IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 15, No. 6, December 2026, pp. 5512~5521
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i6.pp5512-5521

A two-stage hybrid model for liver tumor detection and segmentation using vision transformer

Ananda R. Kumar Mukkala[1], G. Nivedita[2], K. V. Balaramakrishna[3], Sujan Babu Vadde[4], Nandkishor Narkhede[5], Suneel Kumar Asileti[6]

[1]Department of Computer Science and Engineering, Sreyas Institute of Engineering and Technology, Hyderabad, India
[2]Department of Computer Science and Engineering, Geethanjali College of Engineering and Technology, Hyderabad, India
[3]Department of Electronics and Communication Engineering, Aditya University, Surampalem, India
[4]Department of Computer Science and Engineering, Koneru Lakshmaiah Educational Foundation Deemed to be University, Vijayawada, India
[5]Department of Electronics and Computer Science, Shah and Anchor Kutchhi Engineering College, Mumbai, India
[6]Department of Electronics and Communication Engineering, Usha Rama College of Engineering and Technology, Telaprolu, India

Article Info

Article history:

Received Sep 19, 2025
Revised Aug 9, 2026
Accepted Sep 19, 2026

Keywords:

CT scans
Liver cancers
Swin U-shaped network with transformers
TCGA-LIHC
Vision transformer

ABSTRACT

Liver cancer is one of the fastest-growing and most life-threatening malignancies, making early and precise detection essential for improving patient outcomes. This work proposes a two-phase hybrid deep learning framework that integrates swin U-shaped network with transformers (UNETR) for segmentation and a vision transformer (ViT) for classification of liver tumors from histopathological images. The system begins with a preprocessing stage that standardizes image resolution and enhances contrast to ensure consistent input quality. Local intensity pattern transformation (LIPT) and statistical descriptors are then used to extract spatial and textural features. Swin UNETR performs hierarchical self-attention–based segmentation to localize tumor regions, after which the ViT model classifies the segmented areas by capturing both global and local dependencies. Performance evaluation using dice similarity coefficient (DSC), Jaccard coefficient (JC), accuracy, precision, recall, specificity, intersection over union (IoU), Hausdorff distance (HD), and volume overlap error (VOE) demonstrates the proposed model's superiority, achieving a DSC of 95.47%, JC of 0.93, recall of 98.12%, accuracy of 99.76%, and specificity of 99.89%.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Ananda R. Kumar Mukkala
Department of Computer Science and Engineering, Sreyas Institute of Engineering and Technology
Hyderabad, India
Email: anandranjit@gmail.com

1. INTRODUCTION

Liver is considered to be the largest solid organ present within the body of humans that fulfills the role of important metabolic and digestive functions. Liver cancer is said to be the third most common cause of deaths due to cancers throughout the world, highlighting the importance of this disease. Moreover, liver is also considered to be one of the common sites where the metastasis occurs from colon cancer, rectal cancer, pancreatic cancer, breast cancer, and lung cancer. Proper segmentation of the liver becomes very important for the development of treatment plans and surgeries. Modern technology in medical imaging such as computed tomography (CT) scans and magnetic resonance imaging (MRI) has helped healthcare professionals in segmenting the liver effectively [1].

The failure to diagnose liver tumor at its early stage results in eventual patient death thus making liver cancer the deadliest form of cancer that can lead to death. The tumor travels through the liver area while

aging toward a larger configuration resulting in reduced intestinal dimensions that generate a major medical issue. The detection and diagnostic and treatment of liver cancer proves difficult early on thus improving patient survival outcomes. The accuracy of diagnosis evaluation together with successful treatment depends on correct liver tumor segmentation using imaging data [2]. The exact demarcation of liver tumors holds crucial importance for medical practitioners since it enables them to assess tumor features along with shape characteristics and therapy outcomes. Early diagnosis and accurate liver tumor segmentation serve essential roles in treatment planning and patient survival enhancement. The correct identification of liver tumors depends on medical imaging techniques such as CT scans and ultrasound and MRI because these methods detect presence and growth levels of tumors with high diagnostic confidence [3]. Semi-automatic classification is commonly assigned to these tools because their processing needs humans to step in both for complex parameter adjustments and noise management.

Machine learning methods achieve excellent segmentation results. The use of manually created features presents a major problem because it straightly impacts these methods' execution precision. The application of deep learning methods has resulted in exceptional progress through the use of convolutional neural networks (CNNs). This automated feature engineering capability of deep learning methods has enabled numerous successes across target detection and image segmentation and classification operations [4]. Vision transformers (ViT) have accomplished the highest level of performance across different computer vision and medical image analysis applications in recent times. The architectural design of ViTs enables them to maintain uniform representation learning capabilities which excel at recognizing distant dependencies across sequences. The developers of swin transformers created this model to overcome ViT-based models' weaknesses through their hierarchical design which combines patch merging layers with relative position embedding. Multiple studies have investigated the application of ViTs as stand-alone components integrated into CNN-based U-shaped architectures for medical image analysis. U-shaped network with transformers (UNETR) proved the effectiveness of combining ViT-based encoder architecture with CNN-based decoder when processing 3D tokens. UNETR faces limitations from its fixed visual token dimensions because it needs extra convolutional layers to adapt representations to specific resolutions which adds both computational complexity and increases parameter counts [5].

