International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 4 (2026)



ijerst.editor@gmail.com

editor@ijerst.com

ISSN 2319-5991 www.ijerst.com
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research Paper

LIVER CANCER DETECTION USING RESNET, SWIN TRANSFORMER AND CONVNEXT TINY NETWORKS

Yedluri Neeraja[1*] and Durga Ganga Rao Kola[2]

1,2 Department of ECE, UCEK, JNTUK, India

*Corresponding author: Yedluri Neeraja

E-mail: yedlurineeraja2002@gmail.com, ganga_408@yahoo.co.in

Address: Department of Electronics and Communication Engineering,
University College of Engineering Kakinada (Autonomous),
Jawaharlal Nehru Technological University Kakinada, Nagamallithota, Kakinada – 533003,
Andhra Pradesh, India

Abstract

Liver tumor detection from CT images is challenging because most CT slices do not contain tumors, while only a small number contain tumor regions. Another challenge is that slices from the same patient can be very similar, which may cause data leakage when splitting the data randomly. This study focuses on classifying liver CT slices into Tumor and No-Tumor classes using deep learning. The LiTS and 3Dircadb datasets are combined and used for the experiments. Three deep learning models, Swin-T, ConvNeXt-Tiny, and ResNet-50, are trained and evaluated. A patient-level data split is used to keep patients in the training, validation, and test sets separate. The three models are also combined using a soft-voting ensemble. The final evaluation is performed on an unseen test set containing 6,323 slices. The soft-voting ensemble achieved the highest accuracy of 92.74%. These results show that the models can detect tumor and no-tumor slices, while the class imbalance remains an important challenge for tumor detection.

Key words: Liver tumor detection, CT images, Deep learning, Swin Transformer, ConvNeXt, Softvoting ensemble.

1. Introduction

The liver is one of the most important organs in the human body because it performs several essential functions, including metabolism, detoxification, nutrient processing, and regulation of various body processes. However, the liver can be affected by different diseases, including liver cancer. Liver cancer can be difficult to detect at an early stage because its symptoms may not be obvious until the disease has been progressed. Therefore, early and accurate detection is important for timely treatment and better patient outcomes (Joshi et al., 2026).

CT images are widely used for examining the liver because they provide detailed content in images of its internal structures and can help identify abnormal regions such as tumors. In clinical practice, however, detecting tumors from CT scans often requires radiologists to examine a large number of image slices. Only a small proportion of these slices usually contain tumors, while most are normal. Another important challenge is data leakage. CT scans contain multiple adjacent slices from the same patient, and these slices are often highly similar. If individual slices are randomly divided into training and testing sets, images from the same patient may appear in both sets. This can lead to overly optimistic results because the model may learn patient specific patterns rather than general tumor characteristics.

To reduce this problem, the proposed study uses a patient level split ensuring that patients in the test set are completely unseen during training.

With the development of technology, various machine learning and deep learning techniques are used for liver cancer detection and classification to classify liver CT slices into Tumor and No-Tumor classes (Nisha and Mishra, 2026). In this work, mainly three deep learning models: ResNet-50, Swin Transformer, and ConvNeXt-Tiny, are implemented and compared. These models are trained to learn important features from CT images and identify tumor-containing slices.

For experimentation, LiTS and 3Dircadb datasets are used. However, there is strong class imbalance in these datasets makes automatic tumor slice classification challenging, as a model may achieve high overall accuracy while still performing poorly on the minority Tumor class. The predictions from the three models are combined using an ensemble method to improve the overall performance. The final model is tested on unseen patient data using Accuracy, Balanced Accuracy, Macro-F1, and ROCAUC. We also check the model's confidence and uncertainty to understand how reliable its predictions are. Grad-CAM is used to identify the important areas in the CT images that influence the model's prediction. Overall, this study focuses on developing an effective and reliable system for detecting liver tumor slices from CT images.

1.1 Related Work

Deep learning has become widely used in medical image analysis because it can automatically learn useful features from medical images. Different deep learning architectures have been used for image classification, segmentation, and disease detection.

