Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 8, No. 3, July 2026, pp: 983-999 e-ISSN: 2656-8632

RESEARCH ARTICLE

OPEN ACCESS

Hybrid CNN-Transformer Architecture for Robust Liver Tumor Segmentation in 2D CT Slices

Huda Dham Bader[1], and Mohammed Sabah Jarjees[2]

[1] Department of Medical Physiology, College of Medicine, University of Mosul, Mosul, Iraq
[2] Department of Medical Instrumentation Techniques Engineering, Technical Engineering College of Mosul, Northern Technical University, Mosul, Iraq

Corresponding author: Huda Dham Bader (e-mail: huda.badr@uomosul.edu.iq), **Author(s) Email**: Mohammed Sabah Jarjees (e-mail: mohammed.s.jarjees@ntu.edu.iq)

Abstract **Background: Liver tumor segmentation from CT scans is a task affected by class imbalance, low contrast, and small lesion size. Manual segmentation is time-consuming and also suffers from inter-observer variability. Methods: We propose a 2D CNN-Transformer model with 20.3M parameters in an encoder–decoder structure with four transformer layers (8 heads, 2048 feedforward dimension). The model processes 2D axial slices due to GPU memory limits. The loss function combines Cross-Entropy, Dice, and Focal losses with α = 0.25 and γ = 2.0. Preprocessing includes CLAHE (clip limit = 2.0, 8×8 tiles) and gamma correction (γ = 1.2). From the LiTS dataset (131 volumes), 11 volumes with 1,688 slices were selected based on tumor presence, annotation quality, and artifact removal. A patient-level split of 80% for training, 10% for validation, and 10% for testing was used to prevent data leakage. Results: The model achieved liver Dice = 0.916 ± 0.122 and tumor Dice = 0.810 ± 0.304. The 95% confidence intervals using bootstrapping (1,000 resamples) were [0.897–0.934] for liver and [0.765–0.856] for tumor. Best validation results at Epoch 98 were liver Dice = 0.938, tumor Dice = 0.823, and accuracy = 0.992. Pixel accuracy was 99.20% and was not used as the main metric due to class imbalance, where background pixels exceed 90%. An ablation study showed that CLAHE and gamma correction improved tumor Dice by 8.6% and liver Dice by 3.3% compared to a baseline without preprocessing. Conclusion: The model shows performance for liver tumor segmentation on a LiTS subset. External validation on the full dataset and multi-center data is required before clinical use.**

Keywords **Medical image segmentation; CNN; Transformer architecture; Liver CT imaging; Tumor detection.**

I. Introduction

Hepatocellular malignant tumors are a major health issue of concern, and their survival rates are very poor, owing to their diagnosis when the disease is in its advanced stages [1]. In the world, liver cancer was in the sixth position with more than 905,000 cases and 830,000 deaths in 2022. The highest mortality rates were witnessed in Mongolia and Egypt in the entire world [2]. These statistics indicate the importance of early detection, based on suitable liver and tumor segmentation [3]. The studies on hepatic structures and pathological masses have increased significantly over the past few years. Research in this field has grown twice; however, this is a challenging goal [4]. In the computed tomography images, the hepatic parenchyma is barely differentiated from the neighboring anatomies, making it difficult to identify pixels of normal and abnormal hepatic parenchyma. Proper delineation of hepatic pathology is challenging because lesions differ significantly in size, structure, and location among patients [5]. Also, not all pathological masses are well demarcated from their surrounding tissue [6], which makes it even more difficult to identify them using the conventional boundary-detection segmentation method. In recent years, deep learning has revolutionized the analysis of medical images. Automated segmentation is now possible with amazing accuracy using convolutional neural networks. The U-Net architecture served as a base system for biomedical image segmentation [7]. Thereafter, many variants were developed in order to overcome certain constraints. Nevertheless, CNNs are intrinsically suited to long-range spatial dependencies [8]. Transformer-based architectures provide a remedy for the self-attention components. Nonetheless, pure transformers require significant computing power. A combination of the two architectures is promising [9].

Manuscript received 7 March 2026; Revised 30 April 2026; Accepted 20 May 2026; Available online 2 June 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1602

Several studies have addressed liver and tumor segmentation with varying success. Özcan et al. [10] introduced AIM-Unet for automatic segmentation from CT images. Their model achieved 97.86% Dice for liver and 75.6% tumor Dice on the LiTS dataset. Zheng et al. [11] proposed a 4D model based on 3D CNN and ConvLSTM for DCE-MRI, achieving 82.5% tumor Dice. Shao et al. [12] introduced AC-Net, incorporating axial attention and vision transformer modules. Their model achieved 90.0% Dice score for CT segmentation and 80.0% for MRI segmentation. Sabir et al. [13] combined residual blocks with the U-Net architecture in ResU-Net. They achieved 83% tumor Dice on the 3D-IRCADb01 dataset. Budak et al. [14] proposed cascaded deep encoder-decoder CNNs using a two-stage approach. Their method reported 95.22% liver Dice on the 3DIRCADb dataset. Hettihewa et al. [15] developed MANet with multi-attention mechanisms. This architecture achieved 74.8% tumor segmentation accuracy. More recently, Wang et al. [16] introduced HyborNet, combining Gabor attention with transformer interactions. Despite achieving 92.5% liver Dice, tumor performance remained limited at 55.5%. Research Gap and Novelty: Despite recent progress, limitations still exist in this field. Many models fail to detect small liver tumors due to an imbalance between the background, liver, and tumor regions. TransUNet and Swin-Unet are designed for general medical image segmentation and are not tailored for liver CT scans[17][18] . Previous studies have used basic preprocessing methods that do not improve CT contrast. Common loss functions such as Dice and Cross-Entropy do not handle small lesions well or balance pixel accuracy with region overlap.

This study addresses these gaps through: (1) a 2D CNN-Transformer architecture optimized for slice-

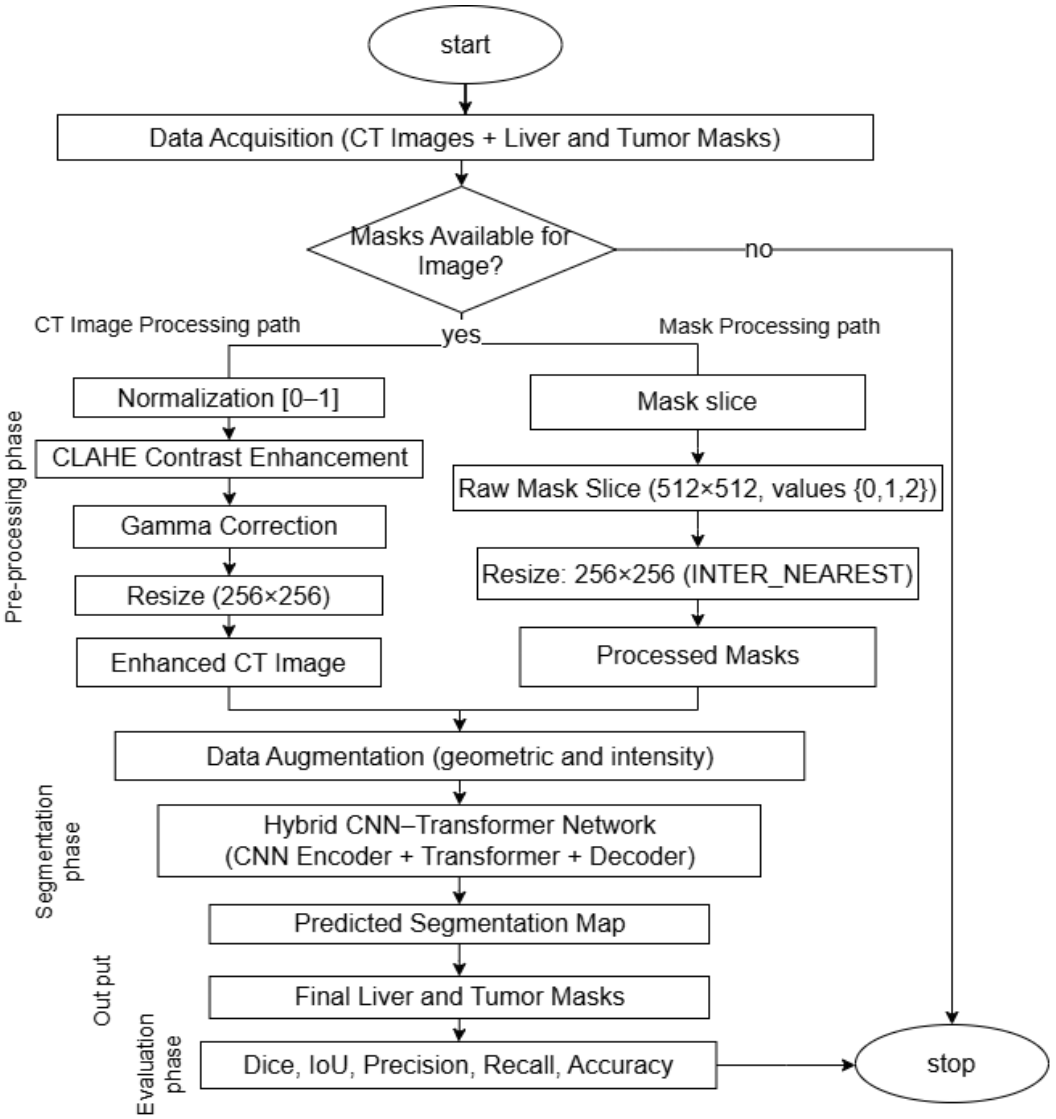


Fig. 1. workflow of the proposed liver tumor segmentation model.