The proposed approach brings multiple essential enhancements to liver tumor applications through a combination of swin UNETR and ViT which boosts both accuracy levels and operational stability. The approach improves characteristic identification through the implementation of statistical analysis and local intensity pattern transformation (LIPT) for effective detection of spatial and textural and intensity-based regional variations in liver tumors. Swin UNETR segmentation allows the model to detect tumors with precise boundaries because it maintains hierarchical spatial relationships throughout the process. The transformer-based architecture of swin UNETR improves segmentation accuracy in challenging cases since it surpasses conventional CNN-based models that have difficulties with complex tumor structures. The tumor classification process employs ViT which uses its self-attention mechanism to improve diagnosis results by understanding relationships between global contexts thus outperforming traditional feature-based classifiers.

The paper follows this organization: section 2 reviews previous research work. Section 3 describes the proposed architecture. The experimental section along with result analysis and discussions appears in section 4. The paper concludes with section 5 which presents closing remarks and future research guidelines.

2. LITERATURE SURVEY

Dharaneswar and Kumar [6] produced a deep learning network-based framework named Sholingapuram and Santosh that detects and heals liver cancer during its early stages. The system gathers images for segmenting and classifying liver tumors as its initial task. The segmentation procedure occurs through Trans-UNet. Segmented image outputs from the process are sent to the liver tumor classification system. The proposed ViT-based residual attention network (RAN) serves to perform the classification [6]. Shao *et al.* [7] developed the attention connect network (AC-Net) which serves as an automated solution for liver tumor segmentation. This approach uses the U-shaped network structure but integrates two vital attention elements which substitute traditional skip-connections to merge spatial characteristics. The system consists of an axial attention module and a ViT module. A loss function that suits this method specifically has been implemented. This approach by first training AC-Net with LiTS2017 data and then performing fine-tuning with CT and MRI data [7]. The ResTransUNet framework by Ou *et al.* [8] combines transformer remote interaction capabilities with original U-Net architecture performance for end-to-end automatic liver segmentation. The main contribution of this work presents a new network design that brings together UNET and transformer structures [8].

Cheng *et al.* [9] developed a new edge-guided segmentation network which they named EG-UNETR. This model consists of three components: TransBridge module (TB) and the feature semantic enhancement module (FSE) together with an edge-guided block (EGB). The TB utilizes multi-head

position-aware attention (MHPA) together with fusion attention (FA) for multi-scale cross-level contextual feature integration that eliminates irrelevant information. Additional processing power is needed to handle high-resolution encoding features so the designers created a lightweight version of TransBridgemodule (LTB) to handle these tasks efficiently. The FSE uses parallel dilated convolutions along with hybrid attention mechanisms to obtain complete semantic information. The non-local block (NLB) aligns and refines these features through its operation. The EGB contains edge-guide attention (EGA) to refine tumor edges leading to precise segmentation outcomes. The performance of EG-UNETR is evaluated through multiple experiments on LiTS2017 and 3D-IRCADb datasets which are described [9]. Phan *et al.* [10] used machine learning algorithms on actual patient data to both detect and categorize liver lesions while aiming for higher screening and diagnosis efficiency. The research contains neural network model implementation using dense convolutional network (DenseNet-121) together with visual geometry group (VGG-19) and ViT. Evaluated three types of liver lesions through analysis of 2008 CT liver images which were separated into arterial, delay, plain, and venous phase groups.

The proposed method applies a two-stage deep learning framework with swin UNETR for segmentation and ViT for classification to detect liver tumors and segment pathology images. Standardization of input images starts with preprocessing steps that include resizing together with contrast enhancement techniques. The subsequent step includes LIPT and statistical feature extraction methods that detect tumor patterns in texture and position. Swin UNETR executes detailed tumor boundary detection through hierarchical self-attention combined with a patch-based processing system.

