Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 8, No. 2, April 2026, pp: 553-571 e-ISSN: 2656-8632

RESEARCH ARTICLE OPEN ACCESS

A Hybrid Deep Ensemble Model for Precise Liver and Tumor Segmentation Using U-Net and W-Net Architectures

B. Sravani and M. Sunil Kumar

Department of CSE Mohan Babu University, A. Rangampeta, Tirupati, Andhra Pradesh, India
Corresponding author: Sravani. B (e-mail:bukkesravani1992@gmail.com, **Author (S) Email:** M. Sunil Kumar (e-mail: sunil.malchi@mbu.asia)

Abstract **The identification of the liver with the hepatic tumors on the computed tomography (CT) scans is a major compulsion to the earliest diagnosis, treatment planning, and surgery in the case of hepatocellular carcinoma. However, automated segmentation is not an easy job due to the non-homogeneous appearance of tumors, blurry boundaries, small size of annotated datasets, and high inter-slice variability. Existing single deep learning models are known to suffer from prediction variance and low generalization in complex clinical conditions. The primary goal of the study is to develop an effective, highly accurate segmentation model that improves the accuracy, consistency, and explanability of liver and tumor borders in CT images. In this paper, an original hybrid deep ensemble model is proposed that leverages the advantages of U-Net and W-Net. This is the primary contribution; one can combine the strong spatial localization ability of U-Net and the reconstruction-driven unsupervised learning ability of W-Net in minimizing the variance and maximizing the generalization. In addition, soft probability fusion, uncertainty modelling, and entropy-based confidence estimation are also introduced to improve reliability and clinical interpretation. The preprocessing of CT images is performed mathematically by normalizing and resizing to 256x256. U-Net and W-Net are trained separately using the pixel-wise probability maps, which are soft-averaged and thresholded. Benchmark liver CT datasets are tested with the ensemble using the Dice coefficient, accuracy, precision, recall, F1-score, Intersection over Union (IoU), ROC-AUC, and statistical significance tests. The results of the experiment show that the suggested ensemble performs better with an accuracy of 95.4, a precision of 94.3, a recall of 93.9, an F1-score of 94.1, IoU of 89.8, and an average ROC-AUC of 0.9615 than the models of the U-Net and W-Net, which differ in a huge number. Statistical confirmation that the improvements are relevant ($p < 0.01$) will be provided. In summary, the proposed deep ensemble segmentation can accurately, reliably, and effectively segment the liver and tumor, showing strong potential for clinical use and subsequent extension to multi-organ and multi-modal medical imaging.**

Keywords **Hepatic segmentation, Neoplasm segmentation, Deep learning, Ensemble learning, Medical image analysis, Computed Tomography (CT), Semantic segmentation**.

I. Introduction

Liver cancer is still one of the most lethal cancers worldwide because most cases remain asymptomatic and have a high mortality rate. Early diagnosis is important because therapy is much more effective and patient survival is higher. Imaging technologies, computational diagnosis, and molecular biology driven development of numerous detection assays [1]. This report provides a detailed overview of established and emerging techniques used in the detection of liver cancer, with a special focus on their methodologies, advantages, and limitations. Imaging is essential in the diagnosis and monitoring of HCC. The following imaging methods are commonly used in routine clinical practice. It is widely used ultrasound in the surveillance of HCC, especially in the high-risk population. It is a non-invasive, inexpensive tool that allows live imaging of liver morphological structure and is also frequently the first screening technique used to detect hepatic issues [2]. CT allows finely detailed cross-sectional imaging of the liver and therefore provides information about the size and position of the lesion, as well as any possible metastasis. It is commonly used for further evaluation after an initial ultrasound. This review underlines the crucial role of imaging in the early and accurate diagnosis of HCC. Both modalities have their own advantages; sometimes, both methods are used together for better diagnostic accuracy. The integration of these imaging

Manuscript received 24 July 2025; Revised January 2026; Accepted 20 March 2026; Available online 2 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1089

modalities into routine clinical care is critical for improving prognostication and guiding effective treatment strategies [3]. Recent developments in artificial intelligence (AI) and deep learning have greatly enabled the automated analysis of medical images. CNNs and encoder-decoder models, in particular, the U-Net have demonstrated good results in biomedical segmentation. U-Net not only maintains spatial information by using skip connections but also effectively learns hierarchical features, which is why it is a popular segmentation of organs. The extension of U-Net to W-Net, which are two cascade U-Nets, also makes it possible to perform unsupervised reconstruction-based learning which assists in retrieving structural information even when the labelled data is scarce. Even with this development, the problem of applying a single model with regard to the complexity and heterogeneity of liver lesions has been unsuccessful in terms of addressing multi-phase CT scans, as well as small lesions and irregular tumor boundaries [4].

The significant findings of the research can be summarized as follows. To begin with, the paper will propose a new hybrid deep ensemble segmentation system that efficiently combines the U-Net and W-Net architectures to address liver and tumor segmentation in CT images [5]. The proposed framework offers the benefits of high spatial localization of U-Net with reconstruction-based unsupervised learning power of W-Net and thus significantly lowers prediction variance and enhances generalization, especially when small, irregular, and low-contrast tumor areas are involved, which are challenging to segment with single models. Second, an ensemble fusion strategy that uses uncertainty modelling is presented, in which the pixel probability maps of both networks are soft-averaged with thresholding. The design provides stronger segmentation and allows entropy-based uncertainty estimation and visualization of the heatmap of activated areas, thus making the model more interpretable and clinically trustworthy, which is typically not a quality of current deep learning-based medical segmentation methods. Third, the suggested framework is widely tested on publicly available benchmark datasets using detailed quantitative metrics and statistical significance analysis. Its results show that the performance is consistent with single models and competitive with state-of-the-art approaches, with higher accuracy, Dice, F1-score, IoU, and ROC-AUC. All these make the proposed ensemble an accurate, reliable, and clinically viable solution to automated liver and tumor segmentation, and it has great potential to be extended to multi-organ and multi-modal medical imaging applications.