Manuscript received 7 March 2026; Revised 30 April 2026; Accepted 20 May 2026; Available online 2 June 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1602

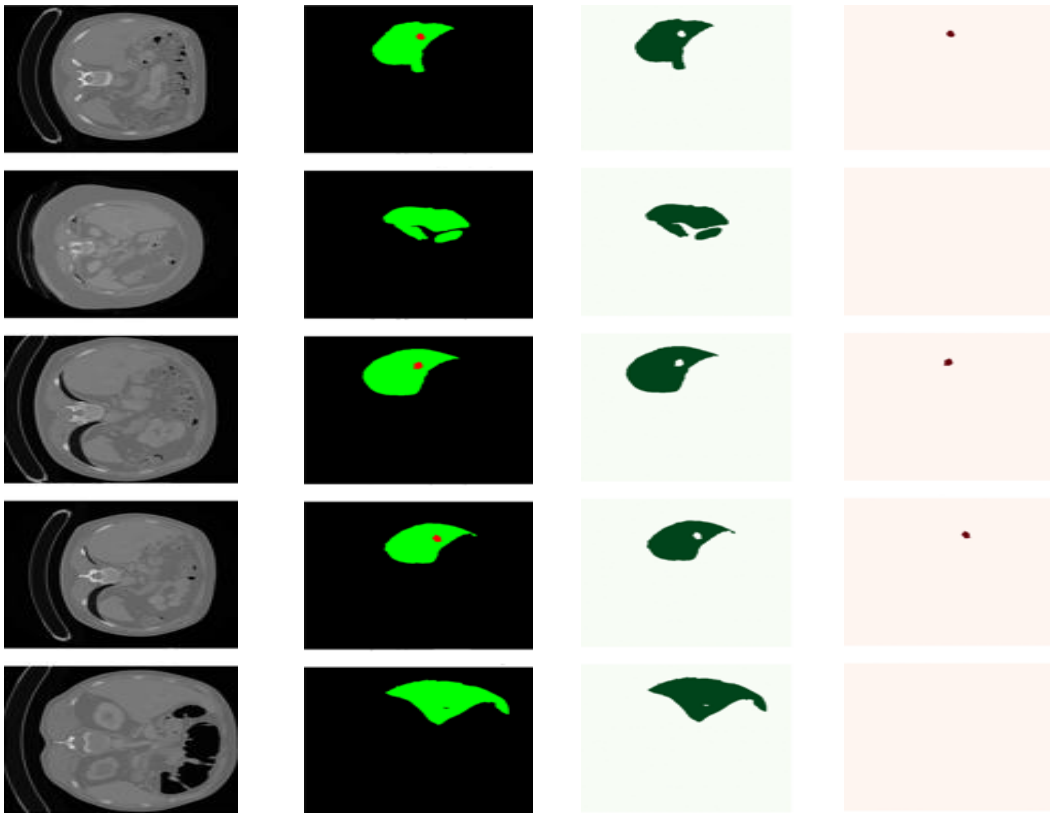


Fig. 2. Samples from the dataset: (a) original CT image, (b) combined segmentation mask (green: liver, red: tumor), (c) liver mask, and (d) tumor mask.

based liver tumor segmentation; (2) CLAHE and gamma correction specifically tuned for hypo-dense tumor visualization; (3) a mixed loss function with focal loss parameters $\alpha = 0.25$ and $\gamma = 2.0$, and (4) rigorous patient-level validation. While the LiTS dataset provides true volumetric CT scans, the proposed architecture operates on individual 2D axial slices rather than full 3D volumes. This design choice is motivated by computational constraints (a Tesla T4 GPU with 16 GB of memory) and clinical practicality, as many PACS systems display and process CT scans slice-by-slice. Therefore, the term "volumetric" in the title has been revised to "2D CT slices" to accurately reflect the implementation.

In this work, we propose a better hybrid architecture between a CNN and a Transformer for more accurate segmentation of liver lesions and tumors by taking advantage of CNN's convolutional-based feature extractor layer combined with the multi-head self-attention mechanism in Transformer. Moreover, advanced pre-processing techniques, including normalization, CLAHE, and gamma correction, enhance tumor visibility. An augmentation pipeline is used for model generalization. Class imbalance can be minimized by utilizing the mixed loss function of CE, Dice, and Focal losses, as rewards are not even for each class. We proposed a hybrid of the CNN and Transformer for key points inference that outperformed state-of-the-art methods, with the dice score of 82.26% for tumor segmentation. Normalization, CLAHE, and gamma rescaling are incorporated into the preprocessing to improve tumor visibility and feature separability. More than 10 augmentations are adopted to boost the robustness of models. Focal loss, Dice, and Cross-Entropy are combined in our loss function to deal with the class imbalance. The model's ability to retain the delineation accuracy of tumors for different types of lesions, with a dice measured at 93.82% for liver segmentation is demonstrated.

II. Method

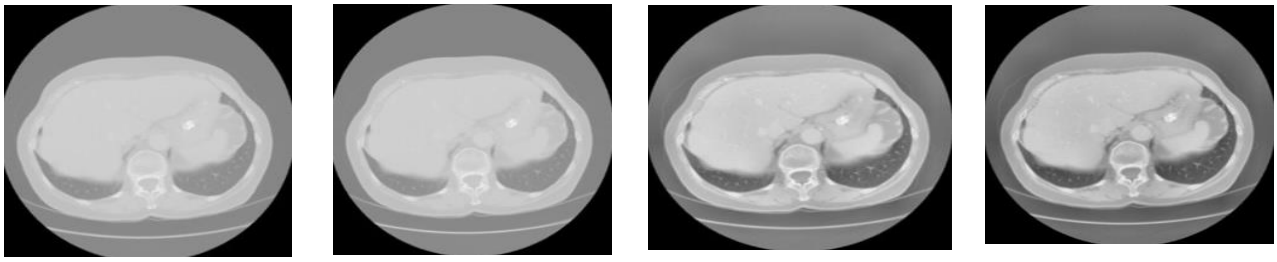
This section covers the dataset, preprocessing steps, segmentation approach, and evaluation metrics. Fig. 1 shows the overall model pipeline.

A. Dataset

The LiTS (Liver Tumor Segmentation) dataset [17] serves as a standard benchmark for liver and tumor segmentation, containing 131 contrast-enhanced abdominal CT scans from clinical institutions worldwide. Each CT scan is stored in NIfTI format. Ground truth annotations include pixel-wise labels for three classes: background (0), liver parenchyma (1),

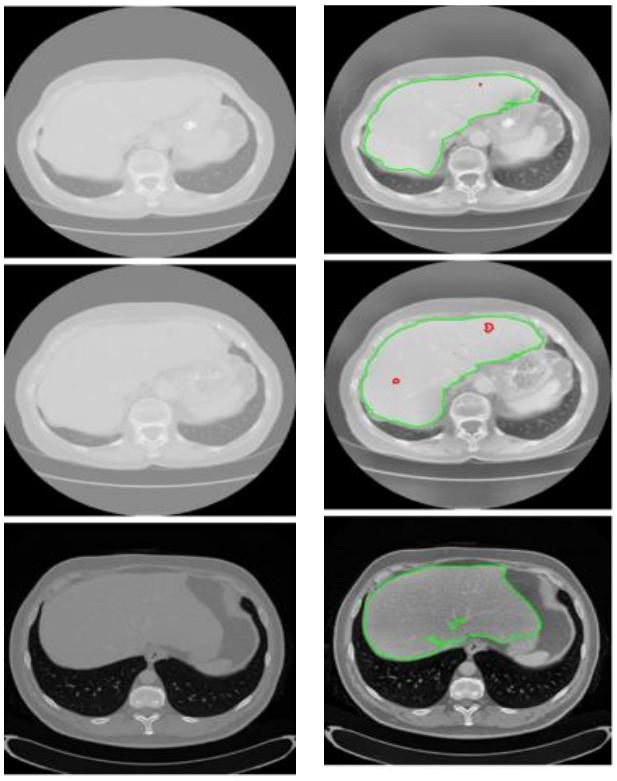
Manuscript received 7 March 2026; Revised 30 April 2026; Accepted 20 May 2026; Available online 2 June 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1602

and tumor tissue (2). Selection Criteria and Rationale for Subsampling: From the full 131 volumes, only 11 volumes (1,688 axial slices) were selected based on explicit inclusion criteria: (1) presence of at least one tumor annotation (n=108 volumes had tumors); (2) absence of severe motion artifacts or incomplete liver coverage upon visual inspection; (3) complete alignment between CT volume and segmentation mask slice-by-slice. The remaining volumes were excluded due to: empty slices without any annotation (to prevent artificial class imbalance), inconsistent mask quality (misaligned or missing slices), or insufficient tumor representation (<5 tumor slices per volume). A total of 2,847 slices (62.8% of all slices across the 11 selected volumes) were excluded because they contained no liver or tumor annotations. The authors acknowledge that the results are not directly comparable to those of methods evaluated on the full LiTS benchmark. Fig. 2 shows an example of the raw dataset representing the anatomical diversity and quality of annotations in various patients. The visualization involves CT slices, combined segmentation masks with liver drawn in green and tumors in red, and a separate class mask, which shows the spatial distribution of the hepatic structures. Patient-Level Splitting Protocol: To prevent data leakage and ensure clinically valid generalization estimates, the dataset was split at the patient (volume) level rather than the slice level. The 11 CT volumes were randomly partitioned (stratified by tumor size) as follows: Training: 8 volumes (1,350 slices, 80%), Validation: 1 volume (168 slices, 10%), Testing: 2 volumes (170 slices, 10%). This ensures no patient



(a) (b) (c) (d)

Fig. 3. CT image enhancement pipeline: (a) original CT image, (b) normalized image, (c) CLAHE-enhanced image, and (d) gamma-corrected image.



(a) (b)

Fig. 4. Comparison of original and preprocessed CT images: (a) raw CT slices and (b) enhanced slices.

Manuscript received 7 March 2026; Revised 30 April 2026; Accepted 20 May 2026; Available online 2 June 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1602

overlap between splits. A fixed seed (42) was used for reproducibility, replacing slice-level splitting.

