# Automated Segmentation of Liver and Liver Tumors with SwinUNETR and UNET Neural Networks on 18F-FDG PET/CT

*SwinUNETR ve Residual UNET Nöral Ağlar ile 18F-FDG PET/BT Karaciğer ve Karaciğer Tümörlerin Otomatik Segmentasyonu*

● Burak Demir, ● Hatice Kübra Yurtçu, ● Merve Ağcıoğlu Atalay, ● Fikret Ertek

*University of Health Sciences Türkiye, Şanlıurfa Mehmet Akif İnan Training and Research Hospital, Clinic of Nuclear Medicine, Şanlıurfa, Türkiye*

## Abstract

**Objectives:** To develop and evaluate automated segmentation models for the liver and hepatic tumors on 18F-fluorodeoxyglucose positron emission tomography/computed tomography (18F-FDG PET/CT) using SwinUNETR and residual UNET architectures, and to assess their accuracy in complex clinical cases.

**Methods:** In this single-center retrospective study, 100 patients (48 males, 52 females; mean age 61±14 years) with 18F-FDG-avid hepatic lesions from various primary malignancies were included. Liver segmentation was performed on non-contrast CT images using pairs of SwinUNETR and residual UNET models, and tumor segmentation was performed on masked PET images using separately trained pair of SwinUNETR and residual UNET model. Model performance was evaluated using the dice similarity coefficient (DSC), volumetric bias, and Bland-Altman analysis for metabolic tumor volume (MTV) and total lesion glycolysis (TLG).

**Results:** For liver segmentation, SwinUNETR achieved a median DSC of 97.59% (range: 95.41-98.93%) with a median volumetric bias of -0.94% (LoA: -3.76% to +0.50%), while residual UNET achieved a median DSC of 97.85% (range: 94.81-98.80%) with a median volumetric bias of -0.34% (LoA: -2.63% to +1.16%). For tumor segmentation, SwinUNETR achieved a median DSC of 92.62% (range: 80.75–97.46%), an MTV bias of -8.60% (LoA: -31.62% to +1.21%), and a TLG bias of -6.40% (LoA: -25.58% to +0.76%). Residual UNET achieved a median DSC of 93.07% (range: 80.74–98.18%), MTV bias of -4.33% (LoA: -24.36% to +10.12%), and TLG bias of -11.10% (LoA: -30.8% to +4.52%). Most MTV and TLG measurements were within ±10% of reference values.

**Conclusion:** Both SwinUNETR and Residual UNET achieved excellent liver segmentation accuracy and clinically acceptable tumor segmentation performance on 18F-FDG PET/CT, with SwinUNETR showing slightly better performance in liver volumetric measurements. These open-source models could be integrated into clinical workflows to automate segmentation tasks, facilitate treatment planning for liver-directed therapies, and support reproducible quantitative imaging analyses.

**Keywords:** Deep learning, SwinUNETR, positron-emission tomography, molecular imaging, fluorodeoxyglucose, image processing

## Öz

**Amaç:** Bu çalışmanın amacı SwinUNETR ve Residual UNET mimarilerini kullanarak 18F-florodeoksiglukoz pozitron emisyon tomografisi/bilgisayarlı tomografinin (18F-FDG PET/BT) görüntülerinde karaciğer ve hepatik tümörlerin otomatik segmentasyon modellerini geliştirmek ve değerlendirmek; ayrıca bu modellerin karmaşık klinik olgulardaki doğruluğunu incelemek olarak belirlenmiştir.

**Yöntem:** Tek merkezli, retrospektif bu çalışmaya, çeşitli primer ve metastatik ^{18}F-FDG tutulumu gösteren karaciğer tümörleri bulunan 100 hasta (48 erkek, 52 kadın; ortalama yaş 61±14 yıl) dahil edildi. Karaciğer segmentasyonu kontrastsız BT görüntülerinde, tümör segmentasyonu ise maskeleme yapılmış PET görüntülerinde gerçekleştirildi. SwinUNETR ve Residual UNET modelleri karaciğer ve tümör segmentasyonu için ayrı ayrı eğitildi. Model performansı dice benzerlik katsayısı (DSC), volumetrik bias ve metabolik tümör hacmi (MTV) ile total lezyon glikolizi (TLG) değerleri Bland-Altman analizi ile karşılaştırıldı.

**Bulgular:** Karaciğer segmentasyonunda SwinUNETR modeli ile %97,59 (aralık: %95,41-98,93) medyan DSC ve -%0,94 (LoA: -%3,76 ila +%0,50) medyan volumetrik bias elde edilmiştir. Residual UNET modelinde ise %97,85 (aralık: %94,81-98,80) medyan DSC ve -%0,34 (LoA: -%2,63 ila +%1,16) bias değerleri izlenmiştir. Tümör segmentasyonunda SwinUNETR modelinde %92,62 (aralık: %80,75-97,46) medyan DSC, -%8,60 MTV bias (LoA: -%31,62 ila +%1,21) ve -%6,40 TLG bias (LoA: -%25,58 ila +%0,76) değerleri gözlenmiştir. Residual UNET modeli ise %93,07 (aralık: %80,74-98,18) medyan DSC, -%4,33 MTV bias (LoA: -%24,36 ila +%10,12) ve -%11,10 TLG bias (LoA: -%30,8 ila +%4,52) değerlerine sahip olarak izlenmiştir. MTV ve TLG ölçümlerinin çoğu referans değerlerin ±%10 aralığında yer aldı.