Recent studies have shown that deep learning can be useful for liver tumor detection and analysis using CT images. Different deep learning models, such as CNNs and Transformers, have been used for classification and segmentation. Data augmentation techniques are also used to improve model performance, especially when the available medical data is limited. Some of the important works in liver cancer detection using deep learning are reviewed as follows:

Joshi et al. (2026) proposed a method called CustomLiverNet for liver cancer detection using CT images. Their method combines GAN, ResNet, and Vision Transformer. GAN is used to generate additional images, while ResNet and Vision Transformer are used to learn useful features from the images. Grad-CAM is also used to show the important areas that affect the model's prediction. Nisha and Mishra (2026) reviewed different deep learning methods used for liver and liver tumor segmentation. Their study explains that having enough high-quality medical data is important for training deep learning models. They also discussed the challenges caused by limited and poorly annotated data.

Qiu et al. (2026) proposed LiverFormer for liver segmentation. Their method combines CNN and Transformer models to learn both local and overall information from medical images. They also used data augmentation to improve the model when the available training data was limited.

He et al. (2024) used a deep learning method with CT images to predict microvascular invasion and pathological differentiation in liver cancer. Their method uses 3D U-Net for liver segmentation and Swin Transformer for classification. This study shows that Transformer models can be useful for learning important information from liver CT images.

Wang et al. (2026) proposed a method for liver cancer diagnosis using CT and MRI images. Their approach combines image features with radiological information. The study shows that using different types of information can help in liver cancer analysis.

Jest and Daniel (2024) proposed a Differential CNN with KELM for liver cancer detection. Their method focuses on improving the classification of liver cancer images. The study also discusses problems such as wrong classification, parameter selection, and computational complexity.

Tejaswi and Rachapudi (2024) developed a computer-aided liver cancer diagnosis system using an improved SegNet and a deep stacking ensemble model. Their method uses preprocessing and segmentation before classification. This study shows that segmentation and ensemble methods can be useful for liver cancer detection.

Gowda and Manjunath (2025) proposed a model called UNet70 for liver tumor classification using the 3Dircadb dataset. The model classifies CT images into Tumor and No-Tumor classes. They also used preprocessing and data augmentation to improve the training process.

Hamghalam et al. (2026) developed a liver cancer segmentation model for CT images. Their method also provides confidence scores to identify cases where the model may not be very reliable. This shows the importance of confidence and reliability in medical image analysis.

Kumar et al. (2026) proposed HETLesion for liver lesion segmentation using CT images. The model combines CNNs, attention mechanisms, and Transformers to learn information from different types of lesions. The study also uses explainable AI to understand the model's predictions.

2. Materials and Methods

2.1 Proposed Work

In this work, a deep learning-based method for liver tumor detection using CT scan images is proposed. The input images are preprocessed to improve their quality and make them suitable for the model. GAN is then used to generate additional training images and reduce the problem of limited data. ResNet, Swin Transformer and ConvNeXt-Tiny are used to extract useful features from the CT images. ResNet mainly captures detailed spatial features, while Swin Transformer learns broader information from different image regions. ConvNeXt-Tiny helps in learning additional high-level visual features from the CT images. Multi Head Attention is used to give more importance to the useful features. The features obtained from these networks are combined using a Customized Feature Fusion Layer. These fused features are passed through a Fully Connected Layer and a Softmax Layer for final classification. The system finally identifies the input image as Tumor or Non-Tumor. The overview of the proposed framework is shown in Figure 1.