B. Preprocessing

To improve the quality of CT images, a preprocessing pipeline was developed that addresses imaging issues and increases tumor visibility. Every CT image was processed in a series of sequential steps. To prevent class imbalance during training, the dataset was filtered to include only images with segmentation masks. Images without masks were not included in the dataset [18]. After that, all of the images were normalized for intensity. In order to guarantee consistency, the normalization uses min-max scaling to scale pixel values to the [0, 1] range. Additionally, it makes stable processing easier in later stages. Better input for CLAHE enhancement is provided by normalized images. CLAHE was applied to enhance local contrast [19]. This method prevents excessive noise in homogeneous regions by computing histograms for small image tiles with a clip limit of 2.0. The tile grid was set to 8 × 8 pixels. CLAHE highlights anatomical features by addressing uneven illumination across various acquisitions. This strategy shows subtle differences between tumors and livers. The CLAHE enhancement phase was followed by gamma correction. Using the transformation equation, $I' = I^{\gamma}$, where I represents the input intensity and I' represents the output intensity [20], a gamma value of 1.2 was applied to the pictures, where Values greater than 1.0 increase the contrast of mid-tones, make features more visible, and make hypo-dense tumors easier to see. In the original images, such tumors are often subtle. Intensity variations are more pronounced under gamma correction. Ablation results showed that removing gamma correction led to a drop in segmentation performance, indicating its role in improving tumor/liver contrast. To ensure the input dimensions remain consistent, the final step is to resize images to 256 x 256 pixels. For the CT images, bilinear interpolation was applied to maintain smooth intensity variations. The full pipe generates improved CT images. These images are trainable and evaluable. The entire steps of the pipeline are illustrated in Fig. 3.

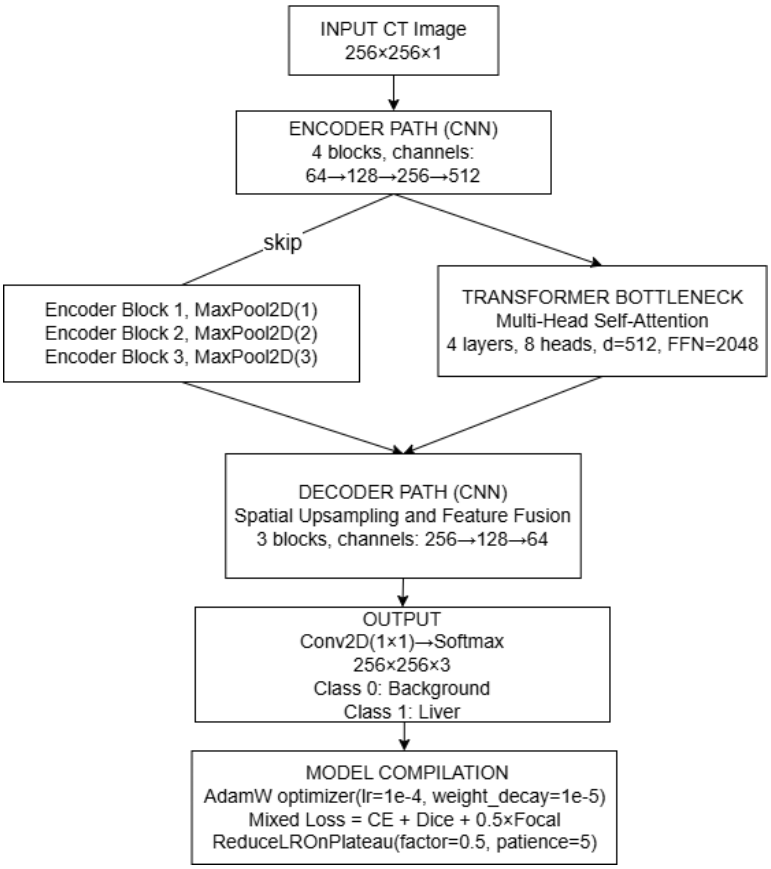


Fig. 5. Encoder, transformer, and decoder proposed model flowchart.

It shows normalization, CLAHE, and gamma. Fig. 4 demonstrates crude and processed comparisons. The original buys are clearly shown in the left columns, and the enhanced preprocessed images are presented in the right columns with anatomical faithfulness.

C. Data Augmentation

Data augmentation is essential to deep learning. It eliminates overfitting as the problem of neural networks is prone to. The medical imaging is of particular interest

Manuscript received 7 March 2026; Revised 30 April 2026; Accepted 20 May 2026; Available online 2 June 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1602

to this technique. Medical data is also usually small and costly to obtain. An elaborate augmentation pipeline was applied here. This was done by using the Augmentations library [21]. It is a library that provides quick, adaptable image manipulation. Transformations were applied probabilistically during training only. Both geometric and intensity transformations were included. Validation and testing data remained unaugmented throughout. Geometric transformations simulate spatial variations in clinical practice. Clinical Justification of Augmentation Transformations: Each augmentation was selected to reflect clinically realistic variations:

a) Horizontal Flip (p=0.5): Simulates supine vs. prone positioning
b) Vertical Flip (p=0.3): Accounts for variable patient orientation (applied conservatively)
c) Rotation (±25°, p=0.5) and Shift Scale Rotate (p=0.5): Model patient positioning differences
d) Elastic Transform (α=1, σ=50, p=0.3): Simulates non-rigid deformation from breathing motion
e) Random Brightness and Contrast (±20%, p=0.5): Replicates CT scanner and tube current variations
f) Gaussian Noise (p=0.3): Models quantum noise in low-dose CT protocols
g) Gaussian Blur (p=0.2): Simulates slight patient motion
h) Coarse Dropout (p=0.3): Encourages reliance on contextual information

All transformations were applied only during training. Validation and testing data remained unaugmented.

D. Proposed Architecture

The architecture operates on 2D slices (256×256×1 input) and produces pixel-wise class predictions (256×256×3). This study proposes a hybrid model combining CNNs and transformers. Local and global features are captured simultaneously through this design. The architecture follows an encoder-decoder paradigm with skip connections. Fig. 5 illustrates its three main components: encoder, transformer, and decoder.

Encoder: Hierarchical features are extracted from CT images in four stages. Every block includes a double convolution and max pooling, which are based on U-Net [22]. U-Net inspired our encoder [23]. Block convolution doubling of the basic building unit of our encoder. The two identical 3×3 convolutions are applied one right after the other, batch normalization [24], and ReLU activation [25] come after each convolution. Batch normalization is defined in Eq. (1), while ReLU activation is given in Eq. (2):

$$\hat{x} = \frac{x-\mu_B}{\sqrt{\sigma_B^2+\epsilon}}, y = \gamma\hat{x} + \beta \quad (1)$$

where x is the input, μ_B is the batch mean, σ_B^2 is the batch variance, ε = 1×10⁻⁵, and γ, β are learnable. The output **y** is the normalized and transformed feature map.

$$\mathrm{ReLU}(x) = \max(0, x) \quad (2)$$

where **x** is the input value, and the output is the maximum of zero and the input value. The convolution operation is formulated in Eq. (3):

$$(I * K)(x, y) = \sum_{i=-k}^{k} \sum_{j=-k}^{k} I\,(x+i, y+j) \cdot K(i, j) \quad (3)$$

where I is the input feature map, K is the convolution kernel, k is the kernel size (k = 3 in our implementation). This combination will make the training stable and non-linear. 256x256 pixel CT scans go into the first stage. Channels expand from 1 to 64 through convolution, and 2×2 max pooling then halves the spatial dimensions. In the second stage, channels increase from 64 to 128 at 128×128 resolution, while the third stage produces 256 channels at 64×64 resolution. The fourth stage outputs 512 channels at 32×32, representing the most abstract semantic features.

Transformer Module: A fundamental limitation of CNNs is their localized receptive fields [26]. Transformers enhance medical segmentation [27]. Local features are captured well, but long-range dependencies remain unreachable. Self-attention mechanisms in the transformer module address this limitation [28]. At the network bottleneck, features with 32×32 spatial dimensions and 512 channels are reshaped into 1024 sequential tokens. Each token has 512 dimensions, corresponding to a single spatial position. The transformer encoder consists of four stacked layers, each containing multi-head self-attention and a feed-forward network. Attention enables every position to attend to all others through Query, Key, and Value matrices. The scaled dot-product attention is computed as shown in Eq. (4):

$$\mathrm{Attention}(Q, K, V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (4)$$

where Q, K, and V denote query, key, and value matrices. Eight parallel attention heads capture diverse relationship types independently. Their outputs are concatenated to form the final representation as expressed in Eq. (5):

$$\mathrm{MultiHead}(Q, K, V) = \mathrm{Concat}(head_1, \ldots, head_8)W^O \quad (5)$$

where head_i is the output of the i-th attention head, and W^O is the output matrix. A feed-forward network with 2048 hidden dimensions follows the attention mechanism. Residual connections [29], layer normalization [24], and dropout at a 0.1 rate [30] ensure stable training. It was experimentally demonstrated that four layers and eight heads are the most balanced in terms of global context capturing and computational cost. The feed-forward network is defined in Eq. (6):

$$\mathrm{FFN}(x) = \mathrm{GELU}(xW_1 + b_1)W_2 + b_2 \quad (6)$$

where W_1, W_2 are weights and b_1, b_2 are biases.

Layer normalization is defined in Eq. (7).

Manuscript received 7 March 2026; Revised 30 April 2026; Accepted 20 May 2026; Available online 2 June 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1602

Table 1. Summary of the Proposed Hybrid CNN–Transformer Network Components.