II. Related Work

The accurate differentiation between liver and neoplastic regions in CT images plays a crucial role in clinical diagnosis, therapy planning, and monitoring of patients. This activity is gaining a lot of attention in the community of medical image analysis due to the central role it plays in hepatic oncology. Conventionally, segmentation has been performed annually or using classical image processing techniques. However, these methods were time-consuming, prone to errors, and often lacked the robustness for clinical reality . The emergence of machine learning/ deep learning opened the way for automatic, scalable, and highly accurate segmentation systems.

A. Legacy and Pre-Deep Learning Approaches

Recently, developments in artificial intelligence (AI) and deep learning (DL) have profoundly altered the processes of liver cancer diagnosis, segmentation, and prognosis. Convolutional neural networks (CNNs), hybrid networks, and nature-inspired optimization methods have been widely researched to build automated, efficient, and explainable medical imaging systems. In the MRI multi-scale, multi-level fusion framework for diagnosing liver lesions, there was high visual interpretability, which is important for clinical validation. Likewise, a liver tumor-specific segmentation model based on CNN showed high precision in tumor boundaries demarcation[6]. The combination of optimized backbone networks with liver-specific architectures also enhanced classification accuracy, reflecting the success of hybrid learning approaches in internalizing varying feature representations [7].

B. The Rise of Deep Learning

CNN-based feature extraction with traditional machine learning classifiers, namely support vector machines, has demonstrated the feasibility of integrating deep and shallow learning paradigms for liver tumor classification [8], [9]. The models of deep learning have also been used in prognosis and recurrence prediction of hepatocellular carcinoma post-surgical resection or transplant and have demonstrated promising prediction accuracy [10], [11]. AI-based methods using multimodal clinical and imaging data have also been used to further improve outcome prediction in liver cancer [12]. Also, AI-powered biomarker discovery and disease-stage models have helped in better prognostic assessment of clinical settings [13], [14], [15]. Premature convergence has been reduced by nature-inspired optimization methods to segmentation models, improving both speed and accuracy, especially with swarm-based optimization methods [16], [17], and [18].

C. Unsupervised, Semi-Supervised Learning

A combination of diverse data has demonstrated high potential for enhancing liver cancer diagnosis, especially with multimodal learning systems that incorporate both imaging and genomic data [19], [20].

Manuscript received 24 July 2025; Revised January 2026; Accepted 20 March 2026; Available online 2 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1089

Although these improvements have been made, AI systems are still perceived as black boxes, thereby restricting their use in clinical practice without proper explainability features. Segmentation architectures based on encoder-decoder models with hybrid segments have shown better results on heterogeneous datasets [21]. AI paradigms that are explainable and privacy-preserving have been gaining more and more attention. Federated and decentralized models have been presented to guarantee that data privacy during model training is guaranteed [22], [23]. To enhance clinician confidence and clinical usability, explainable AI methods that prioritize transparency and interpretability have been given priority [24], [25], [26]. AI-based systems have also been applied to the study of liver fibrosis and the examination of liver pathology in general [27], and visual attention systems have been applied to the interpretation of MRI to enhance explainability [28].

D. Group Learning of Medical Image Segmentations

Deep learning pipelines for detecting and segmenting liver cancer in CT and MRI images have shown great applicability and flexibility in real-life clinical practice[29], It has also been shown that transfer learning and pretrained CNN models enhance early-stage liver disease detection, especially in situations with limited labeled data Pretrained architectures have been found to be effective on small datasets reliably In addition to imaging, AI has been applied to genomic analysis and digital pathology, where it can provide precision oncology through molecular profiling and interpretation of biopsy slides . Taken together, these advancements shed light on an increasing tendency to apply AI across the liver disease workflow, with a strong focus on accuracy, explainability, and clinical relevance. Transparency is one of the factors found to increase clinical trust and adoption . The first machine learning methods to classify liver tumors provided the baseline, which supported the fast development of the deep learning-based methods [30].

E. Data set Tests and Real-World Challenges

The publicly available datasets, such as LiTS (Liver Tumor Segmentation) and 3DIRCADb, have become standard benchmarks for liver and tumor segmentation algorithms. Although these datasets have advanced methods, they also present practical difficulties. Definite delimiting of tumor boundaries has continued to be a challenge, especially in small, diffuse, or infiltrative lesions. Also, the segmentation models tend to have low generalization when used in other institutions, scanners, or imaging protocols because of domain changes. [31]Also, most deep learning systems are black-box systems, and without proper explainability systems, they cannot be adopted in clinical environments. Recent studies have revealed that incorporating advanced learning paradigms can promote diagnostic reliability and robustness Radiomics-based models have been shown to predict histopathological grades of hepatocellular carcinoma and to enable the incorporation of imaging biomarkers into deep learning pipelines Other advances are multimodal and multi-omics model methods, deep transfer learning models to enhance feature reuse and generalization [32], and AI-based biomarker learning models to enhance disease characterization . The technologies of automated disease staging, federated learning with privacy-conscious model training, and integration of the clinical workflows also enhanced the practicality of AI-based liver analysis systems in practice [33],

F. Proposed Framework and Rationales

To overcome the above, the current study develops a new deep ensemble model and combines the U-Net and W-Net frameworks. It is a hybrid design that leverages the strong ability of U-Net to localize features spatially and the ability of W-Net to learn features in an unsupervised, reconstruction-driven manner to provide more robust and context-sensitive segmentations[34] The framework also integrates the mechanisms of advanced visualization and uncertainty quantification, such as activation heatmaps and entropy-based uncertainty maps, to improve the interpretability and aid clinical decision-making.