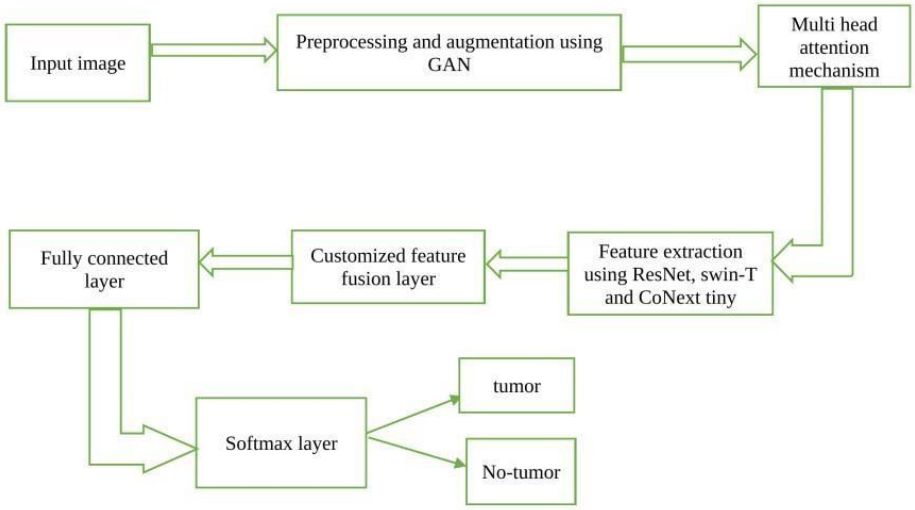

Figure 1. Overview of the proposed framework

2.2 Preprocessing

The proposed system takes a CT scan image of the liver as the input. The image contains information about the liver and the regions where a tumor may be present. This input image is given to the next stage for processing. Unwanted noise is reduced and the quality of the image is improved. The pixel values are also normalized so that the image can be processed easily by the deep learning model. After

preprocessing, GAN (Generative Adversarial Network) is used to increase the training data. GAN generates additional images similar to the original images. This helps when the available dataset has fewer images and also helps to reduce the difference between the classes. The processed images are passed through the Multi-Head Attention mechanism. This mechanism helps the model focus on the important information in the image. It gives more attention to useful features that can help in identifying a tumor.

2.3 Feature Extraction Using ResNet, Swin-T and ConvNeXt-Tiny

ResNet-50: It is a deep convolutional neural network that is commonly used for image classification and feature extraction. It is introduced with the idea of residual learning, where skip connections are used between layers. These connections help the network learn deeper features while reducing problems during training (Joshi et al., 2026).

ResNet-50 can learn different types of visual information from an image, such as edges, textures, shapes, and patterns. In medical image analysis, these features can be useful for identifying abnormal regions in CT images. In this work, ResNet-50 is used to extract important features from liver CT slices that can help distinguish between Tumor and No-Tumor images.

Swin Transformer: It is a Transformer-based deep learning model designed for computer vision tasks. Compared with regular CNN models, it uses an attention mechanism to understand relationships between different parts of an image. The image is divided into small windows and the model learns useful information from these regions.

Swin Transformer can capture both local and global image information. This is useful for CT images because tumor-related information may appear in different regions of the liver. In this work, Swin-T is used to extract important spatial and contextual features from the CT slices for tumor classification (Nasr Esfahani et al., 2026).

ConvNeXt-Tiny: It is a modern convolutional neural network developed by improving the traditional CNN architecture. It combines the advantages of convolution-based feature extraction with several design improvements inspired by Transformer models. The model can learn useful low-level and high-level features from images.

ConvNeXt-Tiny can identify patterns such as textures, shapes, edges, and other important visual information. Its relatively lightweight architecture also makes it suitable for image classification tasks. In this paper, ConvNeXt-Tiny is used along with ResNet-50 and Swin Transformer to extract complementary features from liver CT images. By combining these different architectures, a novel framework is proposed which learn a wider range of features for Tumor and No-Tumor classification. ResNet-50, Swin Transformer, and ConvNeXt-Tiny are considered because each model learns image features in a different way. These models are used to extract useful information from liver CT images for tumor and no-tumor classification.

Important features from the CT images are extracted using three different deep learning models: ResNet-50, Swin-T, and ConvNeXt-Tiny. ResNet-50 helps in learning detailed features such as edges, textures, and shapes. Swin-T captures information from different regions of the image and helps understand spatial features. ConvNeXt-Tiny learns visual patterns and high-level features from the CT images. Using these three models allows the system to learn different types of useful features, which can help in accurately classifying the images into Tumor and No-Tumor classes.

The features obtained from ResNet-50, Swin-T, and ConvNeXt-Tiny are combined using the Customized Feature Fusion Layer. This layer brings the useful information from these networks together. The combined features provide a better representation of the input CT image.

The fused features are given to the Fully Connected Layer. This layer processes the combined information and learns the relationship between the extracted features and the output classes. The output

from the Fully Connected Layer is passed to the Softmax Layer. It calculates the probability of the input image belonging to each class (Nasr Esfahani et al., 2026).

Based on these probabilities, the model makes the final classification. Finally, the system classifies the CT image into one of two classes: Tumor/No-Tumor.

3. Results and Discussion

The proposed study uses two publicly available liver CT datasets: 3Dircadb and LiTS for training and evaluating the liver tumor detection model. Both datasets contain CT images with corresponding liver and tumor mask information. The two datasets are combined to provide a larger number of images for Tumor and No-Tumor classification.