Component	Layer/Operation	Details	Output Shape	Purpose
Encoder	Input Layer	Input CT image (grayscale)	256 × 256 × 1	Receives input
	Encoder Block 1	DoubleConv (Conv2D 3×3, BN, ReLU) × 2, 64 filters	256 × 256 × 64	Low level feature extraction
	MaxPool2D (1)	Kernel: 2×2, Stride: 2	128 × 128 × 64	Spatial downsampling
	Encoder Block 2	DoubleConv (Conv2D 3×3, BN, ReLU) × 2, 128 filters	128 × 128 × 128	Extracts mid features
	MaxPool2D (2)	Kernel: 2×2, Stride: 2	64 × 64 × 128	Downsamples
	Encoder Block 3	DoubleConv (Conv2D 3×3, BN, ReLU) × 2, 256 filters	64 × 64 × 256	Extracts high features
	MaxPool2D (3)	Kernel: 2×2, Stride: 2	32 × 32 × 256	Downsamples
	Encoder Block 4	DoubleConv (Conv2D 3×3, BN, ReLU) × 2, 512 filters	32 × 32 × 512	Deepest feature extraction
Skip Connections	enc1, enc2, enc3	Feature maps from encoder blocks 1, 2, 3	Various	Keeps spatial info
Transformer	Flatten + Permute	Reshape spatial features to sequence	1024 × 512	Prepares attention
	Positional Encoding	Learnable position embeddings	1024 × 512	Adds position data
	Transformer Encoder L1, L2, L3, L4	Multi-Head Attention + FFN, 8 heads, d_model=512	1024 × 512	global context capturing
	Permute + Reshape	Restore spatial structure	32 × 32 × 512	Returns to spatial
Decoder	ConvTranspose2D (1)	Kernel: 2×2, Stride: 2, 256 filters	64 × 64 × 256	Upsamples
	Concatenate (1)	Concatenate with enc3 skip connection	64 × 64 × 512	Combines encoder data
	Decoder Block 1	DoubleConv (Conv2D 3×3, BN, ReLU) × 2, 256 filters	64 × 64 × 256	Refines features
	ConvTranspose2D (2)	Kernel: 2×2, Stride: 2, 128 filters	128 × 128 × 128	Upsamples
	Concatenate (2)	Concatenate with enc2 skip connection	128 × 128 × 256	Combines encoder data
	Decoder Block 2	DoubleConv (Conv2D 3×3, BN, ReLU) × 2, 128 filters	128 × 128 × 128	Refines features
	ConvTranspose2D (3)	Kernel: 2×2, Stride: 2, 64 filters	256 × 256 × 64	Upsamples
	Concatenate (3)	Concatenate with enc1 skip connection	256 × 256 × 128	Combines encoder data
	Decoder Block 3	DoubleConv (Conv2D 3×3, BN, ReLU) × 2, 64 filters	256 × 256 × 64	Final refinement
Output	Conv2D	Kernel: 1×1, 3 filters	256 × 256 × 3	Generates logits
	Softmax	Class probability normalization	256 × 256 × 3	Outputs mask

$$\text{LayerNorm}(x) = \frac{x-\mu}{\sqrt{\sigma^2+\epsilon}} \odot \gamma + \beta \quad (7)$$

where μ and σ^2 denote mean and variance.

Decoder: Segmentation masks at high resolution are reconstructed gradually across three stages of decoders. The 2x2 and 2x2 kernel transposed convolution with a stride of 2 is used to learnable up-sampling [31], doubling the spatial dimensions at each step. Connection between the similar encoder stages, as well as the decoder features, is concatenated [23]. The output size of upsampling is given in Eq. (8):

$$H_{out} = (H_{in} - 1)s - 2p + k \quad (8)$$

where H_{in} and H_{out} represent input and output sizes. Skip connections are defined in Eq. (9):

$$F_{dec} = \text{Concat}(F_{enc}, F_{up}) \quad (9)$$

where F_{enc} and F_{up} denote encoder and upsampled features. This design maintains finer boundary knowledge that is lost in the course of encoding. The transformer output consists of 512 channels, which the first stage of decoding upsamples to 256 channels at 64. The features of the encoder stage 3 are concatenated and double convolution is used to reconstruct the concatenated 512 channels to 256. The same operation is repeated in the second phase (128

Table 2. Loss Function Components.

Loss	Formula	Purpose
Cross-Entropy	$L_{CE} = -\sum_c y_c \log(\hat{y}_c)$	Measures prediction error a true label with stable early gradients
Dice Loss	$L_{Dice} = 1 - \frac{2 \mid S_p \cap S_g \mid}{\mid S_p \mid + \mid S_g \mid}$	$P \cap g$: Maximizes overlap between predicted and ground-truth masks
Focal Loss	$L_{Focal} = -\alpha(1-p_t)^{\gamma}\log(p_t)$	Focuses on hard examples
Total	$L_{total} = L_{CE} + L_{Dice} + 0.5 \times L_{Focal}$	Combined optimization

$\boldsymbol{y_c}, \hat{\boldsymbol{y}}_{\boldsymbol{c}}$: true and predicted probabilities for class $\boldsymbol{c}$
$\boldsymbol{P}, \boldsymbol{G}$: predicted and ground truth masks
$\boldsymbol{\alpha}, \boldsymbol{\gamma}$: focal loss parameters (weighting and focusing)

Manuscript received 7 March 2026; Revised 30 April 2026; Accepted 20 May 2026; Available online 2 June 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1602

by 128, channels 128) and the third phase (256 by 256, channels 64). The output layer is a 1by 1 convoluted layer that reduces 64 channels into 3 probabilities of classes. Softmax activation is used to differentiate background, liver, and tumor classes [32]. The entire architecture is summed up in Table 1.

E. Mixed Loss Function

The proposed mixed loss function is defined in Eq. (10).

$$L_{total} = L_{CE} + L_{Dice} + 0.5 \times L_{Focal} \quad (10)$$

and every component plays a definite and significant role in the optimization process [33]. The conventional Cross-Entropy loss (L_{CE}) gives consistent gradients in the initial stages of the training process when the network outputs are considerably different from the ground truth, which guarantees stable convergence [34]. The Dice loss (L_{Dice}) directly optimizes the overlap between predicted and ground-truth segmentations, making it effective for imbalanced datasets where foreground classes constitute a small fraction of the total pixels and contribute minimally to cross-entropy gradients [35].

Table 2 shows the loss components for training the segmentation model, where γ is the focusing parameter and α is the balancing factor. The focusing parameter is set to γ = 2.0 in the Focal Loss (L_{Focal}), while the class balancing factor is set to α = 0.25 as a constant weight applied uniformly across classes. The model does not use explicit class weights in Cross-Entropy loss because class imbalance is already handled using Focal Loss (α = 0.25, γ = 2.0) and Dice Loss, which reduces the dominance of background pixels and focuses learning on difficult tumor region. A smoothing factor of $\epsilon = 1\times10^{-6}$ is added to the Dice Loss to prevent division by zero. This combination addresses class imbalance and improves segmentation accuracy for small, low-contrast tumors.

F. Evaluation Metric

Segmentation quality and computational efficiency. Accuracy was primarily measured with the Dice Similarity Coefficient (DSC) Eq. (11) [36], and the Jaccard Index, also known as Intersection over Union (IoU) Eq. (12) [36], which quantify how well the predicted segmentation S_p matches the ground truth S_g. Dice evaluates the overall overlap between the predicted and true regions, while Jaccard gives a stricter measure by comparing the shared area to the total combined area. In our experiments, pixel accuracy was 99.20%; however, this is not the preferred evaluation metric due to class imbalance (which most pixels are in the jbg). In this way, the model can achieve high accuracy while failing to properly detect tumors. So the Dice coefficient and IoU(Intersection over Union) are mainly used due to a more objective evaluation of segmentation performance.

$$\text{Dice} = \frac{2|S_p \cap S_g|}{|S_p| + |S_g|} \quad (11)$$

where $|S_p \cap S_g|$ is the intersection area between the predicted mask ($|S_p|$) ,and ground truth mask ($|S_g|$)

$$\text{Jaccard} = \frac{|S_p \cap S_g|}{|S_p \cup S_g|} \quad (12)$$

where $|S_p \cup S_g|$: is the union area between the predicted and ground truth masks.In addition, Sensitivity (Recall), Specificity, Precision, and F1-score Eqs. (13) to (15) were calculated to evaluate how well the model correctly identifies positive and negative pixels, and how reliable its positive predictions are. Sensitivity captures the model's ability to detect true positives, Specificity measures correct detection of negatives, Precision reflects the accuracy of positive predictions, and the F1-score balances Precision and

Table 3. Dataset Split Summary.

Split	Volume IDs	No. of Slices
Train	8 volumes	1,350
Val.	1 volume	168
Test	2 volumes	170

Sensitivity for a single comprehensive value.

$$Specificity = TN / (TN + FP) \quad (13)$$

$$Precision = TP / (TP + FP) \quad (14)$$

$$F1 - Score = 2TP / (2TP + FP + FN) \quad (15)$$

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively.

III. Results

A. Experimental Setup

We performed our experiment on a laptop with Windows 11 Pro, an Intel Core i7-13620H CPU, and 16 GB of RAM. The experiments were conducted on the Kaggle platform with free initialization of a GPU for deep learning. A single NVIDIA Tesla T4 GPU with 16 GB memory handled training and evaluation, and PyTorch 2.0 implemented the model. Training ran 100 epochs batch size was 8 samples based on GPU memory limits. Total training time was 4.02 hours. The AdamW optimizer managed weight updates throughout the training process. Initial learning rate was set to 1×10^{-4} with weight decay of 1×10^{-5} for regularization.

Table 4. Preprocessing Ablation Results.