3. PROPOSED METHOD

A new two-step hybrid deep learning system has been developed which unites swin UNETR with ViT to execute accurate medical image analysis during the detection and segmentation of liver tumors from pathology images as shown in Figure 1. A complete preprocessing pipeline begins by standardizing dimensions and enhancing contrast of input histopathological images directly from the database. The extracted features result from using statistical techniques and LIPT to detect spatial and textural patterns in liver tumor structures. The deep learning model incorporates swin UNETR to perform hierarchical representation learning of extracted features for segmenting tumor regions through patch-based processing with hierarchical self-attention mechanisms. The testing phase starts by processing input images through preprocessing steps followed by extracting features to maintain consistency with training phase procedures. ViT receives extracted features to perform the classification of tumor regions where it determines if tissue segments are benign or malignant. The trained ViT model advances tumor classification through self-attention mechanisms which extract local and global dependencies from tumor microenvironment. Final prediction evaluations consist of precision accuracy metrics together with recall sensitivity measures. The main innovation of the proposed method combines swin UNETR's feature learning hierarchy with ViT's global contextual model to enhance tumor detection precision while improving stability. Textual pattern recognition of medical images enables this diagnostic tool to help clinical staff with quick diagnosis of liver cancer patients and therapeutic planning. The research will accumulate medical image databases and create quick computational frameworks for real-time medical applications through multi-spectral imaging.

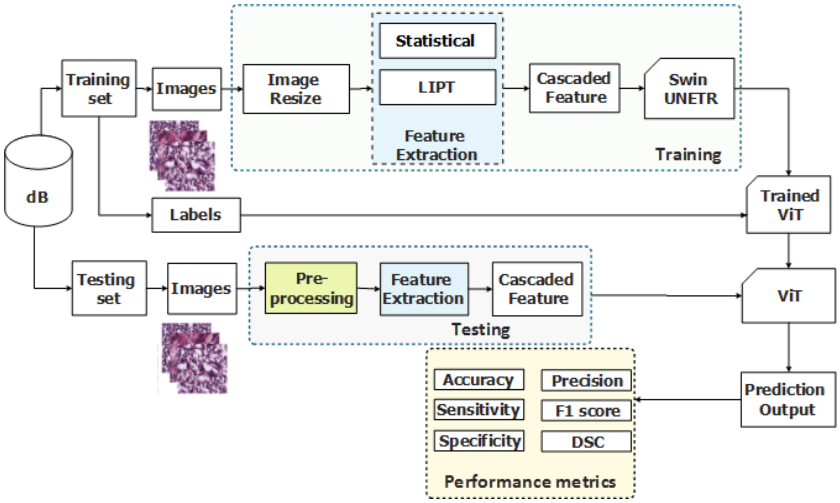


Figure 1. The block diagram of the proposed method

3.1. Image resizing

A picture becomes altered in its pixel details when someone resizes it. Picture resizing offers different functionality than image resampling since it enables users to adjust the image size without pixel number modification by varying pixel spacing [11], [12]. The image received a 128 × 128 pixels reduction as part of this work.

3.2. Statistical feature extraction

Statistical features are extracted to characterize image texture and intensity variations. The following features are considered: energy, entropy, inertia, correlation, inverse difference moment, and difference entropy. Their mathematical formulations are presented in (1) to (6), respectively [13], [14].

i) Energy: it is controlled by the summation of squared parts [13], [14].

$$EG = \sum_i^{N_G} \sum_j^{N_G} \{p(i,j)\}^2 \quad (1)$$

ii) Entropy

$$EN = -\sum_i^{N_G} \sum_j^{N_G} p(i,j) \log(p(i,j)) \quad (2)$$

iii) Inertia

$$IN = -\sum_i^{N_G} \sum_j^{N_G} (i-j)^2 \, p(i,j) \quad (3)$$

iv) Correlation: it measures how the pixel is identified with one another in the entire picture.

$$CO = \frac{\sum_i^{N_G} \sum_j^{N_G} (i-\mu_x)(j-\mu_y) \; p(i,j)}{\sigma_x \sigma_y} \quad (4)$$

v) Inverse difference moment

$$idm = \sum_i^{N_G} \sum_j^{N_G} \frac{1}{1+(i-j)^2} \; p(i,j) \quad (5)$$

vi) Difference entropy

$$DE = -\sum_{K=0}^{N_G-1} P_{x-y}(k) log_2 \, P_{x-y}(k) \quad (6)$$

3.3. Local intensity-pattern tensor

The local intensity pattern tensor is used to represent local intensity variations and capture textural information in the image [15]–[17]. The mathematical formulations used to construct the tensor are presented in (7) to (14).

$$I_p = \{\mathrm{I(p+d)} \mid \mathrm{d} \in \mathrm{D}_k\} \quad (7)$$

$$H_p(b) = \sum_{i=1}^{N} \delta\left(\mathrm{floor}\left(\frac{\mathrm{I(p+}\, d_i)-\mathrm{I_{min}}}{\Delta}\right) = \mathrm{b}\right) \quad (8)$$

$$T_p = [\mathrm{T_{b1,b2,b3}}] \quad (9)$$

$$\mathrm{T_{b1,b2,b3}} = \sum_{(d_i,d_j)\epsilon\, \mathrm{D}_k} \delta\left(\mathrm{floor}\left(\frac{\mathrm{I(p+}\, d_i)-\mathrm{I_{min}}}{\Delta}\right) = and \;\; \left(\mathrm{floor}\left(\frac{\mathrm{I(p+}\, d_i)-\mathrm{I_{min}}}{\Delta}\right) = \mathrm{b_2}\right)\right) \quad (10)$$