**Sonuç:** SwinUNETR ve Residual UNET modelleri, ^{18}F-FDG PET/BT görüntülerinde yüksek derecede karaciğer segmentasyon doğruluğu ve klinik olarak kabul edilebilir tümör segmentasyonu performansı sağlamıştır. SwinUNETR modeli ise karaciğer segmentasyonunda Residual UNET modeline göre daha iyi sonuç vermiştir. Bu açık kaynaklı modeller, klinik iş akışlarına entegre edilerek segmentasyon görevlerini otomatikleştirebilir, karaciğere yönelik tedavi planlamasını kolaylaştırabilir ve tekrarlanabilir nicel görüntüleme analizlerini destekleyebilir.

**Anahtar kelimeler:** Derin öğrenme, SwinUNETR, pozitron emisyon tomografisi, moleküler görüntüleme, florodeoksiglikoz, görüntü işleme

## Introduction

Primary liver malignancies, particularly hepatocellular carcinoma (HCC), represent a major global health burden, ranking as one of the leading causes of cancer-related mortality worldwide (1). In addition to primary tumors, the liver is a frequent site of metastatic spread from various malignancies, including colorectal, breast, and pancreatic cancers (2,3,4). Early detection and accurate characterization of hepatic lesions are essential, as the prognosis of patients with liver involvement depends heavily on timely diagnosis and appropriate therapeutic intervention. Proper treatment planning—whether through surgical resection, transplantation, systemic therapy, or locoregional approaches—can significantly improve survival outcomes in both primary and secondary hepatic malignancies.

Accurate delineation of the liver and its tumors plays a pivotal role in several advanced treatment strategies. For therapies such as selective internal radiation therapy (SIRT) (SIRT, also known as radioembolization) and stereotactic body radiotherapy, precise volumetric and spatial characterization of tumor burden is required to optimize dosimetry, minimize healthy tissue damage, and maximize therapeutic efficacy (5-7). Furthermore, quantitative imaging biomarkers that have been shown to be reliable prognostic factors after radioembolization, such as metabolic tumor volume (MTV) and total lesion glycolysis (TLG) on ^{18}F-fluorodeoxyglucose positron emission tomography/computed tomography (^{18}F-FDG PET/CT), rely on precise segmentation to ensure reproducibility across clinical and research settings (8,9).

Over the past decade, deep learning–based methods have revolutionized medical image segmentation, with convolutional neural networks (CNNs) and, more recently, transformer-based architectures delivering state-of-the-art performance (10-12). Tools such as TotalSegmentator have demonstrated the potential of generalized pre-trained models to achieve high accuracy in multi-organ segmentation tasks (13). In liver imaging, these approaches have significantly reduced the need for labor-intensive manual contouring, thus accelerating clinical workflows and enabling large-scale quantitative studies.

The SwinUNETR architecture, a transformer-based model incorporating hierarchical shifted-window self-attention and UNet-style encoder–decoder design, has shown strong performance in complex 3D segmentation tasks (10,11). By leveraging global contextual information while preserving fine anatomical details, SwinUNETR has the potential to outperform conventional CNN-based architectures in challenging segmentation scenarios. In clinical reality, diseased livers often present with anatomical distortions caused by ascites, postoperative changes, large tumor burdens, or extensive metastatic infiltration. Such conditions may degrade the performance of general-purpose segmentation models, underscoring the need for disease-specific model training tailored to these complex cases.

Previous studies on liver segmentation using neural networks have generally employed fully convolutional architectures such as Residual UNET and have been performed on contrast-enhanced CT images. For example, in a recent study, Yashaswini et al. (14) evaluated the performance of Residual UNET models for liver and tumor segmentation on CT imaging and reported a Dice score of 91.44% for liver segmentation. Additionally, several other studies have investigated liver and tumor segmentation using CNNs (15,16). However, the utility and potential superiority of

SwinUNETR for liver and tumor segmentation, compared to Residual UNET models, have not yet been explored. Furthermore, although there are multiple studies on tumor segmentation in PET imaging, research combining PET and CT imaging for segmentation remains rare.

In this study, we aimed to develop and evaluate automated segmentation models, using both SwinUNETR-V2 and residual U-Net architectures, to segment the liver and hepatic tumors from ^{18}F-FDG PET/CT images. Our goal was to assess their accuracy in the context of challenging clinical cases and to explore the feasibility of disease-specific segmentation models that can maintain robust performance in anatomically complex livers. The developed models are also intended for use in conjunction with the previously developed radioembolization dosimetry module for 3D Slicer (17).

## Materials and Methods

### Patients and Study Design

This single-center, retrospective study included patients with ^{18}F-FDG-avid hepatic lesions from various malignancies who underwent ^{18}F-FDG PET/CT imaging from January 2025 to July 2025. Written informed consent was obtained from all patients before imaging. Exclusion criteria were: (1) significant respiratory artefacts; (2) artefacts secondary to patient motion; and (3) artefacts secondary to metallic objects or prostheses on CT imaging. The study was approved by the Ethics Committee of Harran University (approval no: HRÜ-25.11.02, date: 16.06.2025), with additional approval from the institutional review board. The developed segmentation models and training scripts (18,19), images of four patients for testing (20), and the SlicerAether segmentation module for 3D Slicer (18) are available in public repositories.

### ^{18}F-FDG PET/CT Protocol and Preprocessing of the Data

Imaging was performed using a Siemens Biograph Horizon™ 4R system. Patients fasted for at least 6 hours before imaging, and blood glucose levels were checked prior to the scan. Those with a blood glucose level above 200 mg/dL did not undergo scanning. Images were acquired from the vertex to the proximal femur with the patient in the supine position. Whole-body ^{18}F-FDG PET/CT imaging was performed approximately 1 h after an intravenous injection of ^{18}F-FDG at 3.7 MBq/kg. For PET/CT imaging, PET images were acquired for 90 seconds per bed position and were reconstructed using attenuation correction measured from non-contrast CT images. For reconstruction of PET images, the TrueX+TOF (UltraHD-

PET) algorithm was used with 4 iterations and 10 subsets, a 5-mm post-processing Gaussian filter, and a 180 × 180 matrix. The resulting voxel size was 4.11392 × 4.11392 × 1.50 mm. All PET images were converted to standardized uptake values (SUVs), normalized to body weight, and resampled to an isotropic voxel size of 2 × 2 × 2 mm prior to training. No further normalization other than conversion to SUV values was used.