Although liver and tumor segmentation has advanced significantly compared to traditional image processing methods and is now being solved by deep learning, there are still limitations to individual models with regard to data variability, complexity, and prediction instability. The problem of ensemble learning is that it provides a good way of addressing these problems by removing variance and leveraging the strengths of the complementary models.[35]The U-Net and W-Net synergistic integration is also a comparatively unexploited direction in the current literature. It is hoped that the proposed ensemble structure will address this gap by showing better segmentation results on complicated CT images, and also encourage the creation of reliable, explainable, and clinically useful automated systems to hepatic cancer.

1. Liver Cancer Detection Techniques

Liver cancer continues to be one of the deadliest cancers globally, mostly because of its asymptomatic characteristics in the early stages and the elevated mortality linked to late-stage detection. Prompt detection is essential, as early identification markedly increases the likelihood of effective therapy, thereby boosting survival rates and patient outcomes.[36] Progress in imaging technologies, computational diagnostics, and molecular biology has led to the development of a wide array of diagnostic instruments. This section delineates both conventional and

Manuscript received 24 July 2025; Revised January 2026; Accepted 20 March 2026; Available online 2 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1089

contemporary liver cancer detection techniques, assessing their principles, benefits, and drawbacks.

2. Techniques for Detection Based on Imaging

Medical imaging is pivotal in the detection and assessment of liver cancer.[49] It enables the identification, characterization, and staging of hepatic neoplasms.[37]The most often employed imaging modalities comprise. Ultrasound is a prevalent, non-invasive imaging technique utilized as a primary screening method, especially in high-risk groups. It facilitates real-time visualization of hepatic structures and is commonly used to detect hepatic anomalies and to assist in biopsy procedures.

3. Serum Biomarkers

Biomarkers are also measurable substances found in body fluids that reflect the presence of a disease. In the liver cancer context, multiple blood-based biomarkers are widely used in screening detection, diagnostic and monitoring. Glycosylation of AFP produces a highly cancer-related form, AFP-L3, which has significantly increased specificity for diagnosis of hepatic cancer compared to AFP. This abnormal prothrombin is frequently elevated in hepatocellular carcinoma patients and represents a supplement to alpha-fetoprotein when this is not elevated. [38]The ability of these novel markers to improve early detection and diagnostic [specificity in combination with established biomarkers is currently under investigation. Liver biopsy is the gold standard to diagnose HCC, which involves tissue sample submission and microscopic review. With recent innovations in AI and machine learning, liver cancer diagnosis has been greatly enhanced in terms of accuracy and efficiency. Computer-Aided Diagnosis (CAD) systems are designed to support radiologists by highlighting suspicious areas in medical imaging examinations, pressing potential decision-making, and reducing human error.

CNNs for liver cancer. Computer-aided detection has become a major tool in liver cancer diagnosis, based on which the automatic liver cancer identification, classification, and segmentation - particularly in CT or MRI it is used.[39] Radiomics is the process of obtaining a large number of quantitative features from medical imaging and using these features in predictive models. Combining radiomics features and clinic pathologic variables, such as age, sex, liver function tests or AFP, can enhance the diagnostic and prognostic efficiency. Molecular profiling is transforming the diagnosis of liver cancer - leading to early detection and personalized treatment strategies. Non-invasive means of identifying circulating tumor DNA (CTDNA) or RNA in blood samples. Next Generation Sequencing (NGS) methods can be used to identify specific genetic mutations and modifications associated with liver carcinogenesis. There has been identification of potential RNAs expressed abnormally that are associated with HCC. These small non - coding RNAs are currently being studied as potential candidates for diagnosis as well as prognosis.[40] DNA methylation is one of the major epigenetic changes that occurs during cancer origin and progression. These can be detected in a tissue or blood sample for diagnostic purposes. New non-invasive devices provide additional insight into hepatic tissue and neoplastic biology.

1) Transient Electrography (Fibro Scan) quantifies liver stiffness to evaluate the degree of fibrosis or cirrhosis, both of which are risk factors for hepatocellular carcinoma (HCC).
2) Contrast-Enhanced ultrasonography (CEUS) enhances lesion detection and characterization compared to conventional ultrasonography by viewing real-time vascular dynamics .
3) Optical and Photo acoustic imaging remain predominantly experimental, providing high-resolution structural and molecular imaging with the capacity to identify malignancies at an early stage.

