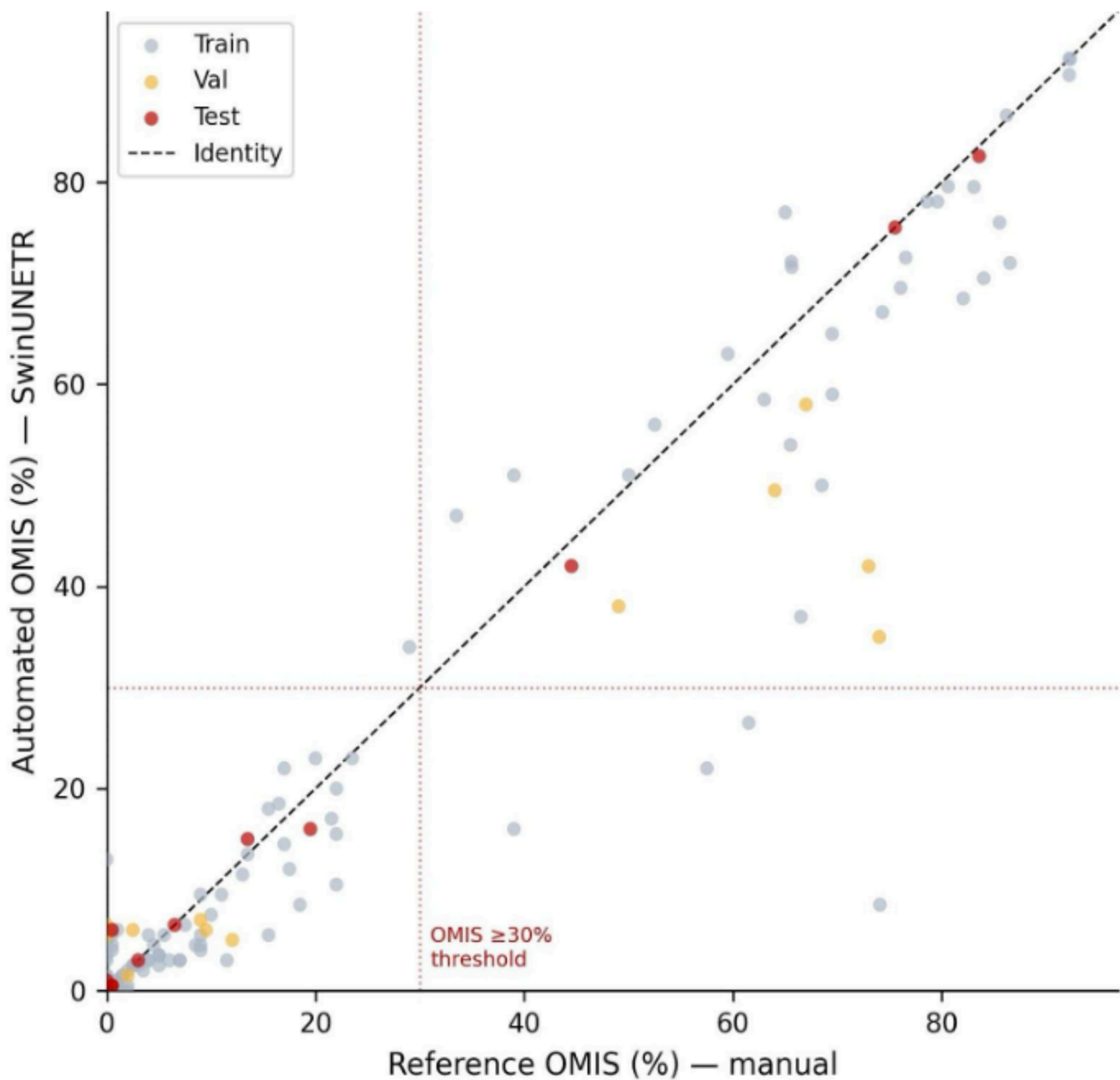


Figure 3

Correlation between automated and manual OMIS across all 188 examinations. The dashed red lines indicate the clinically validated haematotoxicity risk threshold (OMIS $\geq 30\%$). Colour coding reflects dataset partition: grey = training, yellow = validation, red = test.