Non-contrast-enhanced CT images were acquired at 130 kV with a variable tube current modulated according to patient weight using the CareDose4D (Siemens Healthineers) and reconstructed with a 512 × 512 matrix. The resulting voxel size was 1.367 × 1.367 × 1.50 mm. Similarly, the CT images were resampled to an isotropic voxel size of 2×2×2 mm before training. For CT images, voxel intensities were normalized using a linear scaling transformation in which values between -135 hounsfield unit (HU) and +215 HU were mapped to the range 0.0-10.0; values outside this range were clipped to the nearest boundary.

For preprocessing, PET and CT volumes were cropped to a bounding box encompassing the upper abdomen to remove empty voxels and reduce computational load. The liver was manually segmented on CT images for all patients, and the resulting liver masks were used to zero out voxels outside the liver in the PET volumes. A spherical reference volume of interest was placed in the non-tumoral liver parenchyma, and a threshold equal to 1.5 times the liver reference $SUV_{mean}$ was used for manual tumor segmentation. Afterwards, the CT images, masked PET images, and liver and tumor segmentation masks were saved for further processing.

Training volumes were split into overlapping 96 × 96 × 96-voxel patches, yielding 680 training pairs. No patching was applied during testing; instead, a sliding-window inference with the same patch size was used. Preprocessing was performed using 3D Slicer (version 5.9) and custom Python scripts (21,22).

### Model Architecture Loss Function and Training Parameters

A volumetric segmentation model based on the SwinUNETR architecture, originally proposed by Hatamizadeh et al. (10) and later extended by He et al. (11) as SwinUNETR-V2, was implemented. SwinUNETR-V2 integrates the Swin Transformer with a UNET-style encoder–decoder and residual convolutional blocks at the start of each Swin stage, enabling high representational capacity for 3D medical images (10,23). In this study, one SwinUNETR-V2 model was trained to segment the liver in CT images, and another model with identical parameters was trained to segment tumors in masked PET images. Both models used a feature size of 24, transformer depths of (2, 2, 2, 2),

attention heads of (3, 6, 12, 24), a dropout path rate of 0.0, input volumes of 96 × 96 × 96 voxels, and gradient checkpointing to reduce memory usage. Each model had approximately 18.3 million trainable parameters and was trained with a batch size of 1.

We also utilized UNET-structured models with residual blocks for comparison with SwinUNETR models (12). The 3D residual UNET model was configured with an input batch size of 4 and a total of 76.8M trainable parameters. This network employed five resolution levels with channel sizes of 64, 128, 256, 512, and 1024; two residual units per level; strides of (2, 2, 2, 2) for down- and up-sampling; and 3 × 3 × 3 convolution kernels.

The implementation was based on the PyTorch and MONAI frameworks and executed on a graphics processing unit (GPU)-enabled system, allowing efficient handling of 3D volumetric data (20,21,22). The GPU and central processing unit models used for training were an NVIDIA GeForce RTX 4060 with 8 GB of VRAM and an Intel Core i3-9100F (3.60 GHz). The Dice similarity coefficient (DSC) was calculated as follows (24,25,26):

$$Dice\ score = \frac{2 \times |X \cap Y|}{|X| + |Y|}$$

Here, X denotes the set of voxels in the predicted segmentation; Y denotes the set of voxels in the ground-truth segmentation; and $|X \cap Y|$ denotes the number of overlapping voxels. Dice loss was defined as follows:

*Dice loss = 1 – Dice score.*

In addition, cross-entropy loss values were calculated, and a hybrid loss function was used for training:

*Training Loss Function = 0.5 × Dice Loss + 0.5 × Cross Entropy Loss*

### Testing of the Models and Performance Evaluation Metrics

A total of four models were developed for the segmentation of the liver and liver lesions. For liver segmentation, the reference liver volume, the predicted liver volume, and their intersection were computed. Model performance was assessed using the DSC, where a value of 1 indicates perfect overlap between the predicted and reference segmentation, and a value of 0 indicates no overlap.

For tumor segmentation on PET images, the DSC was also used as the primary evaluation metric. In addition, MTV and TLG were calculated for both the reference and model-predicted segmentations. TLG was defined as:

*TLG = MTV × SUV$_{mean}$*

### Statistical Analysis

Descriptive statistics were reported as counts and percentages for categorical variables, and as mean ± standard deviation and median (range) for continuous variables. A p-value less than 0.05 was considered statistically significant for all analyses. Dice scores obtained from the SwinUNETR and residual UNET models were compared using the Wilcoxon signed-rank test.

For tumor segmentation, predicted and reference MTV and TLG values were compared using Bland-Altman plots. The bias, along with 95% confidence intervals (CIs) and limits of agreement (LoA), was calculated for both models. All statistical analyses were performed using RStudio (version 2025.05.1), IBM SPSS Statistics (version 27), and BA-plotteR (27,28).