The integration of many diagnostic modalities markedly enhances the precision and dependability of liver cancer diagnosis.[41]PET/MRI integrates metabolic and anatomical imaging, improving lesion characterization, while AI-driven platforms that combine radiomics, genomics, and clinical factors provide a more thorough diagnostic profile. Integrative techniques are especially beneficial in hepatic oncology, where accurate imaging is essential for identification, therapy planning, and monitoring disease development.[42] Computed Tomography (CT) has emerged as a fundamental method in hepatic diagnosis among medical imaging modalities. Its capacity to provide high-resolution, cross-sectional pictures renders it essential for the identification and assessment of liver tumours, including hepatocellular carcinoma (HCC), and hepatic metastases from extra hepatic malignancies.[43] Precise delineation of hepatic components and neoplastic margins is essential for clinicians to ascertain tumour extent, devise therapies such as resection or radiotherapy, and evaluate therapeutic results. Although manual segmentation by professional radiologists is considered the gold standard, it is inherently time-consuming, labour-intensive, and susceptible to inter-observer variability. The aforementioned restrictions have spurred increasing interest in automated and semi-automatic segmentation techniques, designed to improve workflow efficiency and uniformity in clinical practice[44]Nonetheless, despite progress in computer vision and medical image analysis, automated segmentation of the liver and tumours continues to be a challenging endeavour,[45] principally because of Heterogeneous Intensity Profiles: Tumour locations frequently display varied intensity and uneven textures, hindering differentiation from surrounding tissues.

Manuscript received 24 July 2025; Revised January 2026; Accepted 20 March 2026; Available online 2 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1089

However,[46] CT scans may exhibit motion blur, diminished signal-to-noise ratios, and insufficient contrast between lesions and healthy tissue.[47]Diverse Tumour Morphology: Hepatic tumours exhibit significant variability in size and shape, necessitating flexible models that generalize effectively.[48]The liver's adjacency to other abdominal organs results in indistinct boundaries, particularly in advanced stages of tumours. The dearth of high-quality labelled data, necessitating considerable time and expertise for annotation, impedes the advancement of powerful deep learning models.[49]Convolutional Neural Networks (CNNs) have revolutionized medical image processing by achieving superior performance in several segmentation tasks. The U-Net design has emerged as a fundamental model owing to its encoder-decoder framework and skip connections, which maintain spatial detail throughout the down-sampling and up sampling processes.U-Net has been effectively utilized across multiple domains, such as brain MRI, lung CT, and retinal imaging. Various adaptations, including 3D U-Net, attention U-Net, and multi-scale U-Net, have been created to improve contextual learning and feature extraction.The W-Net is a robust architecture that combines two U-Nets in a stacked configuration. This model facilitates unsupervised learning by integrating image reconstruction with segmentation, thereby enhancing feature representation without complete annotations.[50] The two-pass structure is particularly efficient in semi-supervised contexts, where labelled data is limited. No single design consistently outperforms the others across all datasets and imaging settings, despite their respective capabilities.[51]Model performance is frequently influenced by image quality, anatomical diversity, and dataset complexity Ensemble learning, a technique that integrates many models, has proven to be an excellent approach for enhancing predictive accuracy and generalization. Our methodology utilizes the synergistic strengths of U-Net and W-Net. U-Net demonstrates superior spatial localization, whereas W-Net enhances feature learning via its reconstruction-oriented purpose.[52]The ensemble combines predictions from both networks to provide enhanced and precise segmentation, especially in intricate and unclear clinical scenarios, such as multifocal lesions or tumours adjacent to organ borders.

4. Computed Tomography (CT)

Table 1. Summary of serum biomarkers used in liver cancer diagnosis. CT imaging provides detailed cross-sectional images of the liver, facilitating thorough anatomical evaluation. It is frequently used for subsequent assessment after abnormal ultra-monograph results and is essential for staging liver cancer.

Table 1. Serum Biomarkers

Author(s)	Year	Title	Methodology	Imaging/Techniques	Dataset Used	Key Findings
Wang et al. [24]	2024	Liver classification	Pretrained CNN	MRI	Public, Private	Effective with small datasets
Liu et al. [9]	2016	Fibrosis assessment	Transient Elastography	Ultrasound (US)	Hospital dataset	Accurate fibrosis staging
Park et al. [46]	2022	PET/MRI in oncology	Hybrid PET/MRI	PET/MRI	Clinical	Improved diagnosis via hybrid imaging

III. Method

After completion of liver cancer detection techniques, this methodology section summarizes the deep ensemble framework with representation of robust liver and tumour segmentation. The method has four large steps, namely, pre-processing of the data, design of the dual-model architecture, ensemble prediction, and analysis of the performance. All its components are properly designed in such a way that they guarantee better accuracy in segmentation and generalization. This study presents a deep ensemble segmentation framework that combines U-Net and W-Net architectures for the automated segmentation of liver and tumours in CT scans. The primary contributions of the study encompass. Dual-Model Ensemble Architecture. Ensemble fusion the ensemble fusion takes the pixel-wise probability map of U-Net and W-Net and averages them together by soft averaging, which is computed as Eq. (1)(2)[1].

$$P_{ensemble}(x) = \frac{1}{2}\left(\left(P_{U-Net}(x)\right) + P_{W-Net}(X)\right) \quad (1)$$

where P(x) refers to the probability which is predicted at pixel. $P_{U-Net}(x)$ is probability predicted by U-Net and $P_{W-Net}(X)$ is probability predicted by W-Net, $P_{ensemble}(x)$ is final averaged prediction and $\frac{1}{2}$ refers to weighting of both models. Another thresholding of the fused probability map at 0.5 is then

Manuscript received 24 July 2025; Revised January 2026; Accepted 20 March 2026; Available online 2 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1089

used to obtain the final binary mask. This can be added to figure 1 to have better reproducibility of this fusion stage. We employ a hybrid framework that consolidates U-Net and W-Net outputs through soft averaging and thresholding to provide final segmentation masks.