$$\mathrm{Tp} \approx \mathrm{U1} \cdot \mathrm{S} \cdot \mathrm{U2} \cdot \mathrm{U3} \quad (11)$$

$$F_p = vec(\mathrm{U1 \text{ or } U2 \text{ or } U3 \text{ or } S}) \quad (12)$$

$$F_{Region} = \frac{1}{|R|} \sum_{\boldsymbol{p}\in \boldsymbol{R}} F_p \quad (13)$$

$$F_{Region,norm} = \frac{F_{Region}-\mu}{\sigma} \quad (14)$$

3.4. Swin-U-shaped network with transformers

The swin UNETR model uses patch embedding and transformer-based attention mechanisms to extract hierarchical features from input images. The mathematical formulations describing patch embedding and the attention mechanism are presented in (15) to (21). The swin UNETR is a deep learning model for medical image segmentation, using swin transformer for the encoder part and U-Net-like architecture for the decoder. For an assumed input CT scan image, X of dimension H × W × C, the swin UNETR divides it into patches of dimension P × P. Each patch is then mapped to an embedding space.

$$z_0 = [X_p^1 E;\ X_p^2 E; \ldots . X_p^N E + E_{POS} \tag{15}$$

Each patch representation is fed into the swin transformer encoder for hierarchical feature extraction. For a given patch embedding Z, queries, keys, and values are computed as:

$$Q = ZW_Q, K = ZW_K, V = ZW_v \tag{16}$$

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{17}$$

$$\{Z_1, Z_2, Z_3, Z_4\} \tag{18}$$

$$X_{\mathbf{skip}} = \text{Concat}(Z_i, X_{\text{up}}) \tag{19}$$

$$X_{\text{final}} = \sigma(W_{conv}, * X_{\mathbf{skip}} + b_{conv}) \tag{20}$$

$$Y = \sigma(W_{\text{out}}, * X_{\text{final}} + b_{\text{out}}) \tag{21}$$

Figure 2 shows the swin UNETR architecture. Table 1 shows the parameter specification. The proposed swin UNETR-based segmentation model operates with selected hyperparameters that maximize its performance output.

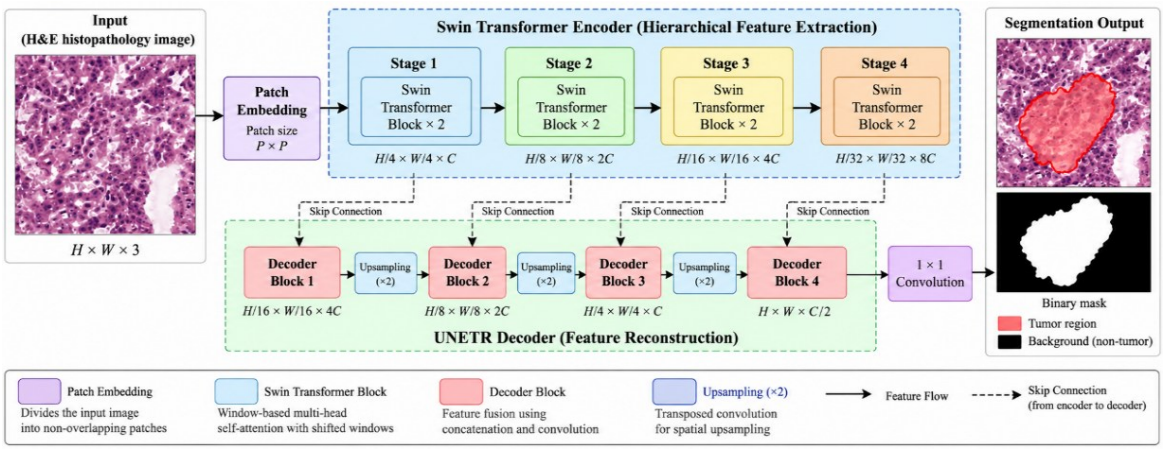


Figure 2. Schematic architecture of the swin UNETR model for liver tumor segmentation

Table 1. Parameter specification of swin UNETR

Parameter	Value
Batch size	8
Normalization	Layer normalization
Patch size	4 × 4 × 4
Loss function	Dice loss
Optimizer	AdamW
Learning rate scheduler	Cosine annealing
Learning rate	1e-4

3.5. The vision transformer

ViT modifies the architecture of the transformer network by treating the image patches as tokens, as defined in (21) [18]–[20].

$$z_0 = [x_1, x_2, \ldots x_N]W_e + b_e \quad (21)$$

These patches are flattened and embedded. All the patches that have been embedded are treated as tokens and are fed to the layers of transformer network. ViT model employs multi-headed self-attention mechanism and feed-forward networks, as mentioned above. Parameters used in the ViT model are listed in Table 2.