## Results

### Patients and General Characteristics

A total of 110 patients were initially considered for inclusion. Six patients were excluded due to respiratory artifacts, and four were excluded due to metallic artifacts in the upper abdominal CT images. Consequently, 100 patients (48 males, 52 females) with various malignancies were included in the study. The mean age was 61±14 years. The most common primary malignancies were breast cancer (28%), colorectal carcinoma (23%), lung cancer (13%), gastric cancer (8%), and pancreatic cancer (6%). The remaining patients had HCC, lymphoma, ovarian cancer, esophageal cancer, gallbladder cancer, cervical cancer, soft tissue sarcoma, tumors of unknown origin, or thyroid cancer. More than half of the patients (55%) had more than five FDG-avid liver lesions, 22% had 2-5 FDG-avid lesions, and 23% had a single FDG-avid lesion. The liver reference SUV$_{mean}$ was 2.17±0.48 g/mL, and the mean tumor SUV$_{max}$ was 10.52±7.50 g/mL.

Patients were randomly assigned to a training set (n=85) and a test set (n=15). In the test set, nine patients were female and six were male. Primary malignancies in this group included breast cancer (n=6), colorectal carcinoma (n=3), lung cancer (n=2), and lymphoma (n=2). The remaining patients had pancreatic cancer, tumors of unknown origin, or esophageal cancer. The mean age of the test group was 59±15 years, and the mean reference liver SUV$_{mean}$ was 2.25±0.32 g/mL.

### Segmentation of Liver on CT Images

In the test group, the median reference liver volume was 1679 mL (range: 887.6-2536.3 mL). The SwinUNETR model achieved a median Dice score of 97.59% (range: 95.41%-98.93%). The median liver volume estimated by SwinUNETR

Demir et al. Automated Segmentation of Liver and Tumors

Mol Imaging Radionucl Ther 2026;35(1):10-18

was 1672.2 mL (range: 872.9-2414.4 mL). Bland-Altman analysis demonstrated a median bias of –0.94% (95% CI: -1.05 to -0.64), with lower and upper LoA of -3.76% and +0.50%, respectively (Figure 1). These results indicate that SwinUNETR slightly underestimated the liver volume but maintained high segmentation accuracy.

The Residual UNET model achieved a median dice score of 97.85% (range, 94.81-98.80%). The median liver volume estimated by Residual UNET was 1693.17 mL (range: 891.24-2361.7 mL). Bland-Altman analysis revealed a median bias of -0.34% (95% CI: -0.58 to -0.17); LoA: -2.63% to +1.16%. When comparing the dice scores of the two models, SwinUNETR had higher scores in 13 patients (87%) and lower scores in 2 patients (13%) (p=0.036; Figure 2). The DSC values for each patient and the differences between SwinUNETR and Residual UNET models in liver segmentation are given in Table 1.

### Segmentation of Tumors on Masked PET Images

In the test group, the median number of liver tumors was 6 (range: 1-39), and the median reference MTV was 58.71 mL (range: 2.20-374.20 mL). The median $SUV_{max}$, $SUV_{mean}$, and TLG values in the reference segmentations were 9.98 g/mL (range: 5.46-18.65 g/mL), 4.76 g/mL (range: 3.20-9.21 g/mL), and 337.92 g (range: 8.76-3447.90 g), respectively.

The SwinUNETR model achieved a median dice score of 92.62% (range: 80.75%-97.46%). The median MTV and TLG estimated by SwinUNETR were 50.84 mL (range: 1.62-343.36 mL) and 287.11 g (range: 6.86-3334.18 g), respectively. In the Bland-Altman analysis, the SwinUNETR model demonstrated a median bias of -8.60% (95% CI: -16.8 to -2.15) for MTV, with lower and upper LoA of -31.62% and +1.21%, respectively. Similarly, SwinUNETR model had a median bias of -6.40% (95% CI: -10.08 to -2.13) for TLG, with lower and upper LoA of -25.58% and +0.76%, respectively.

The Residual UNET model achieved a median dice score of 93.07% (range: 80.74-98.18%). The medians of MTV and TLG estimated by Residual U-Net were 56.22 mL (range: 1.70-400.15 mL) and 269.20 g (range: 6.40-3015.43 g), respectively. In the Bland-Altman analysis, the Residual UNET model demonstrated a median bias of -4.33% (95% CI: -10.62% to -1.59%) for MTV, with lower and upper LoA of -24.36% and +10.12%, respectively. Similarly, the Residual UNET model showed a median bias of -11.10% (95% CI: -16.87 to -6.22) for TLG, with lower and upper LoA of -30.8% and +4.52%, respectively. When dice scores were compared, SwinUNETR outperformed Residual UNET in 8 patients (53%) and scored lower in 7 patients (47%) (p=0.570). Examples of patient segmentation results are shown in Figures 3 and 4. The DSC values for each patient
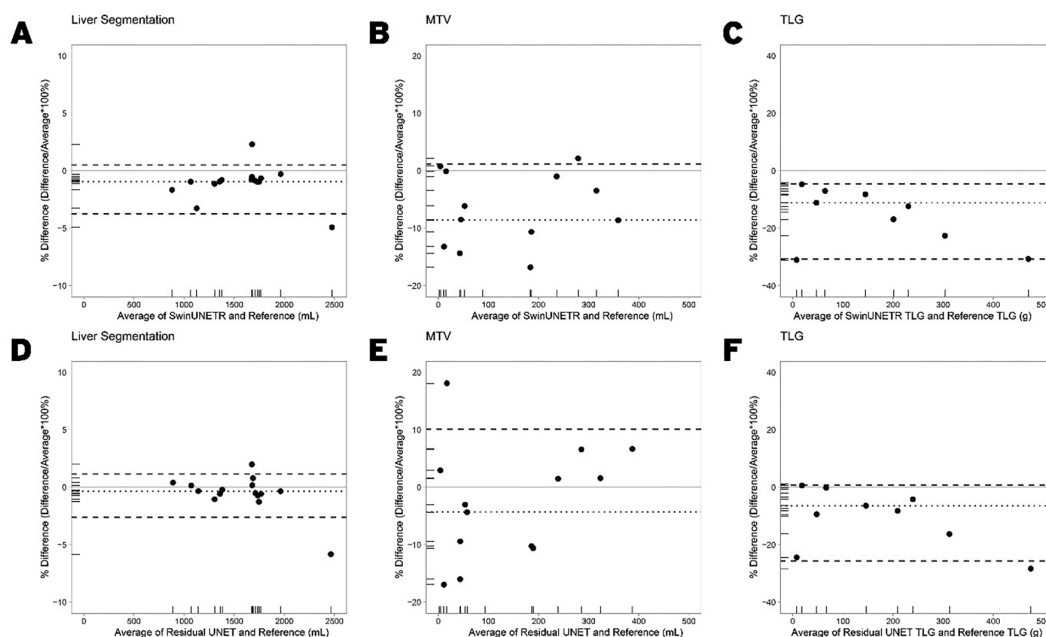