Configuration	Liver Dice	Tumor Dice
Baseline (No CLAHE, No Gamma)	0.887	0.746
Ours (CLAHE + Gamma)	0.916	0.810
Improvement	+3.3%	+8.6%

Manuscript received 7 March 2026; Revised 30 April 2026; Accepted 20 May 2026; Available online 2 June 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1602

This optimizer combines adaptive learning rates with decoupled weight regularization effectively. Such a combination significantly enhances the generalization capability on medical imaging tasks. A ReduceLROnPlateau scheduler monitored validation loss during training continuously. When loss stopped improving for 5 consecutive epochs, the learning rate decreased by a factor of 0.5 automatically. This adaptive strategy helps the model escape local minima and achieve better convergence. Gradient clipping with a maximum norm of 1.0 prevented exploding gradients during backpropagation. Three complementary loss components formed the optimization objective function. Cross-entropy loss provided stable gradients during early training phases. Dice loss directly maximized overlap between predictions and ground truth masks. Focal loss with α=0.25 and γ=2.0 addressed class imbalance by emphasizing hard examples. The total loss was computed as Eq. (10). The LiTS benchmark dataset provided all experimental data for this study. This dataset contains 11 volumetric CT scans with expert annotations available. A total of 1,688 annotated slices were carefully extracted from these volumes. Slices without segmentation masks were excluded to prevent class imbalance issues. Random splitting partitioned the dataset into three separate subsets appropriately. The training data comprised 80% of the total slices, totaling approximately 1,350 samples. Each validation and test set received 10% of the data, or approximately 169 slices. This provides sufficient training data. The dataset split was performed at the patient (volume) level to avoid data leakage between training, validation, and testing sets, ensuring that slices from the same patient do not appear in multiple subsets. The dataset split is summarized in Table 3.

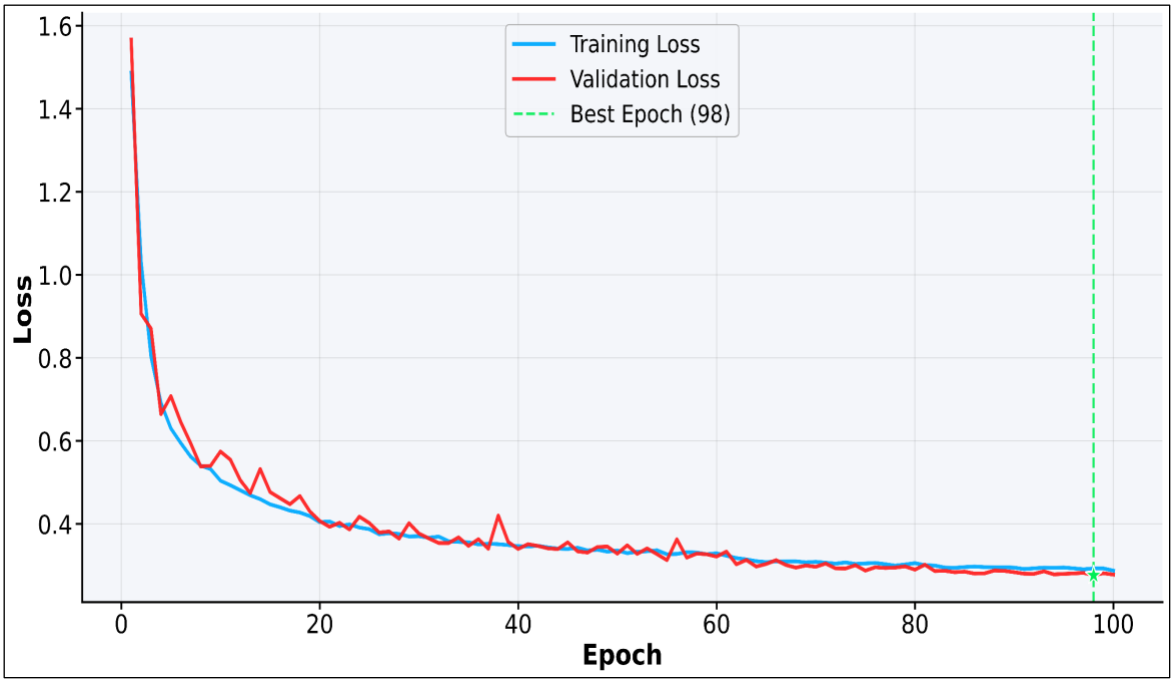


Fig. 6. Training and validation loss curves over 100 epochs.

B. Preprocessing Results

Our CT image Preprocessing pipeline was created. This addresses low-contrast medical imaging. Fig. 3 indicates raw and pre-processed samples. The process begins with normalization to [0, 1] by applying Min-max scaling to the pixel values. This ensures the steps that follow are consistent; otherwise, CLAHE and gamma correction do not work. CLAHE enhanced local contrast in the images Clip limit: 2.0, tile grid: 8×8 pixels. This reveals variations between hepatic tissue and tumor. The histograms are presented in Fig. 4. Brightness was modified using γ=1.2, and lastly, the images were adjusted to a uniform size of 256x256 pixels. Bilinear interpolation maintained smooth gradients, whereas nearest-neighbor respects the mask integrity, which enabled batch processing. As shown in Table 4, the proposed preprocessing (CLAHE + gamma correction) improved tumor Dice by 8.6% and liver Dice by 3.3%, demonstrating its effectiveness in enhancing low-contrast liver tumor segmentation.

C. Training Results

Measures were tracked over 100 training epochs, and the training process and analysis were stable. Summaries were saved after each epoch to analyze. Training and validation loss curves are in Fig. 6. In early periods, the two curves lowered gradually. Training loss reduced from 0.288 in epoch 1 to 0.288 in epoch 100, whereas Validation loss reduced from 1.567 to 0.278 within the same epochs. Minimal validation loss

Manuscript received 7 March 2026; Revised 30 April 2026; Accepted 20 May 2026; Available online 2 June 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1602

was achieved at epoch 98 with a value of 0.277, which was pointed out by a green marker. This is why the stabilized around 0.93 after epoch 80, indicating convergence was achieved. The tumor Dice score

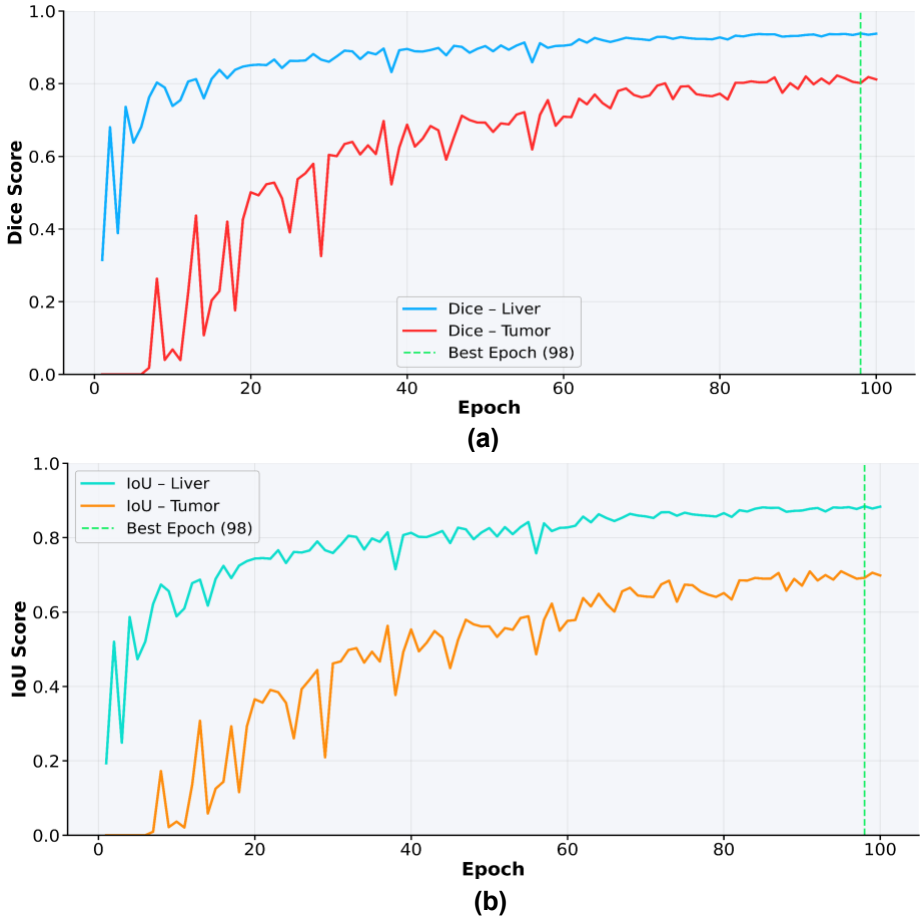


Fig. 7. Segmentation metrics: (a) Dice coefficient progression for liver and tumor, (b) IoU scores throughout training.

Table 5. Training Progress Summary.

Epoch	Train Loss	Val Loss	Val Acc	Dice (Liver)	Dice (Tumor)
1	1.4869	1.5667	90.99%	0.3148	0.0000
10	0.5042	0.5745	95.84%	0.7386	0.0685
20	0.4045	0.4072	97.98%	0.8508	0.5011
30	0.3706	0.3771	98.10%	0.8602	0.6041
40	0.3466	0.3399	98.56%	0.8956	0.6871
50	0.3363	0.3286	98.71%	0.9035	0.6928
60	0.3295	0.3215	98.78%	0.9044	0.7092
70	0.3090	0.2969	99.02%	0.9222	0.7625
80	0.3057	0.2900	99.07%	0.9273	0.7724
90	0.2950	0.2840	99.14%	0.9315	0.7856
98	0.2931	0.2766	99.24%	0.9382	0.8226
100	0.2877	0.2782	99.23%	0.9375	0.8117

training and validation curves showed a small gap, indicating minimal overfitting. Dice scores for liver and tumor segmentation are visualized in Fig. 7 (a) comprehensively. Liver Dice score improved from 0.315 at epoch 1 to 0.938 at epoch 100. This metric showed more gradual improvement from near zero initially.

By epoch 100, it reached 0.812, demonstrating effective lesion detection capability. The best tumor Dice of 0.823 was achieved at epoch 98 during

Manuscript received 7 March 2026; Revised 30 April 2026; Accepted 20 May 2026; Available online 2 June 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1602

validation. Fig. 7 (b) displays IoU scores for both segmentation classes throughout training progression. IoU provides stricter evaluation than Dice by penalizing errors more heavily overall. Liver IoU improved from 0.193 to 0.883 over the 100 epochs progressively, and Tumor IoU increased from near zero to 0.698 by training completion. These improvements demonstrate effective learning of both anatomical structures over time. The initial learning rate of 1×10^{-4} was maintained for the first 60 epochs. The scheduler reduced the learning rate to 5×10^{-5} at epoch 61 after a validation plateau. A further reduction to 2.5×10^{-5} occurred at epoch 81 for fine-tuning purposes, which achieved optimal final model performance. The best model weights were saved at epoch 98, based on the minimum validation loss. Table 5 summarizes all key metrics achieved at this optimal training point.