Table 2. Parameter specification of ViT

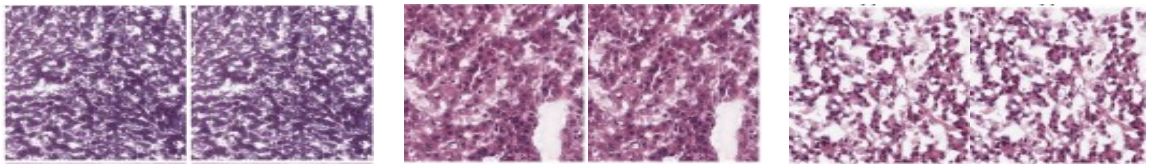
Parameter	Value
Batch size	16
Number of attention heads	12
Number of layers	12
Learning rate scheduler	Cosine annealing
Patch size	16 × 16
Learning rate	3e - 4
Dropout rate	0.1
Loss function	Cross-entropy loss

4. RESULTS AND DISCUSSION

This section present the tests of the suggested method of categorizing breast cancer images, which is done in the programming language Python using Google Colab. Google Colab provides us with computing facilities for free, including an Intel Xeon CPU with 2 cores and 13 GB of RAM memory. For difficult computations such as deep learning, Google Colab also provides us with an NVIDIA Tesla K80 GPU with 12 GB of VRAM.

4.1. Dataset

The cancer genome atlas-liver hepatocellular carcinoma (TCGA-LIHC) is one of the largest publicly available data sets for investigating the issue of liver cancer, in particular, the most common type of liver cancer is hepatocellular carcinoma (HCC). The developed hybrid model based on deep learning techniques was trained on publicly available TCGA-LIHC database with 491 whole slide images (WSI) [21], [22]. Three classes of liver HCC tumors were established on 491 slides and these include solid tissue normal as class 0, primary tumor as class 1 and recurrent tumor as class 2. There were 89 slides of solid tissue normal along with 399 slides of primary tumors and 3 slides of recurrent tumors. In deep learning applications such as classification it becomes difficult to input full resolution WSI with large dimensions into CNN since it will greatly increase the computation effort and the number of parameters. Reduction in size of histopathology images using downsampling or reshaping techniques will result in huge loss of information from the image. Every WSI image undergoes a tiling process which generates 1,024 × 1,024 patches of images as per procedure in [23]. We conducted the patch extraction process while keeping sufficient tissue visibility within each extracted patch. A total of 813, 893, and 680 valid type 0, type 1, and type 2 patches thus resulted from this procedure and served as the foundation for creating the hybrid deep learning model. Figure 3 shows the sample images of the three types in the dataset: type 0 as in Figure 3(a), type 1 as in Figure 3(b), and type 2 as in Figure 3(c).



(a) (b) (c)

Figure 3. Sample images from the dataset: (a) type 0, (b) type 1, and (c) type 2

4.2. Performance metrics

Performance metrics for segmentation evaluate how well a model segments an image or data into distinct parts or regions [24], [25]. The segmentation performance metrics used to evaluate the proposed

model are defined mathematically in (22) to (30). These metrics assess segmentation accuracy, overlap, and boundary similarity.

$$DSC = \frac{2TP}{FP + 2TP + FN} \tag{22}$$

$$IoU = \frac{|A \cap B|}{|A \cup B|} \tag{23}$$

$$HD(A,B) = max(\sup_{a \in A} inf_{b \in B}\, d(a,b), \sup_{b \in B} inf_{b \in A}\, d(a,b)) \tag{24}$$

$$VOE = 1 - \frac{|A \cap B|}{|A \cup B|} \tag{25}$$

$$MSD(A,B) = \frac{1}{|A|} \sum_{a \in A} \min_{b \in B} d(a,b) \tag{26}$$

$$Sensitivity = \frac{TP}{TP+FN} \tag{27}$$

$$Precision = \frac{TP}{TP+EN} \tag{28}$$

$$Specificity = \frac{TN}{TN+FP} \tag{29}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{30}$$

Table 3 compares UNET, UNETR, and swin UNETR, demonstrating the superior segmentation performance of swin UNETR. Its intersection over union (IoU) reaches 0.96, compared with 0.64 for UNET, indicating better overlap between predicted and ground-truth regions. The Hausdorff distance (HD) also decreases from 22 mm to 9 mm, indicating improved boundary accuracy. Table 4 evaluates training testing ratios of 40/60, 50/50, 60/40, and 70/30. Performance improves as the training proportion increases. Table 5 compares the proposed method with existing approaches, further highlighting its effectiveness in liver tumor segmentation based on the reported evaluation metrics.