**Figure 1.** Bland-Altman plots of the segmentations predicted with SwinUNETR (A, B, C) and residual UNET (D, E, F) compared to the reference segmentation

MTV: Metabolic tumor volume, TLG: Total lesion glycolysis

and the differences between SwinUNETR and Residual UNET models in tumor segmentation are given in Table 2.
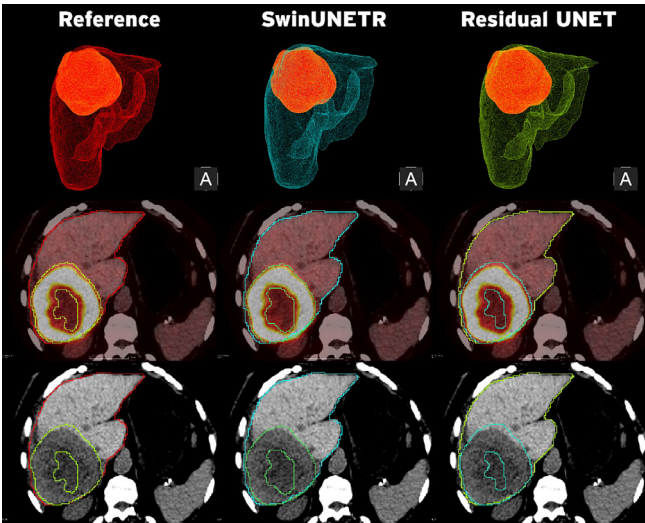
## Discussion

In this study, both the SwinUNETR and residual UNET models achieved excellent performance in liver segmentation on CT images, with median dice scores exceeding 97% and narrow LoA. Although the SwinUNETR model slightly outperformed the residual UNET in terms of dice score, the difference was modest, and both approaches demonstrated highly reliable volumetric agreement with
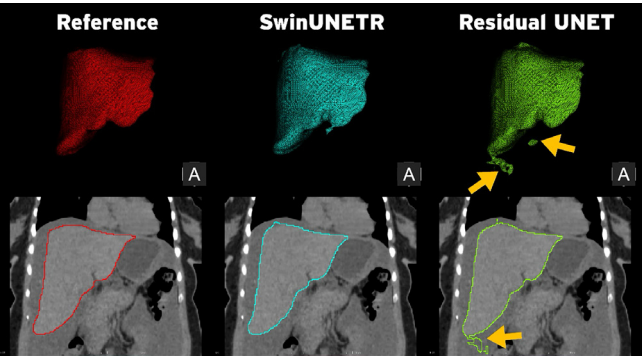


**Figure 2.** Reference and predicted segmentation results of a 36-year-old female patient with lymphoma. Significant ascites can be observed; SwinUNETR and Residual UNET models achieved dice scores of 97.65% and 97.36%, respectively, for liver segmentation. In the Residual UNET model, several extrahepatic regions are falsely identified as liver because of ascites (yellow arrows)
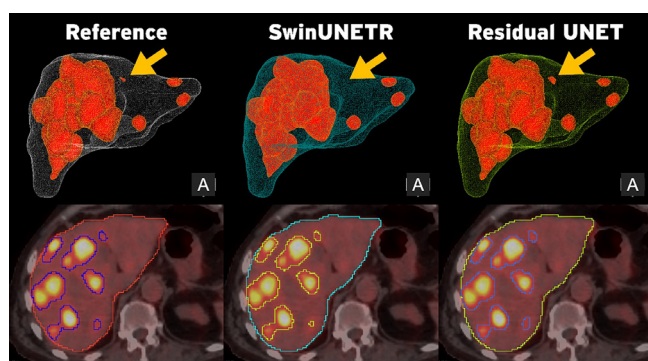


**Figure 3.** Reference and predicted segmentation results of an 81-year-old female patient with diffuse large B cell lymphoma. A large, strongly fluorodeoxyglucose-avid mass with areas of necrosis can be observed in the right lobe of the liver. SwinUNETR achieved dice scores of 98.71% for liver segmentation and 95.66% for tumor segmentation, while residual UNET achieved 98.50% for liver segmentation and 96.36% for tumor segmentation