1. **Soft probability Averaging**

A hybrid modeling approach P_E (x) that combines the results of several separate models to enhance overall predictive performance is represented by the ensemble prediction function. The ensemble output in this formulation is calculated as the mean of two component model predictions, specifically $P_{U\ (x)+}P_{W\ (x)}$. The strategy seeks to maximize each model's advantages while reducing each one's shortcomings by combining these forecasts. In complicated tasks like disease prediction and medical image analysis, where single-model performance may be limited, this averaging method improves resilience, lowers variation, and frequently produces more accurate and consistent findings (Eq.(2))

$$P_E\ (x) = \frac{1}{2}\left(P_{U\ (x)+}P_{W\ (x)}\right) \quad (2)$$

where P_E (x) final ensemble probability output and (x)is input image P_U(x) probability predicted by the U-Net model, P_W(x) probability predicted by the W-Net model, $\frac{1}{2}$ refers to the weighting of both models. A binary mask $M_{E(x)}$then becomes known by thresholding Eq. (3) [15]

$$M_{E(x)=}\begin{cases}1, \text{if} P_E\ (x) \geq \tau \\ 0, \text{otherwise},\end{cases} \quad (3)$$

where $M_{E(x)}$ final binary output, P_E (x) is ensemble probability at x, τ is threshold value,1 is the foreground class and 0 is the background class. Where $\tau = 0$.This averaging reduces the variance of prediction errors between the two networks.

2. Weighted Linear Fusion

To Adaptively balance the confidence of each model, we assign non-negative scalar weights w_U and w_W such that $w_U + w_W = 1$ Eq. (4)[26]

$$P_E\ (x)=w_UP_U(x) + w_WP_W(x) \quad (4)$$

where P_U(x) probability prediction from the U-net model, P_W(x) probability prediction from the W-Net model, w_U weight assigned to the U-Net prediction, w_W weight assigned to the W-Net prediction. P_E(x) refers to the final ensemble probability output. The best weights are obtained by the minimization of the mean-squared error between the fused prediction and the ground truth Y(x) (Eq. (5)) [28].

$$\frac{min}{w_U,w_W}\sum_{x\in\Omega}Y(x)[\text{wUPw}(x)+\text{wWPW}(x) \quad (5)$$

where $\frac{\min}{w_U,w_W}$ optimization goal to find weights of w_U and w_W ,Ω is the set of all input pixels, (Y(x) ground truth for input x, Subject to w_U , $w_W \geq 0 \text{and } w_U + w_W = 1$.

3. Logit Space Fusion

To achieve a more stable combination when predictions are highly confident, the outputs are transformed into log-odds (Eq.).(6)[39].

$$Lu\ (x) = \log\frac{Pu}{1-Pu\ (x)}, Lw(x) = \log\frac{Pw\ (x)}{1-Pw\ (x)} \quad (6)$$

where P_U(x) probability predicted by the U-Net model, P_W(x) probability predicted by the W-Net model, Lu (x)logit value from U-Net, Lw(x)logit value from W-Net. The ensemble is expressed as (7) [37]

$$L_E\ (x) = \alpha_UL_U\ ((x) + \alpha_wLw\ (x) \quad (7)$$

where Lu (x)logit output from U-Net, Lw(x)logit output from W-Net. α_U weight assigned to the U-Net logit, α_w weight assigned to the W-Net logit, L_E (x) final ensemble logit value. And the fused probability becomes Eq. (8) [38]

$$P_E\ (x) = \sigma\big(LE(x)\big) = \frac{1}{1+\exp\ (-L_E\ (x))} \quad (8)$$

where α_U, α_w are learnable fusion coefficients and $\sigma(,)$denotes the sigmoid function.

4. Fusion Loss Function

The ensemble parameters are optimised on a joint loss of Binary Cross-Entropy (BCE) and Dice loss Eq. (9) [14].

$$L_{\text{fusion}=\lambda_1L_{BCE+}\ \lambda_2L_{Dice}} \quad (9)$$

where L_{fusion} final fused loss used for model training L_{BCE+}binary cross entropy loss measuring pixel-wise classification error, $L_{Dice}$is dice loss overlap between predicted and ground truth masks, $\lambda_1$weight controlling contributions of BCE loss, λ_2 weight controlling contribution of DICE loss. (10) (11) [11]

$$£_{BCE}= -\frac{1}{\Omega}\sum_{x\in\Omega}[Y\ (x)\log P_E\ (x)+(1\text{-}Y(x))\ \log P_E\ (x))] \quad (10)$$

where $£_{BCE}$ binary cross entropy loss, Ω is the set of all input pixels, x is a pixel, Y(x) ground truth label, P_E (x)ensemble predicted probability at x and log is natural logarithm.

$$£_{\text{Dice}=1-\frac{2\sum_zY\ (x)P_E\ (x)\ +\epsilon}{\sum_xY\ (x)^2+\sum_zP_E\ (x)^2+\epsilon}} \quad (11)$$

where $L_{Dice}$is dice loss, x is pixel, Y(x) ground truth label, P_E (x)ensemble predicted probability at x , ∈ small constant to avoid division by zero, $\sum_x$summation of all the pixels, the constants λ_1, λ_2 control the trade-off and ∈ prevents division by zero.