Table 3. Comparative performance of the proposed method based on UNET

Performance	UNET	UNETR	Swin UNETR
IoU	0.64	0.85	0.96
HD	22 mm	15 mm	9 mm
Volume overlap error (VOE)	30%	25%	15%
Mean surface distance (MSD)	3.5 mm	4.0 mm	2.0 mm
Dice similarity coefficient (DSC)	83.63	0.90	0..95
Jaccard coefficient (JC)	0.87	0.82	0.93

Table 4. Comparative performance of the proposed method based on training and testing ratio

Performance	40/60	50/50	60/40	70/30
IoU	0.51	0.64	0.80	0.96
HD	22 mm	18 mm	13 mm	9 mm
VOE	25%	23%	20%	15%
MSD	3.3 mm	2.8 mm	2.6 mm	2.0 mm
DSC	0.57	0.68	0.87	0..95
JC	0.61	0.76	0.83	0.93

Table 5. Comparative performance of the proposed method

Performance	Swin-UNet [4]	AC-Net [7]	UNETR [9]	Fusion transformer [11]	This work
DSC	83.63	0.90	84.45	91.77	95.47
JC	0.87	0.82	0.78	0.80	0.93
Recall	99.33	92.0	91.3	92.31	98.12
HD	10.15	11.96	11.25	12.56	9.54
Precision	81.24	89.0	79.36	85.74	98.69
Accuracy	99.64	98.47	97.58	93.63	99.76
Specificity	99.65	97.25	96.47	95.60	99.89

5. CONCLUSION

The dual-stage deep learning model combines swin UNETR for segmentation alongside ViT for classification to establish an effective automated detection system for liver tumors and their segmentation in CT imaging data. The model achieves excellent segmentation outcomes because swin UNETR employs hierarchical self-attention together with ViT's global contextual learning to extract complex spatial and textural tumor features. The experimental findings show how the method delivers substantial enhancement of DSC values to 95.47% combined with higher JC levels at 0.93 and accuracy reaching 99.76% compared to conventional deep learning models. HD reaches a minimum value of 9.54 mm when applying the proposed method which ensures detailed boundary detection. The research demonstrates that uniting segmentation and classification models produces optimal results for liver tumor analysis. The upcoming research will extend the dataset while adding multi-modality imaging techniques and optimize computational operations to enable real-time clinical application. The designed framework can operate as an artificial intelligence diagnosis system which enhances both early liver cancer identification and better treatment strategy development.

FUNDING INFORMATION

The authors state that no funding was involved in this research.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Ananda R. Kumar Mukkala	✓			✓				✓		✓			✓	
G. Nivedita		✓				✓			✓		✓			✓
K. V. Balaramakrishna				✓	✓		✓			✓		✓		
Sujan Babu Vadde	✓		✓		✓			✓		✓				
Nandkishor Narkhede	✓					✓			✓				✓	
Suneel Kumar Asileti		✓			✓				✓			✓		✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors state that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

DATA AVAILABILITY

The data supporting the findings of this study are available from The Cancer Genome Atlas Liver Hepatocellular Carcinoma (TCGA-LIHC) project through the Genomic Data Commons Data Portal at https://portal.gdc.cancer.gov/projects/TCGA-LIHC. Access to certain controlled data may require authorization in accordance with the repository's access policies.

REFERENCES

[1] P. M. Kumar, H. Gohel, J. Selvaraj, and B. P. Kavin, "Livernet based segmentation of lesions from computed tomography scan for liver tumor detection," *Intelligent Data Analysis*, vol. 29, no. 5, pp. 1289–1312, 2025, doi: 10.1177/1088467X241301660.

[2] W. Yu, C. Wang, Z. Sun, S. Cheng, and K. Wang, "Segmentation of liver cancer pathology images using transformer-CNNs with high-resolution semantic features," in *2024 Chinese Intelligent Systems Conference*, Singapore: Springer Nature, 2024, doi: 10.1007/978-981-97-8654-1_35.

[3] K. Vijayaprabakaran, P. Ramalingam, R. Ramalingam, A. Ilavendhan, and R. Vedhapriyavadhana, "CUNet-CLSTM: a novel fusion of CUNet and CLSTM for superior liver cancer detection in CT scans," *IEEE Access*, vol. 13, pp. 66373–66392, 2025, doi: 10.1109/ACCESS.2025.3559592.

[4] M. Faheem, A. H. Khan, N. Zahra, and A. Y. Gill, "Optimizing liver cancer detection: leveraging gabor features and machine learning for enhanced abdominal CT image segmentation," *International Journal of Multidisciplinary Research and Growth Evaluation*, vol. 6, no. 2, pp. 224-229, 2025.