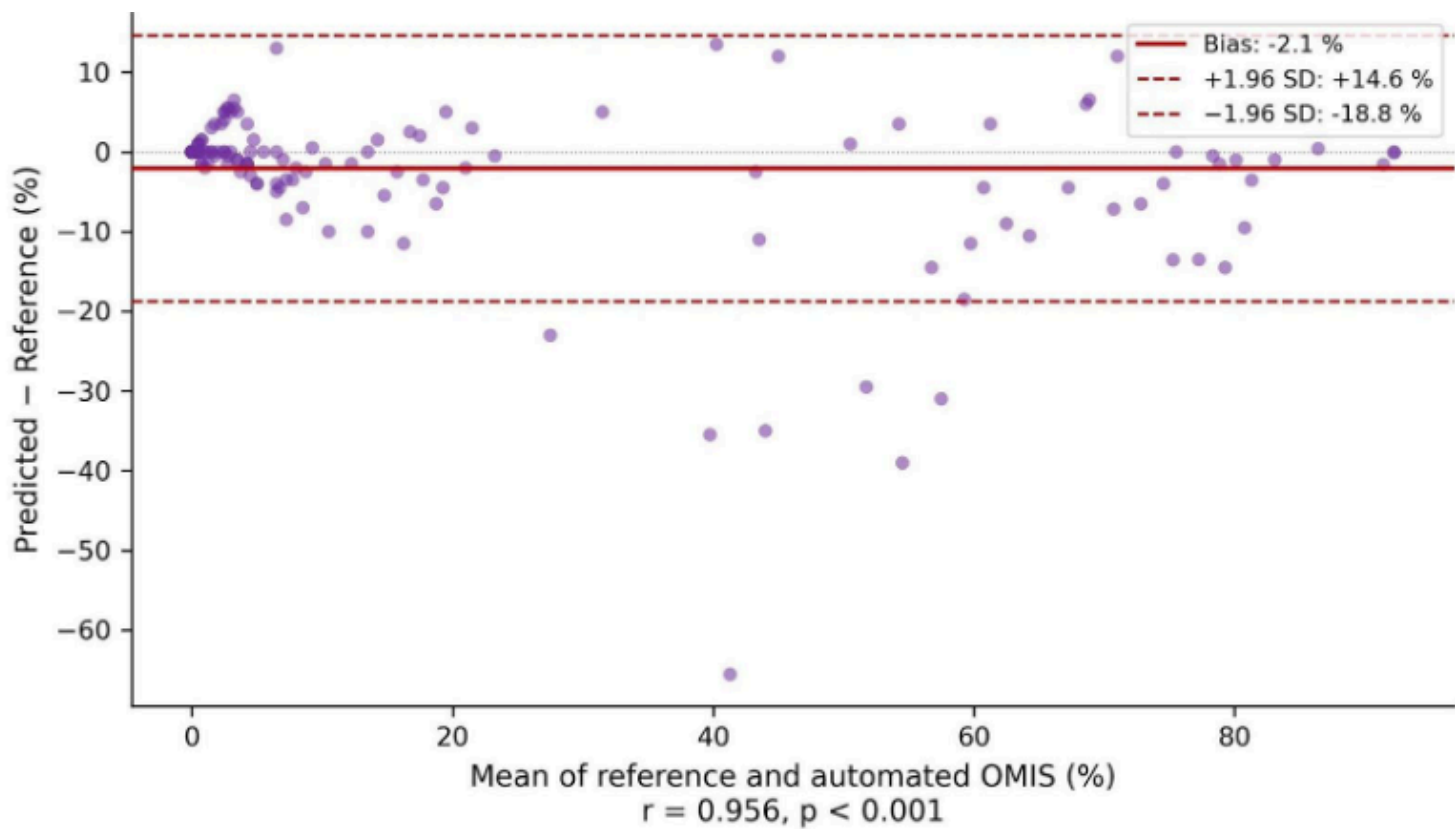


Figure 4

Bland-Altman analysis of the Osteo-Medullary Invasion Score (OMIS) between automated SwinUNETR pipeline and manual reference across all 188 examinations.