3.1 3Dircadb Dataset

The 3Dircadb dataset is a medical CT dataset used for liver analysis. In this study, 2,823 CT images are considered from the dataset (3Dircadb Dataset, Kaggle). The images have a size of 512 × 512 pixels and are provided with corresponding liver and tumor masks. The liver mask shows the liver region, while the tumor mask indicates the presence of a tumor.

Based on the tumor mask, each image is classified into two classes: Tumor and No-Tumor. If a tumor is present in the mask, the image is labeled as Tumor. If no tumor is present, the image is labeled as No-Tumor. The dataset contains more No-Tumor images than Tumor images, which creates a class imbalance problem. Therefore, suitable preprocessing and augmentation techniques are applied during model training. **3.2 LiTS Dataset**

The LiTS (Liver Tumor Segmentation) dataset is another publicly available CT dataset used in this study. It contains a large number of liver CT images along with liver and tumor masks. The dataset provides useful information for identifying tumor regions in liver CT scans (LiTS Dataset, Kaggle).

The LiTS dataset also has a large difference between the number of Tumor and No-Tumor slices. Most of the CT slices do not contain tumors, while only a small number of slices contain tumor regions. The LiTS dataset is combined with the 3Dircadb dataset to increase the amount of training data and improve the learning of the deep learning models.

After combining both datasets, the images are processed and labeled as Tumor or No-Tumor based on the tumor mask. The final data is then used to train and evaluate ResNet-50, Swin Transformer, and ConvNeXt-Tiny models.

3.3 Performance Measures

In this work, the performance of the proposed method is tested using accuracy, precision and recall. Accuracy is the fraction of correctly classified samples:

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

$$Precision = TP / (TP + FP) \quad (2)$$

$$Recall = TP / (TP + FN) \quad (3)$$

F1-score is the harmonic mean of precision and recall, reported both macro-averaged (unweighted mean over classes) and weighted-averaged (weighted by class support):

$$F1 = 2 \times (Precision \times Recall) / (Precision + Recall) \quad (4)$$

3.4 Results

Table 1 reports the checkpoints loaded for final testing and their recorded validation AUC. For ConvNeXt-Tiny, the training log also records a higher stage-1 validation AUC of 0.9037 at epoch 3, whereas the epoch-4 stage-2 checkpoint (validation AUC 0.8889) is loaded for test evaluation. Thus, the table describes the checkpoint actually evaluated, not the global maximum across both training stages. The reported test metrics correspond to the loaded epoch-4 checkpoint; a global-best checkpoint evaluation would require rerunning test inference.

Table 1. Checkpoints loaded for test evaluation and their logged validation AUC.

Model	Loaded Epoch	Validation AUC
Swin-T	4	0.8832
ConvNeXt-Tiny	4	0.8889
ResNet-50	4	0.8817

The ensemble weights are tested using only the validation set. The equal-weight soft-voting ensemble achieved a validation AUC of 0.9024. The optimized weights (0.2324 for Swin-T, 0.7581 for ConvNeXt-Tiny, and 0.0094 for ResNet-50) achieved a lower validation AUC of 0.8947. Therefore, equal weights of one-third for each model are used for the final test evaluation.

Table 2. Equal weights used for the final soft-voting test evaluation.

Model	Weight used for test
Swin-T	0.3333
ConvNeXt-Tiny	0.3333
ResNet-50	0.3333

Table 3 shows the final test results for the three individual models and the equal-weight soft-voting ensemble. The results are calculated on the held-out test set of 6,323 slices. Results are reported separately from the soft-voting ensemble in Table 3.

Table 3. Test-set performance of the three individual models and the soft-voting ensemble.

Model	Class	Precision	Recall	F1
Swin-T	No-Tumor	0.9540	0.9365	0.9500
Swin-T	Tumor	0.3051	0.4439	0.3617
ConvNeXt-Tiny	No-Tumor	0.9568	0.9601	0.9624
ConvNeXt-Tiny	Tumor	0.3750	0.3048	0.3363
ResNet-50	No-Tumor	0.9637	0.9507	0.9572
ResNet-50	Tumor	0.3546	0.4305	0.3889
Ensemble (soft)	No-Tumor	0.9640	0.9649	0.9626
Ensemble (soft)	Tumor	0.3724	0.3316	0.3508

Table 4 shows the accuracy on ensemble on the test set (5,949 No-Tumor and 374 Tumor slices).