**Table 1. The DSC values and differences in segmentation of liver calculated for both SwinUNETR and residual UNET models for each patient are given**

| Patients | SwinUNETR DSC | Residual UNET DSC | Difference |
|---|---|---|---|
| Patient 1 | 98.93% | 98.20% | +0.73% |
| Patient 2 | 98.84% | 98.60% | +0.24% |
| Patient 3 | 98.81% | 98.80% | +0.02% |
| Patient 4 | 98.81% | 98.64% | +0.16% |
| Patient 5 | 98.71% | 98.50% | +0.21% |
| Patient 6 | 98.63% | 98.51% | +0.12% |
| Patient 7 | 97.65% | 97.36% | +0.29% |
| Patient 8 | 97.59% | 97.85% | -0.26% |
| Patient 9 | 97.57% | 98.12% | -0.54% |
| Patient 10 | 97.26% | 97.20% | +0.06% |
| Patient 11 | 97.25% | 97.13% | +0.12% |
| Patient 12 | 96.62% | 96.56% | +0.06% |
| Patient 13 | 95.89% | 95.88% | +0.01% |
| Patient 14 | 95.73% | 95.25% | +0.48% |
| Patient 15 | 95.41% | 94.81% | +0.59% |
| DSC: Dice similarity coefficient | | | |

**Table 2. The DSC values and differences in segmentation of tumors calculated for both SwinUNETR and residual UNET models for each patient are given**

| Patients | SwinUNETR DSC | Residual UNET DSC | Difference |
|---|---|---|---|
| Patient 1 | 97.46% | 97.33% | +0.13% |
| Patient 2 | 97.07% | 98.18% | -1.11% |
| Patient 3 | 96.35% | 89.91% | +6.44% |
| Patient 4 | 96.02% | 93.99% | +2.03% |
| Patient 5 | 95.94% | 94.70% | +1.23% |
| Patient 6 | 95.66% | 96.36% | -0.70% |
| Patient 7 | 93.74% | 94.58% | -0.84% |
| Patient 8 | 92.62% | 90.29% | +2.32% |
| Patient 9 | 91.01% | 85.93% | +5.08% |
| Patient 10 | 90.25% | 93.07% | -2.81% |
| Patient 11 | 90.06% | 93.07% | -3.01% |
| Patient 12 | 87.84% | 94.29% | -6.44% |
| Patient 13 | 86.76% | 89.17% | -2.41% |
| Patient 14 | 82.37% | 87.53% | -5.16% |
| Patient 15 | 80.75% | 80.74% | +0.02% |
| DSC: Dice similarity coefficient | | | |

Demir et al. Automated Segmentation of Liver and Tumors

Mol Imaging Radionucl Ther 2026;35(1):10-18

reference segmentations. While both models may produce errors in patients with liver disease such as hepatosteatosis (Figure 5) or ascites (Figure 2) these results indicate that transformer-based and residual convolutional architectures are viable options for accurate hepatic segmentation in clinical and research settings.

For tumor segmentation on masked PET images, both models also demonstrated high performance, although their accuracy was lower than for liver segmentation.



**Figure 4.** Reference and predicted segmentation results of an 81-year-old male patient with rectal carcinoma. Multiple strongly fluorodeoxyglucose-avid lesions can be observed in both lobes of the liver. While both models successfully predicted most lesions, SwinUNETR missed a small metastasis, and the residual UNET model predicted the same lesion to be larger than the reference (yellow arrows). SwinUNETR achieved dice scores of 97.26% (liver) and 97.07% (tumor); residual UNET achieved 97.20% (liver) and 98.18% (tumor)
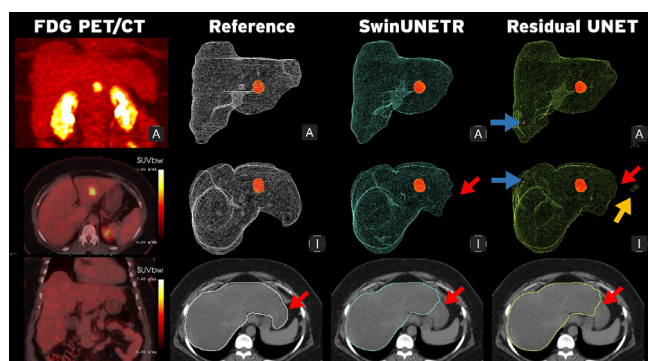


**Figure 5.** Reference and predicted segmentation results of a 56-year-old female patient with breast cancer metastasis of the liver. The patient had significant hepatosteatosis with hepatomegaly, and the lateral segment of the left lobe of the liver extended into the upper left quadrant. As a result, both the SwinUNETR and Residual UNET models failed to recognize the most lateral portions of the liver (red arrows); additionally, the UNET model produced a false-positive region erroneously labeled as liver (yellow arrow). In addition, the residual UNET model produced small false-positive foci in tumor segmentation due to noise in PET images (indicated by the blue arrows). SwinUNETR and Residual UNET models achieved Dice scores of 95.41% and 94.81% for liver segmentation and 91.01% and 85.93% for tumor segmentation, respectively

This is not unexpected, as tumor segmentation in FDG PET is inherently more challenging. Factors such as image noise, heterogeneous tracer uptake, and the presence of physiological uptake in adjacent structures can introduce false-positive voxels. Furthermore, variations in SUV thresholding methods can lead to differences in measured MTV and TLG, even for the same lesion. Despite these challenges, the majority of the predicted MTV and TLG values in our study were within ±10% of reference measurements, a level of agreement that is likely sufficient for many clinical applications, including treatment planning and response assessment. From a practical standpoint, these models could be integrated into clinical workflows to automate time-consuming segmentation tasks, assist in treatment planning for radiotherapy or radioembolization, and provide reproducible volumetric measurements for research studies. Given their open-source availability, they can also serve as a foundation for further development, including fine-tuning for specific scanner protocols or disease subtypes.