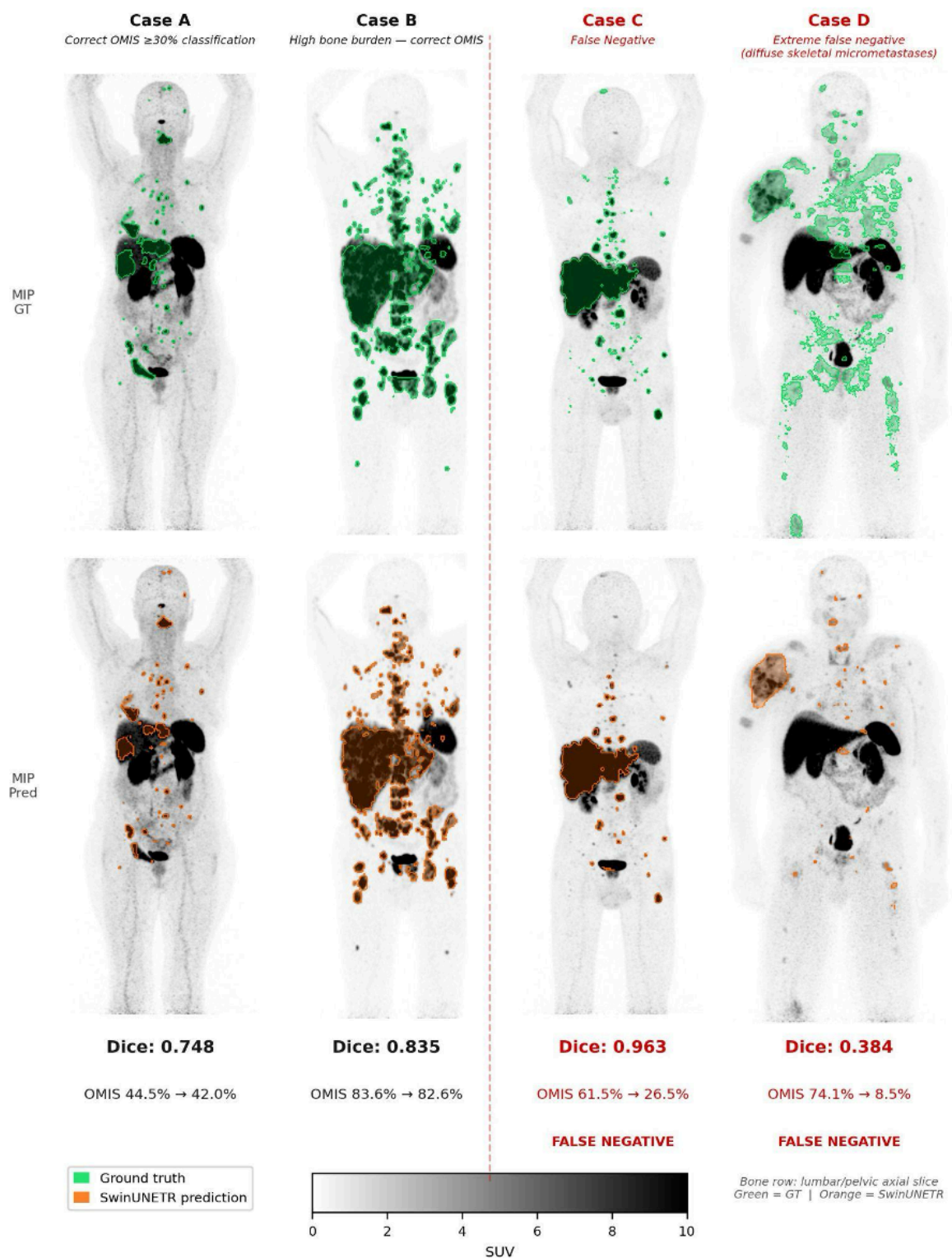


Figure 5

Representative examples of automated whole-body $[^{68}\text{Ga}]\text{Ga-DOTA-TOC}$ PET/CT lesion segmentation by SwinUNETR. Cases A–B: correct OMIS classification. Cases C–D: false-negative OMIS misclassification due to diffuse small-volume skeletal disease below the network's effective detection threshold. Top row: maximum intensity projection (MIP) overlaid with ground-truth contours (green). Bottom row: MIP with SwinUNETR predicted contours (orange).