Table 4. Accuracy of the proposed framework.

Model	Accuracy
Swin-T	0.9073
ConvNeXt-Tiny	0.9212
ResNet-50	0.9201
Ensemble (soft)	0.9274

4. Conclusion

This study developed and evaluated a deep learning system for classifying liver CT slices into Tumor and No-Tumor classes. The LiTS and 3Dircadb datasets are combined, and three models: Swin-T, ConvNeXt-Tiny, and ResNet-50, are trained and tested. A patient-level split is used so that the patients in the test set are not used during training.

The results show that all three models achieved good overall accuracy. The soft-voting ensemble achieved the highest accuracy of 92.74%, outperforming the individual models ConvNeXt-Tiny (92.12%), ResNet-50 (92.01%) and Swin-T (90.73%).

However, the lower balanced accuracy and Tumor-class recall and F1-score values (0.33 and 0.35 for the ensemble) compared with the overall accuracy show that detecting the minority Tumor class is still difficult because of the strong class imbalance. Overall, the results show that deep learning models can be useful for liver tumor slice classification, but further improvements are needed for better detection of tumor slices.

References

Wang Y, Wang S, Pan W, Ye S-P, Wang J-M, Abbas G, Yousef A H and Touti E (2026), "Auxiliary diagnosis of liver cancer using multimodal CT and MRI with deep learning and radiological feature fusion approach", *Ain Shams Engineering Journal*, Vol. 17, No. 3, 104014.

Sai Durga Tejaswi V and Rachapudi V (2024), "Computer-aided diagnosis of liver cancer with improved SegNet and deep stacking ensemble model", *Computational Biology and Chemistry*, Vol. 113, 108243.

Qiu L, Chi W, Xing X, Rajendran P, Li M, Jiang Y, Pastor-Serrano O, Yang S, Ji Y and Wen Q (2026), "Artificial intelligence-enhanced Couinaud segmentation for precision liver cancer therapy", *Biomedical Signal Processing and Control*, Vol. 120, 110100.

Nisha and Mishra P K (2026), "Data efficient deep learning for liver and liver tumor segmentation: A comprehensive survey", *Neurocomputing*, Vol. 663, 131937.

Nasr Esfahani A, Jalili A and Sajedi H (2026), "NeXtSwin-X: dual-branch cross-attention fusion of ConvNeXt and Swin Transformer for accurate brain tumor classification from MRI and CT", *Scientific Reports*, Vol. 16, 19959. https://doi.org/10.1038/s41598-026-50158-1

Maria Jest P and Antony Asie Daniel V (2024), "Differential CNN and KELM integration for accurate liver cancer detection", *Biomedical Signal Processing and Control*, Vol. 95, 106419.

LiTS Dataset, Kaggle. Available at: https://www.kaggle.com/datasets/harshwardhanbhangale/litsdataset

Kumar P, Sivesha V, Suryanarayana T V V, Mohanavel V, et al. (2026), "HETLesion: A hierarchical extended transformer framework for liver lesion segmentation in CT images with XAI", *Neurocomputing*, 133582.

Joshi S, Dwivedi A, Kumar R, Kumar A, Kumar R and Amrita (2026), "Enhancing liver cancer detection: An innovative deep learning approach combining GAN, ResNet, and vision transformer", *Expert Systems with Applications*, Vol. 298, 129734.

He X, Xu Y, Zhou C, Song J, Lin Y, Zhang H, Wang Y, Fan Q, Wang D, Chen W, Wang J and Guo D (2024), "Prediction of microvascular invasion and pathological differentiation of hepatocellular carcinoma based on a deep learning model", *European Journal of Radiology*, Vol. 172, 111348.

Hamghalam M, Peoples J J, Kobayashi K S M, Park G, Kwak E, Bunker E C, Gangai N, Gonen M, Chun Y S, Kang H C, Do R K G and Simpson A L (2026), "Liver cancer segmentator: Metadata-guided confidence scoring for reliable segmentation of colorectal liver metastases in CT", *Computer Methods and Programs in Biomedicine*, Vol. 276, 109233.

3Dircadb Dataset, Kaggle. Available at: https://www.kaggle.com/datasets/priyamsha1717/3dircadbdataset

Gowda Y and Manjunath M R V (2025), "Automatic liver tumor classification using UNet70 a deep learning model", *Journal of Liver Transplantation*, Vol. 18, 100260.