5. Variance Reduction and Uncertainty Modelling

Are the prediction variances of U-Net and W-Net, respectively, and Let σ_U^2 and σ_W^2 denote the prediction variances of the U-Net and W-Net models, respectively, and let ρ represent the correlation coefficient between their predictions. Under a soft-averaging ensemble strategy, the variance of the

Manuscript received 24 July 2025; Revised January 2026; Accepted 20 March 2026; Available online 2 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1089

ensemble prediction σ_E^2 can be expressed as Eq.(12) [5]

$$\sigma_E^2 = \frac{1}{4}\left(\sigma_U^2 + \sigma_W^2 + 2\rho\sigma_U\sigma_W\right) \quad (12)$$

where σ_E^2 variance of the ensemble prediction, σ_U^2 variance of the U-Net prediction, σ_W^2 variance of the W-Net prediction, ρ correlation coefficient between U-Net and W-Net predictions, $\frac{1}{4}$ scaling factor due to averaging two models. This formulation indicates that the ensemble variance depends not only on the individual model variances but also on the degree of correlation between their prediction errors. When the correlation coefficient approaches zero ($\rho \to 0$), the covariance term diminishes, resulting in a lower ensemble variance. Consequently, reduced inter-model correlation yields more stable and reliable ensemble predictions, thereby explaining the improved robustness of the proposed fusion strategy [6]. To further quantify prediction uncertainty at the pixel level, entropy-based uncertainty estimation is employed. n the ensemble probability output $P_E(x)$. The entropy measure is defined as: Eq.(13)[13].

$$H(x) = -P_E(x)\log P_E(x) - (1 - P_E(x))\log(1 - P_E(x)) \quad (13)$$

where x is a pixel, $H(x)$entropy of the ensemble prediction at Y(x) ground truth label, $\mathrm{P_E}(\mathrm{x})$ensemble predicted probability at x and log is the natural logarithm. Higher entropy values correspond to regions of increased uncertainty, typically observed near ambiguous tumor boundaries or low-contrast areas. This uncertainty modeling not only enhances interpretability but also provides a mechanism to identify regions that may benefit from further clinical review. High entropy regions indicate indistinct tumor edges and can be put under manual inspection.

6. Final Decision Rule

The Final segmentation mask is obtained as Eq.(14) [18]. The function is the last binarization decision rule based on the ensemble prediction. according to a predetermined value. τ τ. Precisely, the output will be given a value based upon how the probability given was predicted to be above the threshold hence transforming continuous probabilistic predictions to discrete class labels. This computation is frequently employed in the classification and segmentation processes to guarantee uniform and understandable decision-making, especially in the medical image processing where a clear separation of target and non-target area is the key to the correct diagnosis.

$$\mathrm{Y(x)} = \mathbb{I}\left[\mathrm{P_E}\,(\mathrm{x}) \geq \tau\right] \quad (14)$$

where $\mathbb{I}$ is the indicator function? The resulting mask combines both networks with spatial accuracy and context, with better accuracy and anatomically consistent boundaries, which generate better results in terms of Dice and IoU scores. $\mathrm{P_E}\,(\mathrm{x})$ ensemble predicted, τ is threshold value, $\mathrm{Y(x)}$ binary label at pixel

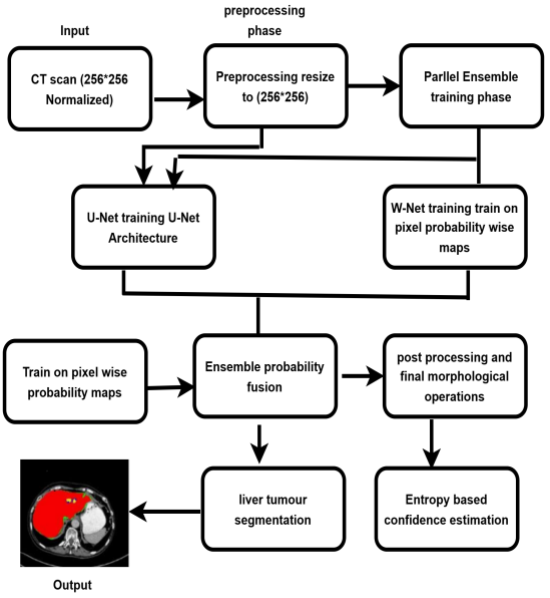

Fig1 Architecture of the proposed deep ensemble framework integrating U-Net and W-Net

In Fig.1. Unsupervised Mask generating: In scenarios devoid of ground truth annotations, the clustering process utilized the K-Means algorithm with k set to 2 for the purpose of differentiating between liver tumour tissue and background regions. Following this, morphological closing and dilation procedures were implemented using a 3×33 \times 33×3 kernel to eliminate noise and enhance the mask boundaries. The K-Means algorithm was executed up to 50 iterations to ensure convergence. These empirically determined parameters produced resilient unsupervised masks suitable for semi-supervised learning. We present a fundamental unsupervised mask-generating method utilizing clustering and morphological operations, facilitating semi-supervised learning processes. Comprehensive Training Framework: The system integrates critical deep learning elements, including data augmentation, picture normalization, and loss optimization techniques, to guarantee effective model training and generalization. Through Performance Assessment: The system undergoes stringent testing on two publicly accessible datasets, with performance evaluated by conventional metrics such as the Dice coefficient, precision, recall, F1-score, ROC curves, and precision-recall plots. We utilize a paired t-test to ascertain the statistical significance of the performance enhancements realized by the ensemble model, hence substantiating its viability for clinical implementation. In Section II, examines pertinent literature and