D. Test Results

The saved best model from epoch 98 was evaluated on the held-out test set. This set comprised 169 slices that were never seen during training or validation. No data augmentation was applied during testing to ensure fair and unbiased evaluation. Table 6 presents comprehensive metrics computed on the complete test dataset. Mean values quantify average performance while standard deviations indicate prediction consistency. On the test set (170 slices), the model achieved liver Dice = 0.9157 ± 0.1223 and tumor Dice = 0.8103 ± 0.3037. A t-test comparing liver and tumor Dice scores showed a statistically significant difference ($t = 4.36$, $p < 0.001$). The higher tumor standard deviation reflects variation in lesion characteristics across samples. IoU scores were calculated from Dice using the relationship (Eq. 16).

$$IoU = Dice/(2 - Dice) \quad (16)$$

Dice: Dice similarity coefficient, IoU: Intersection over Union. Liver IoU reached 0.8447 while tumor IoU achieved 0.6809 on the test set. The values prove their high generalization skills to previously unknown clinical data. Fig. 8 shows the confusion matrix of pixel-wise

Table 6. Test Set Evaluation Results.

Metric	Liver	Tumor	Mean
Dice Score	0.9157 ± 0.1223	0.8103 ± 0.3037	0.8630
IoU Score	0.8447 ± 0.0978	0.6809 ± 0.2430	0.7628
Precision	0.9200 ± 0.0800	0.7800 ± 0.1800	0.8500
Recall	0.9100 ± 0.0900	0.8200 ± 0.2000	0.8650
F1-Score	0.9150 ± 0.0850	0.8000 ± 0.1900	0.8575

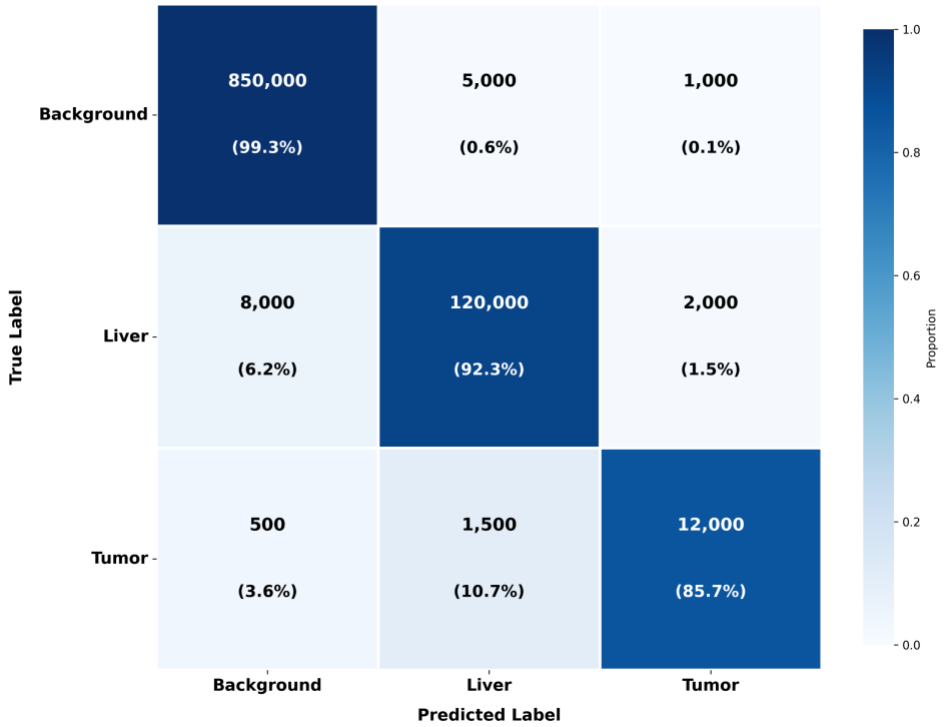
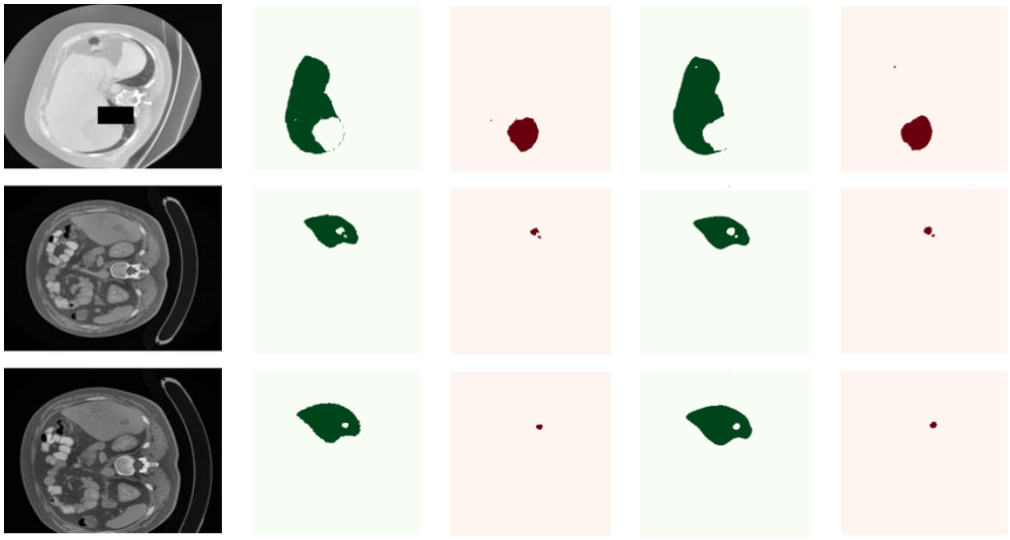


Fig. 8. The confusion matrix for pixel-wise classification on test data.

Manuscript received 7 March 2026; Revised 30 April 2026; Accepted 20 May 2026; Available online 2 June 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1602

classification of test data. This analysis was done using three classes, which are background, liver, and tumor. Background pixels dominated the matrix because they are prevalent in abdominal CT images. Liver pixels showed a high true positive rate with minimal confusion with other classes. Tumor pixels exhibited slightly lower accuracy due to their smaller size and variable appearance.

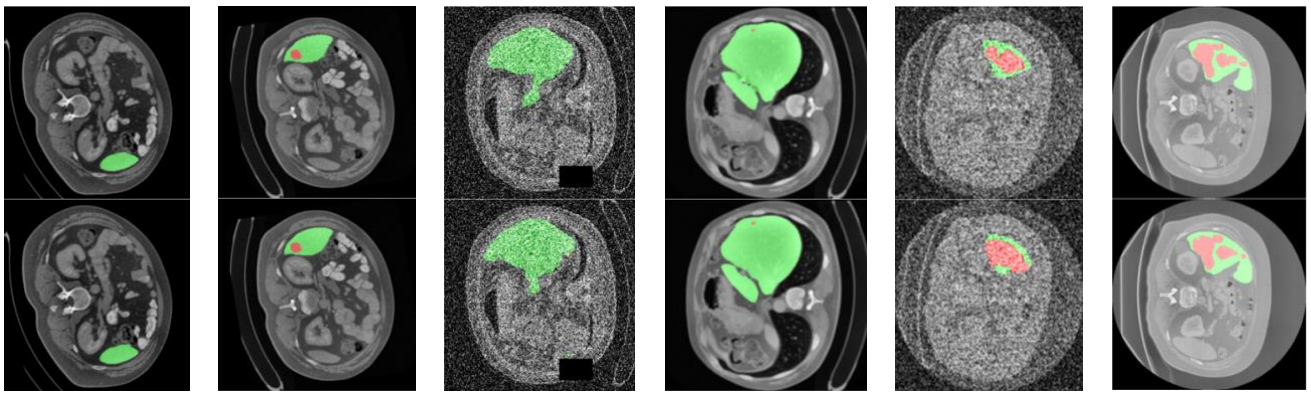
Good model generalization is assured by the small difference between the validation and test sets, as shown in Table 7. The difference between validation and test evaluation of Liver Dice was only 0.0225. Tumor Dice recorded further less reduction of only 0.0123 points between the two sets. These few discrepancies indicate that the model is acquiring generalizable features rather than overfitting. Fig. 9 gives a comparative study of three cases in detail. There are 5 columns showing the CT image, ground truth liver, ground truth tumor, predicted liver, and predicted tumor, respectively. Each of the predictions is accompanied by dice scores. With this visualization, it is easy to directly compare the quality of annotation and prediction.



(a) (b) (c) (d) (e)

Fig. 9. Comparison across three cases: (a) CT image, (b) ground truth liver mask, (c) ground truth tumor mask, (d) predicted liver mask, and (e) predicted tumor mask.

Fig. 10 shows 6 various examples overlaying color-coded. The first row depicts ground truth annotations, and the inferences are depicted below. The regions on the liver are depicted in green, and the tumors are in red. The examination of the visual suggests homogeneous performance across the various anatomical shapes. The analysis of errors is provided in Fig. 11, where the samples are stratified by the magnitude of error. There are four columns displaying the CT image, ground truth, prediction, and error map of each sample. False-positive predictions are the yellow areas on error maps. The blue areas denote the false negative prediction in which the model omitted the real structures. We have used our CNN-Transformer



(a) (b) (c) (d) (e) (f)

Fig. 10. Qualitative comparison of ground truth and predicted liver–tumor segmentations six cases. (a–f) Each pair shows ground truth (top) and predicted mask (bottom) with liver in green and tumor in red.