Our results compare favorably with the literature. Previous studies have reported Dice scores for liver segmentation in the range of 94-97% using deep learning methods (29,30,31), placing both of our models at the higher end of this reported range. In tumor segmentation using deep learning methods, Leung et al. (32) developed models using ^{18}F-FDG PET/CT and Galium-68 prostate-specific membrane antigen PET/CT and showed that median DSCs of up to 0.83 can be achieved for patients with lung cancer, melanoma, lymphoma, and prostate cancer. Although the dice scores achieved by both models (median >92%) indicate a high degree of accuracy, particularly given the heterogeneity of the test cohort, we masked the liver segment to simplify a two-pass algorithm; therefore, direct comparison was not feasible. Our approach differs by being specifically optimized for hepatic tumor segmentation on PET, potentially enhancing performance in cases of complex intrahepatic disease. In this context, Luo et al. (33) investigated the role of deep learning models in the detection and diagnosis of focal lesions in ^{18}F-FDG PET/CT images and achieved a Dice coefficient of 0.740. In addition, the developed models demonstrated high performance in differentiating benign from malignant liver nodules.

## Study Limitations

Our study has several limitations. First, it was conducted at a single center; external validation on datasets from other institutions would be necessary to confirm generalizability. Second, although our models demonstrated high accuracy, tumor segmentation performance was still influenced by PET noise and by the thresholding approach used to

generate ground truth. Third, we were unable to compare our results directly with TotalSegmentator because our ground-truth labels include the intrahepatic segments of the inferior vena cava and the portal vein, which TotalSegmentator delineates as separate structures. Finally, although our test set contained a range of primary and metastatic lesions, sample sizes for certain tumor subtypes were relatively small, which may limit the generalizability of our findings across all disease presentations.

## Conclusion

Both the SwinUNETR and residual UNET models achieved excellent accuracy for liver segmentation and high performance for hepatic tumor segmentation on ^{18}F-FDG PET/CT, with most volumetric measurements falling within clinically acceptable limits. While SwinUNETR demonstrated slightly superior performance, both architectures showed potential for integration into clinical workflows and research pipelines. Given their open-source availability and adaptability, these models could support automated, reproducible segmentation in treatment planning and quantitative imaging.

### Ethics

**Ethics Committee Approval:** The study was approved by the Ethics Committee of Harran University (approval no: HRÜ-25.11.02, date: 16.06.2025), with additional approval from the institutional review board.

**Informed Consent:** Written informed consent was obtained from all patients before imaging.

### Footnotes

### Authorship Contributions

Surgical and Medical Practices: B.D., Concept: B.D., Design: B.D., Data Collection or Processing: B.D., H.K.Y., F.E., M.A.A., Analysis or Interpretation: B.D., H.K.Y., F.E., M.A.A., Literature Search: B.D., Writing: B.D.

**Conflict of Interest:** No conflict of interest was declared by the authors.

**Financial Disclosure:** The authors declared that this study has received no financial support.

### References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71:209-249.

2. Ntanasis-Stathopoulos I, Liatsou E, Dedes N, Triantafyllakis K, Tsilimigras DI, Gavriatopoulou M. Multidisciplinary management of colorectal liver metastases. In: Interdisciplinary cancer research. Springer Nature; 2024.

3. Van Cutsem E, Cervantes A, Adam R, Sobrero A, Van Krieken JH, Aderka D, Aranda Aguilar E, Bardelli A, Benson A, Bodoky G, Ciardiello F, D'Hoore A, Diaz-Rubio E, Douillard JY, Ducreux M, Falcone A, Grothey A, Gruenberger T, Haustermans K, Heinemann V, Hoff P, Köhne CH, Labianca R, Laurent-Puig P, Ma B, Maughan T, Muro K, Normanno N, Österlund P, Oyen WJ, Papamichael D, Pentheroudakis G, Pfeiffer P, Price TJ, Punt C, Ricke J, Roth A, Salazar R, Scheithauer W, Schmoll HJ, Tabernero J, Taïeb J, Tejpar S, Wasan H, Yoshino T, Zaanan A, Arnold D. ESMO consensus guidelines for the management of patients with metastatic colorectal cancer. Ann Oncol. 2016;27:1386-1422.

4. Clark AM, Ma B, Taylor DL, Griffith L, Wells A. Liver metastases: microenvironments and ex-vivo models. Exp Biol Med (Maywood). 2016;241:1639-1652.

5. Weber M, Lam M, Chiesa C, Konijnenberg M, Cremonesi M, Flamen P, Gnesin S, Bodei L, Kracmerova T, Luster M, Garin E, Herrmann K. EANM procedure guideline for the treatment of liver cancer and liver metastases with intra-arterial radioactive compounds. Eur J Nucl Med Mol Imaging. 2022;49:1682-1699.

6. Chiesa C, Sjogreen-Gleisner K, Walrand S, Strigari L, Flux G, Gear J, Stokke C, Gabina PM, Bernhardt P, Konijnenberg M. EANM dosimetry committee series on standard operational procedures: a unified methodology for ^{99m}Tc-MAA pre- and 90Y peri-therapy dosimetry in liver radioembolization with 90Y microspheres. EJNMMI Phys. 2021;8:77.

7. Dawson LA, Ten Haken RK. Partial volume tolerance of the liver to radiation. Semin Radiat Oncol. 2005;15:279-283.

8. Hwang SH, Hong HS, Kim D, Kim GM, Lee HW, Lee M, Kim DY, Park MA, Yun M. Total lesion glycolysis on 18F-FDG PET/CT is a better prognostic factor than tumor dose on 90Y PET/CT in patients with hepatocellular carcinoma treated with 90Y transarterial radioembolization. Clin Nucl Med. 2022;47:e437-e443.

9. Demir B, Soydal C, Kucuk NO, Celebioglu EC, Bilgic MS, Kuru Oz D, Elhan AH, Kir KM. Voxel-based dosimetry with integrated Y-90 PET/MRI and prediction of response of primary and metastatic liver tumors to radioembolization with Y-90 glass microspheres. Ann Nucl Med. 2025;39:31-46.