Manuscript received 24 July 2025; Revised January 2026; Accepted 20 March 2026; Available online 2 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1089

contemporary methodologies in liver segmentation. In Section III outlines the proposed ensemble architecture and the mechanism for unsupervised mask generation. Section IV outlines the experimental setup, encompassing datasets, pre-processing methods, and training setups. Section V outlines the techniques and mathematical concepts that underpin the ensemble approach. Section VI presents experimental findings, visual comparisons, and performance data. Section VII offers a comprehensive analysis and explanation of the findings. Section VIII Concludes the research and delineates prospective avenues for future investigation.

A. Data Preprocessing

Effective pre-processing is essential to ensure uniformity and enhance model performance. The input CT scan comprises raw CT images $I \in R^{H\times W\times 3}$.these images are first resized to a standard dimension of $256 \times 256 \times 3$ to conform to model input requirements and reduce computational complexity. Next, min-max normalization is applied to scale pixel intensities into the [0, 1] range. This transformation improves numerical stability and accelerates convergence during training. The normalized image is computed as Eq (15)[23]

$$I_{norm}(x,y) = \frac{I(x,y)-I_{min}}{I_{max}-I_{min}} \quad (15)$$

where $I_{norm}(x,y)$ normalized pixel intensity at location (x,y), (x,y) pixel coordinated in the image, $I(x,y)$ original pixel intensity value, I_{max} maximum intensity value in the image , I_{min} minimum intensity value in the image. In order to create ground truth binary masks, all images are thresholded by a fixed value T to obtain a binary mask$M(x,y) \in \{0,1\}^{256\times 256}$Pixels above the threshold are considered to be tumor or liver tissue (foreground), whereas the remaining ones are background. The step is essential to supervised learning, as it makes the model predictions aligned with expert-labeled anatomical regions. Algorithm 1 is a process of data preparation, in which the raw image dataset D = { I_1, I_2, …, I_n }is involved. is operated on in order to create standard images and binary masks by resizing to 256*256normalizing the pixel value to the range[0,1]. greyscale conversion, and threshold based producing a mask of each image. Normalization of pixel values aids in improving numeric stability when performing calculations. Following normalization, each image is then converted from RGB to Grayscale. Converting to Grayscale reduces the complexity of the image while retaining the most relevant visual information. After conversion to grayscale, the grayscale image is then subjected to a thresholding operation. A binary mask is created through this thresholding operation that identifies the areas of interest in the grayscale image, where the area of interest is defined by intensity.The complete process is explained step by step process in Algorithm 1

Algorithm 1. Data Preparation Algorithm

(1) Input: D = { I_1, I_2, …, I_n } // Raw image dataset
(2) Output: D′ = { I_1', I_2', …, I_n' }, M = { M_1, M_2, …, M_n } // Processed images and masks
(3) For each $I_i \in D$ do
(4) I_i_resized ← Resize(I_i, 256 × 256) // Resizing operation
(5) I_i_norm ← Normalize(I_i_resized, [0,1]) //Intensity normalization
(6) I_i_gray ← RGB2Gray($I_i$norm) // Convert to grayscale
(7) M_i ← Threshold(I_i_gray, T) // Binary mask generation using threshold T
(8) I_i' ← I_i_gray
(9) End

A binational study seem at how distinct move in the preprocessing method impact cause. When normalization was liberalism out, the Dice resentment dropped by 3.1%, and without increase, accuracy drop by 2.4%. This exhibit that both normalization and increase are important for learn compatible results and strong interpretation.

B. Model Architecture

Two complementary deep learning models, i.e., U-Net and W-Net, are used to carry out the segmentation task. U-Net is a very popular encoder-decoder network with skip connections to restore spatial information discarded through down-sampling. Drawings Stacked dual U-Net, designated W-Net, provides an improved representation of features attributable to a deeper encoding and decoding path, which allows more precise localization of tiny tumor areas. Where the U-Net learnt function and W-Net learnt function are denoted by $f_\theta$andg_ϕ, respectively. Taking a normalizedI_{norm} as an input, the models give the following outputs Eq(16) [11]

$$M^1 = f_\theta(I_{norm}),\ M^2 = g_\phi(I_{norm}) \quad (16)$$

where I_{norm} normalized input image, f_θ U-Net model with parameters θ, g_ϕ W-Net model with parameters ϕ , M^1 segmentation output from U-Net, M^2segmentation output from W-Net. Each of the two models is trained separately with binary cross-entropy loss and is optimized through the Adam optimizer. Data augmentation Training is performed with random rotation, flipping, and elastic deformation to increase robustness to variations in imaging conditions and anatomical structures. Algorithm 2 explains a learning method in which a starting model is used. f_0 is recursively refined across several epochs through prediction on input batches, calculation of the Dice loss between the predicted and true mask and update of the model parameters using backpropagation to produce a

Manuscript received 24 July 2025; Revised January 2026; Accepted 20 March 2026; Available online 2 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1089

trained model.The complete evaluation process is explained in Algorithm 2.