[5] M. Zhang, L. He, Y. Liu, and Y. Chao, "U-Trans Net: a hybrid of CNN and transformer network for liver and liver tumor segmentation," in *Third International Conference on Image Processing, Object Detection, and Tracking (IPODT 2024)*, 2024, doi 10.1117/12.3050404.
[6] S. Dharaneswar and B. P. S. Kumar, "An in-depth method for liver tumor segmentation and classification using trans-UNet and vision transformer-based residual attention network," in *5th International Conference on Electronics and Sustainable Communication Systems, ICESC 2024-Proceedings*, 2024, pp. 954–961, doi: 10.1109/ICESC60852.2024.10689998.
[7] 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," *Technology in Cancer Research and Treatment*, vol. 23, 2024, doi: 10.1177/15330338231219366.
[8] J. Ou, L. Jiang, T. Bai, P. Zhan, R. Liu, and H. Xiao, "ResTransUnet: an effective network combined with transformer and U-Net for liver segmentation in CT scans," *Computers in Biology and Medicine*, vol. 177, 2024, doi: 10.1016/j.compbiomed.2024.108625.
[9] D. Cheng, Z. Zhou, and J. Zhang, "EG-UNETR: An edge-guided liver tumor segmentation network based on cross-level interactive transformer," *Biomedical Signal Processing and Control*, vol. 97, 2024, doi: 10.1016/j.bspc.2024.106739.
[10] T. C. Phan, C. H. Ho, and A. C. Phan, "Detection and classification of liver lesions using vision transformer and active learning," in *Future Data and Security Engineering. Big Data, Security and Privacy, Smart City and Industry 4.0 Applications*, Singapore: Springer, pp. 208–224, 2024, doi: 10.1007/978-981-96-0434-0_15.
[11] W. Zhang *et al.*, "Automatic segmentation of liver tumor from multi-phase contrast-enhanced CT images using cross-phase fusion transformer," in *12th Asian-Pacific Conference on Medical and Biological Engineering*, Cham, Switzerland: Springer, vol. 103, 2024, pp. 121–130, doi: 10.1007/978-3-031-51455-5_15.
[12] E. Goceri, "A hybrid attention-based deep learning model for segmentation of livers and liver tumors from CT scans," *Multimedia Tools and Applications*, vol. 84, no. 37, pp. 46191–46212, 2025, doi: 10.1007/s11042-025-20760-y.
[13] N. Elazab, W. Gab Allah, and M. Elmogy, "Computer-aided diagnosis system for grading brain tumor using histopathology images based on color and texture features," *BMC Medical Imaging*, vol. 24, Jul. 2024, doi: 10.1186/s12880-024-01355-9.
[14] S. Stephe, B. S. Kumar, A. Thirumalraj, and V. Dzhyvak, "Transformer based attention guided network for segmentation and hybrid network for classification of liver tumor from CT scan images," *Eastern Ukrainian Medical Journal*, vol. 12, no. 3, pp. 692–710, 2024, doi: 10.21272/eumj.2024;12(3):692-710.
[15] H. Ma and M. Maimaiti, "U-Net with fully utilize of features for liver and liver-tumor segmentation in CT images," in *Pattern Recognition and Computer Vision,* Singapore: Springer, 2025, pp. 34-47, doi: 10.1007/978-981-97-8499-8_3.
[16] R. Mojtahedi, M. Hamghalam, R. K. G. Do, and A. L. Simpson, "Towards optimal patch size in vision transformers for tumor segmentation," in *Multiscale Multimodal Medical Imaging*, Cham, Switzerland: Springer, pp. 110–120, 2022, doi: 10.1007/978-3-031-18814-5_11.
[17] F. J. P. Montalbo, L. P. Palad, R. L. Castillo, K. I. Marasigan, and A. L. P. D. Ocampo, "CT scan liver and liver tumor 2D segmentation: a deep learning review and empirical analysis," in *2024 IEEE 12th Conference on Systems, Process & Control*, 2024, pp. 35–40, doi: 10.1109/ICSPC63060.2024.10862188.
[18] W. M. Idress *et al.*, "A lightweight Swin-UNet model for accurate liver tumor segmentation on memory-constrained devices," *Preprints Bioengineering,* Mar. 2025.
[19] Y. Wang, T. Wang, H. Li, and H. Wang, "ACF-TransUNet: attention-based coarse-fine transformer U-Net for automatic liver tumor segmentation in CT images," in *2023 4th International Conference on Big Data and Artificial Intelligence and Software Engineering, ICBASE 2023*, 2023, pp. 84–88, doi: 10.1109/ICBASE59196.2023.10303169.
[20] Y. Xia *et al.*, "Deep learning for oncologic treatment outcomes and endpoints evaluation from CT scans in liver cancer," *npj Precision Oncology*, vol. 8, no. 1, 2024, doi: 10.1038/s41698-024-00754-z.
[21] H. Qi, W. Wang, Y. Shi, and X. Wang, "AD-DUNet: a dual-branch encoder approach by combining axial transformer with cascaded dilated convolutions for liver and hepatic tumor segmentation," *Biomedical Signal Processing and Control*, vol. 95, 2024, doi: 10.1016/j.bspc.2024.106397.
[22] Q. Gu, H. Zhang, R. Cai, S. Y. Sui, and R. Wang, "Segmentation of liver CT images based on weighted medical transformer model," *Scientific Reports*, vol. 14, no. 1, 2024, doi: 10.1038/s41598-024-60594-6.
[23] R. Li *et al.*, "DHT-Net: dynamic hierarchical transformer network for liver and tumor segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 7, pp. 3443–3454, 2023, doi: 10.1109/JBHI.2023.3268218.
[24] M. A. Sattari *et al.*, "Liver margin segmentation in abdominal CT images using U-Net and Detectron2: annotated dataset for deep learning models," *Scientific Reports*, vol. 15, no. 1, 2025, doi: 10.1038/s41598-025-92423-9.
[25] A. S. Anwar, K. Amin, M. M. Hadhoud, and M. Ibrahim, "ResTransUNet: a hybrid CNN-transformer approach for liver and tumor segmentation in CT images," *Computers in Biology and Medicine*, vol. 190, 2025, doi: 10.1016/j.compbiomed.2025.110048.