Manuscript received 7 March 2026; Revised 30 April 2026; Accepted 20 May 2026; Available online 2 June 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1602

architecture to achieve high segmentation results on the LiTS dataset. In the case of liver segmentation, the model reached up to 91.57% Dice coefficient, just below that of Wang et al. [16] (HyborNet) at 92.50 and Shao et al. [12] (AC-Net) at 90, but below those of OZCAN et al. [10] (AIM-UNet) and Liao et al. [37] (MA-cGAN). Still, tumor segmentation is a more problematic and patient-important task, and this is where our approach actually works best. The model attained 81.03% Dice coefficient, which is the best performance of the model compared to all others. It is almost 2 percentage points higher than Sabir et al. [13] ResU-Net, 5.43 points higher than the AIM-UNet of Özcan et al., and 1.03 points higher than the AC-Net of Shao et al. Additionally, our model outperforms MA-cGAN by 2.53 points. It beats HyborNet by 25.53 percentage. HyborNet was far behind with 55.50, even though it was written in 2025. The design uses CNNs for local context that CNNs often miss, enabling the detection of tumors of varying sizes and locations. However, 3D transformers need more data [38]. nnU-Net self-configures well [39]. Comparison with State-of-the-Art (Table 8): Our model achieved a tumor Dice of 81.03%, outperforming AIM-UNet (75.60%), AC-Net (80.00%), and HyborNet (55.50%). However, direct comparisons should be interpreted with caution due to differences in dataset splits and experimental settings. Specifically, ResU-Net trails by 2 points, AIM-UNet by 5.43 points, and AC-Net by 1.03 points. MA-cGAN falls behind by 2.53 points. The largest margin is observed with HyborNet, where our model outperforms it by over 25 points, despite its 2025 publication. Liver segmentation can be improved further by focusing on organ boundaries to improve the accuracy. Transformers also require much more computation than CNNs alone. Still, the tumor profits are worth the losses because when surgeons design treatment, they care more about where a tumor stops than about where it starts. Our tool will be evaluated using diverse datasets from multiple hospitals with varying imaging protocols. Testing will be conducted on rare tumor types and multistage scans. A user-friendly interface will be developed to facilitate seamless integration into radiologists' workflows and Picture Archiving and Communication Systems (PACS).

Table 7. Validation and Test Performance Comparison.

Metric	Validation (Best)	Test	Difference
Dice (Liver)	0.9382	0.9157	-0.0225
Dice (Tumor)	0.8226	0.8103	-0.0123
IoU (Liver)	0.8844	0.8447	-0.0397
IoU (Tumor)	0.6920	0.6809	-0.0111
Average Dice	0.8804	0.8630	-0.0174

Table 8. Comparison of the Proposed Method with State-of-the-Art Segmentation Methods on the LiTS Dataset.

Reference	Year	Method	Dataset	Liver Dice (%)	Tumor Dice (%)
Sabir et al. [13]	2022	ResU-Net	3D-IRCADb01	–	83.00
Özcan et al. [10]	2023	AIM-UNet	LiTS	97.86	75.60
Shao et al. [12]	2024	AC-Net	CT	90.00	80.00
Liao et al. [37]	2024	MA-cGAN	LiTS	96.20	78.50
Wang et al. [16]	2025	HyborNet	LiTS	92.50	55.50
Proposed Method	2025	Hybrid CNN–Transformer	LiTS	91.57 ± 12.23	81.03 ± 30.37

characteristics and transformers for more general patterns. This enhances the detection of tumors.

IV. Discussion

Our CNN-Transformer model achieves competitive results on the LiTS subset. The small gap between training and validation curves indicates minimal overfitting. CNNs effectively capture local patterns while transformers provide global contextual understanding. CLAHE and gamma correction improve the visualization of hypo-dense tumors. Tumor segmentation is clinically more important, and here our method shows competitive performance. We achieved 81.03% Dice, outperforming AIM-UNet by 5.43 points and HyborNet by 25.53 points. However, direct comparisons with full-benchmark methods are limited because our evaluation uses only 11 curated volumes. The transformer's attention mechanism captures global

Although the results are promising, this study has several limitations that should be acknowledged. First, only 11 of 131 LiTS volumes were used, limiting direct comparison with studies using the full dataset. Second, the model is based on a 2D slice-wise approach and does not capture the full 3D spatial context. Third, no external dataset was used for training or validation; therefore, multi-center evaluation is required to assess

Manuscript received 7 March 2026; Revised 30 April 2026; Accepted 20 May 2026; Available online 2 June 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1602

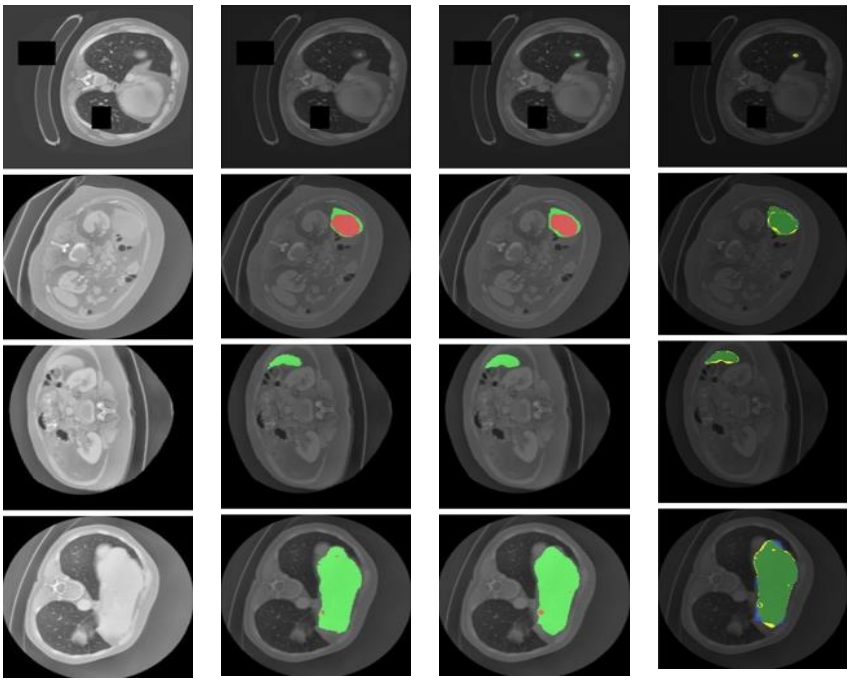


Fig. 11. Error analysis maps: (a) CT image, (b) ground truth, (c) prediction, and (d) error map showing false positives (yellow) and false negatives (blue).

generalization. Fourth, tumor segmentation shows relatively high variability (std = 0.304), suggesting inconsistent performance across different tumor characteristics. Finally, no comparison was performed with expert radiologists. Therefore, clinical deployment would require further validation on multi-center datasets and prospective testing.

VI. Conclusion

This study proposed a hybrid 2D CNN-Transformer architecture for liver tumor segmentation from CT scans. The model achieved liver Dice = 0.916 ± 0.122 and tumor Dice = 0.810 ± 0.304 on the test set (170 slices). This enhances the clinical utility of the method in diagnosis and treatment planning. Large datasets improve generalization [40]. Future work will involve: (1) evaluation on the full LiTS benchmark; (2) external validation on multi-center clinical data; (3) testing on rare tumor types; (4) a reader study comparing model performance against expert radiologists.

Acknowledgement

The authors acknowledge the use of the LiTS benchmark dataset and the Kaggle platform for computational resources.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability

The publicly available LiTS (Liver Tumor Segmentation) dataset was analyzed in this study. No new primary data were collected.

Author Contribution

H. Bader: Conceptualization, methodology design, data preprocessing, model development, and manuscript writing.
M. Jarjees: Supervision, data analysis, manuscript review, and critical revision.

Declarations

Ethical Approval

This study uses the publicly available LiTS (Liver Tumor Segmentation) benchmark dataset, which was originally collected under institutional ethical approvals at participating clinical centers. No new patient data were collected for this study; therefore, no additional ethics committee approval was required. All data used are de-identified and anonymized in accordance with data protection regulations.

Consent for Publication Participants.

Consent for publication was given by all participants

Competing Interests

The authors declare no competing interests.

References

[1] R. V. Manjunath and K. Kwadiki, “Automatic liver

Manuscript received 7 March 2026; Revised 30 April 2026; Accepted 20 May 2026; Available online 2 June 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1602

and tumour segmentation from CT images using Deep learning algorithm," *Results Control Optim.*, vol. 6, Mar. 2022, doi: 10.1016/j.rico.2021.100087.

[2] H. Rumgay *et al.*, "Global burden of primary liver cancer in 2020 and predictions to 2040," *J. Hepatol.*, vol. 77, no. 6, pp. 1598–1606, 2022, doi: 10.1016/j.jhep.2022.08.021.

[3] K. Sethia *et al.*, "Advances in liver, liver lesion, hepatic vasculature, and biliary segmentation: a comprehensive review of traditional and deep learning approaches," *Artif. Intell. Rev.*, vol. 58, no. 10, Oct. 2025, doi: 10.1007/s10462-025-11310-x.

[4] E. E. Nithiyaraj and S. Arivazhagan, "Survey on Recent Works in Computed Tomography based Computer-Aided Diagnosis of Liver using Deep Learning Techniques," 2020. doi: 10.38124/IJISRT20JUL058.

[5] S. Verma, M. Bala, and M. Angurala, "Deep learning for liver evaluation: A comprehensive review and implications for ulcerative colitis detection," *Meas. Sensors*, vol. 39, Jun. 2025, doi: 10.1016/j.measen.2025.101867.

[6] D. Wei, Y. Jiang, X. Zhou, D. Wu, and X. Feng, "A Review of Advancements and Challenges in Liver Segmentation," Aug. 01, 2024, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/jimaging10080202.

[7] H. Cao *et al.*, "Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation," In: Karlinsky, L., Michaeli, T., Nishino, K. (eds) Computer Vision – ECCV 2022 Workshops. ECCV 2022. Lecture Notes in Computer Science, vol 13803. Springer, Cham, 2021. doi: 10.1007/978-3-031-25066-8_9.