10. 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. Arxiv. 2022.

11. He Y, Nath V, Yang D, Tang Y, Myronenko A, Xu D. SwinUNETR-V2: stronger swin transformers with stagewise convolutions for 3D medical image segmentation. Lecture Notes in Computer Science. 2023.

12. Kerfoot E, Clough J, Oksuz I, Lee J, King AP, Schnabel JA. Left-ventricle quantification using residual U-Net. Lecture Notes in Computer Science. 2019.

13. Wasserthal J, Breit HC, Meyer MT, Pradella M, Hinck D, Sauter AW, Heye T, Boll DT, Cyriac J, Yang S, Bach M, Segeroth M. TotalSegmentator: robust segmentation of 104 anatomic structures in CT images. Radiol Artif Intell. 2023;5:e230024.

14. Yashaswini GN, Manjunath RV, Shubha B, Prabha P, Aishwarya N, Manu HM. Deep learning technique for automatic liver and liver tumor segmentation in CT images. Journal of Liver Transplantation. 2025;17:100251.

15. Li X, Chen H, Qi X, Dou Q, Fu CW, Heng PA. H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes. IEEE Trans Med Imaging. 2018;37:2663-2674.

16. Ghofrani F, Behnam H, Motlagh HDK. Liver segmentation in CT images using deep neural networks. 2020 28th Iranian Conference on Electrical Engineering, ICEE. 2020.

17. Demir B, Soydal C, Kucuk NO, Celebioglu EC, Bilgic MS, Kuru Oz D, Elhan AH, Kir KM. Voxel-based dosimetry with integrated Y-90 PET/MRI and prediction of response of primary and metastatic liver tumors to radioembolization with Y-90 glass microspheres. Ann Nucl Med. 2025;39:31-46.

Demir et al. Automated Segmentation of Liver and Tumors

Mol Imaging Radionucl Ther 2026;35(1):10-18

18. 4burakfe/SlicerAether: This extension allows users to utilize SwinUNETR and UNET models to segment multimodality images. Available from: https://github.com/4burakfe/SlicerAether/tree/main

19. Claritas/Segmentation Edition at main · 4burakfe/Claritas. Available from: https://github.com/4burakfe/Claritas/tree/main/Segmentation%20 Edition

20. Release Sample Cases · 4burakfe/SlicerAether. Available from: https://github.com/4burakfe/SlicerAether/releases/tag/Sample_Cases

21. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, Bauer C, Jennings D, Fennessy F, Sonka M, Buatti J, Aylward S, Miller JV, Pieper S, Kikinis R. 3D slicer as an image computing platform for the quantitative ımaging network. Magn Reson Imaging. 2012;30:1323-1341.

22. 3D Slicer image computing platform | 3D Slicer. Available from: https://www.slicer.org/

23. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, Lin S, Guo B. Swin transformer: hierarchical vision transformer using shifted windows. IEEE/CVF International Conference on Computer Vision (ICCV). 2021.

24. Wang Z, Popordanoska T, Bertels J, Lemmens R, Blaschko MB. Dice semimetric losses: optimizing the dice score with soft labels. lecture notes in computer science. 2023.

25. Wang Z, Ning X, Blaschko MB. Jaccard metric losses: optimizing the jaccard ındex with soft labels. Adv Neural Inf Process Syst. 2023.

26. Milletari F, Navab N, Ahmadi SA. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. 2016 Fourth International Conference on 3D Vision IEEE. 2016.

27. Goedhart J, Rishniw M. BA-plotteR - A web tool for generating Bland-Altman plots and constructing limits of agreement. Res Vet Sci. 2021;137:281-286.

28. R Core Team. R: A Language and environment for statistical computing. Foundation for Statistical Computing. 2022.

29. Heimann T, Van Ginneken B, Styner MA, Arzhaeva Y, Aurich V, Bauer C, Beck A, Becker C, Beichel R, Bekes G, Bello F de, Binnig G, Bischof H, Bornik A, Cashman PM, Chi Y, Cordova A, Dawant BM, Fidrich M, Furst JD, Furukawa D, Grenacher L, Hornegger J, Kainmuller D, Kitney RI, Kobatake H, Lamecker H, Lange T, Lee J, Lennon B, Li R, Li S, Meinzer HP, Nemeth G, Raicu DS, Rau AM, Rikxoort EM van, RoussonM, Rusko L, Saddi KA, Schmidt G, Seghers D, Shimizu A, Slagmolen P, Sorantin E, Soza G, Susomboon R, Waite JM, Wimmer A, Wolf I. Comparison and evaluation of methods for liver segmentation from CT datasets. IEEE Trans Med Imaging. 2009;28:1251-1265.

30. Meng L, Tian Y, Bu S. Liver tumor segmentation based on 3D convolutional neural network with dual scale. J Appl Clin Med Phys. 2020;21:144-157.

31. Christ PF, Elshaer MEA, Ettlinger F, Tatavarty S, Bickel M, Bilic P, Rempfler M, Armbruster M, Hofmann F, D'Anastasi M, Sommer WH, Ahmadi SA, Menze BH. Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields. Lecture Notes in Computer Science. 2016.

32. Leung KH, Rowe SP, Sadaghiani MS, Leal JP, Mena E, Choyke PL, Du Y, Pomper MG. Deep semisupervised transfer learning for fully automated whole-body tumor quantification and prognosis of cancer on PET/CT. J Nucl Med. 2024;65:643-650.

33. Luo Y, Yang Q, Hu J, Qin X, Jiang S, Liu Y. Preliminary study on detection and diagnosis of focal liver lesions based on a deep learning model using multimodal PET/CT images. Eur J Radiol Open. 2024;14:100624.