BIOGRAPHIES OF AUTHORS



Dr. Ananda R. Kumar Mukkala an associate professor in the Department of Computer Science and Engineering, Sreyas Institute of Engineering and Technology, India with over 19 years of teaching, research, and academic leadership. He hold a Ph.D. in Computer Science and Engineering, along with M.Tech. and B.Tech. degrees in Computer Science and Engineering. His academic interests include artificial intelligence, machine learning, data structures, cloud computing, cyber security, blockchain, and web technologies. Throughout his career, he have been committed to delivering quality education, mentoring students, guiding research projects, and contributing to curriculum development and NBA/NAAC accreditation activities. He has published more than 33 research papers, authored the book "C–the basics," and hold three patents. He actively participate in faculty development programs, workshops, and professional organizations to enhance my knowledge and teaching practices. His mission is to inspire students through innovative teaching, research excellence, industry collaboration, and continuous learning while contributing meaningfully to the advancement of engineering education. He can be contacted at email: anandranjit@gmail.com.



G. Nivedita completed M.Tech. in Computer Science and Engineering from Geethanjali College of Engineering and Technology. Currently, she pursuing my doctoral research as a part-time research scholar at Amrita Vishwa Vidyapeetham, Bengaluru. Since 2014, she has been serving as an assistant professor in the Department of Computer Science and Engineering at Geethanjali College of Engineering and Technology, Cheeryal. With a decade of academic experience, she deeply committed to teaching, mentoring, and contributing to the academic growth of students. Her areas of interest include machine learning and computer vision, where she focus on applying advanced algorithms to solve real-world problems. She passionate about research and continuously strive to integrate the latest technological advancements into the work. She can be contacted at email: nive.gopigari@gcet.edu.in.



K. V. Balaramakrishna is an assistant professor in the Department of Electronics and Communication Engineering at Aditya University, Aditya Nagar, ADB Road, Surampalem, Kakinada District, Andhra Pradesh 533437, India. His research interests include communication systems, signal processing, and emerging technologies in electronics. He has contributed to various research publications and academic initiatives, focusing on practical and innovative approaches to engineering education. He is dedicated to fostering technical excellence and guiding students toward research-oriented learning. He can be contacted at email: balaram56kv@gmail.com.



Sujan Babu Vadde is an assistant professor in the Department of Computer Science and Engineering at Koneru Lakshmaiah Educational Foundation Deemed to be University, Vaddeswaram, Guntur District, Andhra Pradesh, India. His research interest includes deep learning, machine learning, artificial intelligence, and cyber security. He can be contacted at email: sujanbabu218@gmail.com.



Dr. Nandkishor Narkhede is basically an electronics engineer. He completed his doctorate degree in Electronics and Communications by using artificial intelligence and machine learning in 2023, from Sir Padampat Singhania University. He completed his masters in Electronics and Telecommunications from VESIT Mumbai University. He completed his bachelor's degree in Electronics Engineering from SGGS College of Engineering, Dr. B R Ambedkar Marathwada University. Currently he is working as an associate professor in Shah and Anchor Kutchhi Engineering College, Chembur, Mumbai. He published twelve papers in journals of which 4 are Scopus indexed. He published twelve papers in international conferences and two papers in national conferences. Two papers are published in Springer book chapter. He can be contacted at email: nandkishor.narkhede@sakec.ac.in.



Dr. Suneel Kumar Asileti is working as associate professor in the Department of Electronics and Communication Engineering, in Usha Rama College of Engineering and Technology, an autonomous National Assessment and Accreditation Council, a accredited college in Andhra Pradesh, India. He received Ph.D. for his work in Microstrip Patch Antenna in 2022 from Andhra University, Visakhapatnam, India. His areas of interest include micro strip patch antennas, image processing, electromagnetic fields and transmission lines, and VLSI. He has attended conferences and INUP hands on workshops conducted by IIT Bombay, and IIT Hyderabad. He has keen interests in microwave communication systems. He can be contacted at email: Suneelasileti@gmail.com.