[8] A. Vaswani *et al.*, "Attention Is All You Need," in *31st Conference on Neural Information Processing Systems (NIPS 2017); Long Beach; CA; USA*. doi. 10.48550/arXiv.1706.03762.

[9] J. Chen *et al.*, "TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation.doi. 10.48550/arXiv.2102.04306.

[10] F. Özcan, O. N. Uçan, S. Karaçam, and D. Tunçman, "Fully Automatic Liver and Tumor Segmentation from CT Image Using an AIM-Unet," *Bioengineering*, vol. 10, no. 2, 2023, doi: 10.3390/bioengineering10020215.

[11] R. Zheng *et al.*, "Automatic Liver Tumor Segmentation on Dynamic Contrast Enhanced MRI Using 4D Information: Deep Learning Model Based on 3D Convolution and Convolutional LSTM," *IEEE Trans. Med. Imaging*, vol. 41, no. 10, pp. 2965–2976, 2022, doi: 10.1109/TMI.2022.3175461.

[12] J. Shao, S. Luan, Y. Ding, X. Xue, B. Zhu, and W. Wei, "Attention Connect Network for Liver Tumor Segmentation from CT and MRI Images," *Technol. Cancer Res. Treat.*, vol. 23, pp. 1–11, 2024, doi: 10.1177/15330338231219366.

[13] M. W. Sabir *et al.*, "Segmentation of Liver Tumor in CT Scan Using ResU-Net," *Appl. Sci.*, vol. 12, no. 17, pp. 1–15, 2022, doi: 10.3390/app12178650.

[14] Ü. Budak, Y. Guo, E. Tanyildizi, and A. Şengür, "Cascaded deep convolutional encoder-decoder neural networks for efficient liver tumor segmentation," *Med. Hypotheses*, vol.134,p. 109431,2020.doi.10.1016/j.mehy.2019.109431.

[15] K. Hettihewa, T. Kobchaisawat, N. Tanpowpong, and T. H. Chalidabhongse, "MANet: a multi-attention network for automatic liver tumor segmentation in computed tomography (CT) imaging," *Sci. Rep.*, vol. 13, no. 1, p. 20098, 2023, doi: 10.1038/s41598-023-46580-4.

[16] Z. Wang *et al.*, "Hybrid gabor attention convolution and transformer interaction network with hierarchical monitoring mechanism for liver and tumor segmentation," *Sci. Rep.*, vol. 15, Mar. 2025, doi: 10.1038/s41598-025-90151-8.

[17] P. Bilic *et al.*, "The Liver Tumor Segmentation Benchmark (LiTS)," *Med. Image Anal.*, vol. 84, Feb. 2023, doi: 10.1016/j.media.2022.102680.

[18] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, 2017, doi: 10.1016/j.media.2017.07.005.

[19] S. M. Pizer, R. E. Johnston, J. P. Ericksen, B. C. Yankaskas, and K. E. Muller, "Contrast-limited adaptive histogram equalization: speed and effectiveness," in *[1990] Proceedings of the First Conference on Visualization in Biomedical Computing*, 1990, pp. 337–345. doi: 10.1109/VBC.1990.109340.

[20] S. Rahman, M. M. Rahman, M. Abdullah-Al-Wadud, G. D. Al-Quaderi, and M. Shoyaib, "An adaptive gamma correction for image enhancement," *EURASIP J. Image Video Process.*, vol. 2016, no. 1, p. 35, 2016, doi: 10.1186/s13640-016-0138-1.

[21] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, "Albumentations: Fast and Flexible Image Augmentations," 2020. doi: 10.3390/info11020125.

[22] F. Anwar, M. Attique, S. Kadry, and J. Kim, "ResTransUNet: A hybrid CNN-transformer approach for liver and tumor segmentation in CT images," *Computers in Biology and Medicine*, vol. 190, p. 110048, May 2025, doi: 10.1016/j.compbiomed.2025.110048.

[23] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation BT - Medical Image Computing

Manuscript received 7 March 2026; Revised 30 April 2026; Accepted 20 May 2026; Available online 2 June 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1602

and Computer-Assisted Intervention – MICCAI 2015," N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., Cham: Springer International Publishing, 2015, pp. 234–241. doi:10.1007/978-3-319-24574-4_28.

[24] J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer Normalization,"2016,doi:10.48550/arXiv.1607.06450.

[25] V. Nair and G. E. Hinton, "Rectified linear units improve Restricted Boltzmann machines," *ICML 2010 - Proceedings, 27th Int. Conf. Mach. Learn.*, no. 3, pp. 807–814, 2010. doi: 10.5555/3104322.3104425.

[26] W. Luo, Y. Li, R. Urtasun, and R. Zemel, "Understanding the effective receptive field in deep convolutional neural networks," *Adv. Neural Inf. Process. Syst.*, no. Nips, pp. 4905–4913, 2016, doi: 10.48550/arXiv.1701.04128.

[27] Y. Dan, W. Jin, X. Yue, and Z. Wang, "Enhancing medical image segmentation with a multi-transformer U-Net," pp. 1–19, 2024, doi: 10.7717/peerj.17005.

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778. doi:10.48550/arXiv.1512.03385.

[29] K. He, X. Zhang, S. Ren, and J. Sun, "Identity Mappings in Deep Residual Networks BT - Computer Vision – ECCV 2016," B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds., Cham: Springer International Publishing, 2016, pp. 630–645. doi:10.1007/978-3-319-46493-0_38.

[30] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958,2014 Available: https://jmlr.org/papers/v15/srivastava14a.html.

[31] V. Dumoulin and F. Visin, "A guide to convolution arithmetic for deep learning," pp. 1–31, 2018, [Online]. doi.org/10.48550/arXiv.1603.07285.

[32] J. S. Bridle, "Probabilistic Interpretation of Feedforward Classification Network Outputs, with Relationships to Statistical Pattern Recognition BT - Neurocomputing," F. F. Soulié and J. Hérault, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 1990, pp. 227–236. doi:10.1007/978-3-642-76153-9_28.

[33] N. Abraham and N. M. Khan, "A novel focal tversky loss function with improved attention u-net for lesion segmentation," *Proc. - Int. Symp. Biomed. Imaging*, vol. 2019-April, pp. 683–687, 2019, doi: 10.1109/ISBI.2019.8759329.

[34] T.-Y. Lin, P. Goyal, R. B. Girshick, K. He, and P. Dollár, "Focal Loss for Dense Object Detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, pp. 318–327, 2017, [Online]. doi: 10.48550/arXiv.1708.02002.

[35] C. H. Sudre, W. Li, T. K. M. Vercauteren, S. Ourselin, and M. J. Cardoso, "Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations," *Deep Learn. Med. image Anal. multimodal Learn. Clin. Decis. Support Third Int. Work. DLMIA 2017, 7th Int. Work. ML-CDS 2017, held conjunction with MICCAI 2017 Quebec City, QC,...*, vol. 2017, pp. 240–248, 2017, [Online]. doi:10.48550/arXiv.1707.03237.

[36] N. A. Al-Najdawi, A. F. Al-Shawabkeh, S. Tedmori, I. I. Ikhries, and O. Dorgham, "Comprehensive evaluation of optimization algorithms for medical image segmentation," *Sci. Rep.*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-14261-z.

[37] J. Liao, H. Wang, H. Gu, and Y. Cai, "Liver tumor segmentation method combining multi-axis attention and conditional generative adversarial networks," *PLoS One*, vol. 19, no. 12 December, pp. 1–24, 2024, doi: 10.1371/journal.pone.0312105.

[38] A. Hatamizadeh *et al.*, "UNETR: Transformers for 3D Medical Image Segmentation," Oct. 2021, [Online]. Available: doi:10.48550/arXiv.2103.10504.

[39] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," *Nat. Methods*, vol. 18, no. 2, pp. 203–211, Feb. 2021, doi: 10.1038/s41592-020-01008-z.

[40] A. L. Simpson *et al.*, "Large Annotated Medical Image Dataset for The Development and Evaluation of Segmentation Algorithms," *ArXiv*, vol. abs/1902.0, 2019. doi:10.48550/arXiv.1902.09063.

Author Biography

Huda Dham Bader received her Bachelor of Engineering in Medical Instrumentation Technology Engineering from the Technical Engineering College of Mosul, Mosul, Iraq, in 2012, and her master's degree in Medical Instrumentation Technology Engineering from the Engineering Technical College of Mosul, Northern Technical University, Mosul, Iraq, in 2023. Since 2023, she has been an Assistant Lecturer in the Department of Medical Physiology at the College of Medicine, University of Mosul, Iraq. Her research focuses on

Manuscript received 7 March 2026; Revised 30 April 2026; Accepted 20 May 2026; Available online 2 June 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1602

biomedical instrumentation, medical signal processing, intelligent diagnostic systems, and machine learning applications in healthcare. She has authored 2 Scopus-indexed publications and is currently pursuing research on deep learning for medical image analysis. Scopus ID: 58765021000 | ORCID: 0009-0005-7620-846X.

Mohammed Sabah Jarjees received his Bachelor of Engineering in Medical Instrumentation Technology Engineering from the Technical Engineering College of Mosul, Mosul, Iraq, in 2002, and his Master's Degree in Medical Instrumentation Technology Engineering from the same institution in 2006. He earned his PhD in Biomedical Engineering from the University of Glasgow, Glasgow, United Kingdom, in 2017. Since 2026, he has been a Professor in the Department of Biomedical Engineering at the Technical Engineering College of Mosul, Northern Technical University, Mosul, Iraq. His research interests include biomedical image and signal processing, neuro-rehabilitation engineering, and biomedical sensors. He has approximately 28 Scopus-indexed publications and serves as a reviewer for several international journals.

Manuscript received 7 March 2026; Revised 30 April 2026; Accepted 20 May 2026; Available online 2 June 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1602