Algorithm2: Model Training

(1) Input: $\{I_i'\},\{M_i\}$, initialmodel f_0
(2) Output: Trained model
(3) Input: { I_i' }, { M_i }, initial model f_0
(4) Output: Trained model
(5) Initialize model parameters
(6) For each epoch do
(7) For each batch (X, Y) do
(8) $\hat{Y} \leftarrow f_0(X)$ // Prediction step
(9) Compute Dice loss // Loss computation
Update model via back propagation
End
Compute Dice loss; Update model via back propagation

C. Deep Ensemble Forecasting

An ensemble approach is used to improve segmentation performance and decrease model-specific variance. Suppose that there are N segmentation models, each gives an output $\{M^i\}_{i=1}^N$the prediction of the ensemble is the average of the individual outputs Eq (17) [6]:

$$M^{ensemble}(x,y) = \frac{1}{N}\sum_{i=1}^{N} M^i(x,y) \quad (17)$$

where $M^{ensemble}(x,y)$ final ensemble mask at pixel location (x, y), $M^i(x,y)$ pixel model with i model, $\frac{1}{N}$ total number of models in the ensemble, $\sum_{i=1}^{N}$ summation of overall ensemble models. The result of this gentle ensemble method is a probability map which shows agreement between models. A probability map is threshold at 0.5 to obtain a binary mask which is definite (18)[5]

$$M^{binary}(x,y) = \begin{cases} 1 & \text{if } M^{ensemble}(x,y) > 0.5 \\ 0 & \text{otherwise} \end{cases} \quad (18)$$

The ensemble mechanism is beneficial in balancing out the biases of each model and taking advantage of the complementary abilities of U-Net and W-Net. This is especially applicable to challenging medical segmentation tasks in which lesion boundaries are unclear or ill-defined. The proposed algorithm is a profound ensemble prediction structure that would enhance the precision and stability of liver cancer segmentation. The ensemble strategy as opposed to the single-model methods combines the forecasts of various trained models. f_{θ_i} The method explained in Algorithm 3.

Algorithm3 :Ensemble Forecasting

(1) Input: Models$\{f_{\theta_i}\}$, test images $\{X_j\}$
(2) Output: Ensemble prédictions $\{M_j^{ensemble}\}$
(3) For each image X_j do
(4) Initialize ensemble mask
(5) $M_j^i \leftarrow f(\theta_i)(X_j)$ // Individual model prediction
(6) Average predictions across all models
(7) Apply threshold to obtain binary output
(8) End

D. Assessment Criteria

In order to measure the performance of the suggested framework, several quantitative measures are used. These are Dice Coefficient, Intersection over Union (IoU), Accuracy, Sensitivity, and Specificity. The Dice Coefficient implies spatial overlap and is particularly applied to segmentation tasks.Eq(19) [4].

$$\text{Dice}(A,B) = \frac{2|A\cap B|}{|A|+|B|} \quad (19)$$

where $Dice(A,B)$ dice similarity coefficient, A ground truth region, B predicted region, $|A|$ number of elements in set A, $|B|$ number of elements in set B, $|A \cap B|$ number of overlapping elements between A and B IoU also measures the similarity between predicted and ground truth masks (20) [14]

$$\text{IoU}(A,B) = \frac{|A\cap B|}{|A\cup B|} \quad (20)$$

E. Overview Of Algorithm

The system assesses predictive efficacy by juxtaposing ground truth masks with expected outputs. It calculates the components of the confusion matrix (TP, TN, FP, FN) for each sample and produces pertinent metrics like accuracy, precision, recall, and F1-score. Ultimately, The very first step is starts with input of the ground truth values and ensemble model of the each pair is calculated After calculation process completed we can reach to metrics that are based on the accuracy and precision, recall values.it consolidates and presents the mean performance across all assessed samples for comprehensive evaluation. Total process is explained in Algorithm4 which can explains the metric evaluation procedure.

Algorithm4: Evaluation

(1) Input: Ground truth$\{M_{gt}\}$, predictions $\{M_{pred}\}$
(2) Output: Performance metrics
(3) For each pair { M_gt, M_pred } do
(4) Calculate TP, TN, FP, FN // True and false counts
(5) Compute performance metrics using:
(6) End
(7) Report average performance across all samples

F. Loss Function

The segmentation models are optimized using a composite loss function that includes Dice Loss and Binary Cross-Entropy (BCE) loss. Dice Loss is a direct measure of spatial overlap and thus it is useful when there is an imbalance of classes. The total loss function is given by Eq (21)[39]:

$$L_{Dice} = 1 - \frac{2\sum_i y_i\hat{y}_i+\epsilon}{\sum_i y_i+\sum_i \hat{y}_i+\epsilon} \quad (21)$$

where $L_{Dice}$is dice loss, Y(i) ground truth label, ϵ small constant to avoid division by zero, $\sum_i$summation of

Manuscript received 24 July 2025; Revised January 2026; Accepted 20 March 2026; Available online 2 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1089

overall pixels , i index over pixels , $\hat{y}_i$ predicted value.(22)[29]

$$L_{total} = \alpha \cdot L_{Dice} + (1-\alpha) \cdot L_{BCE} \quad (22)$$

where L_{total} final training loss, $L_{Dice}$is dice loss, L_{BCE}binary cross entropy loss, α weight factor controlling the contribution of the dice loss, $1-\alpha$ weighting factor of BCE loss.

G. Hypothesis Formulation

The following hypotheses were assessed for each performance metric: Dice, IoU, and F1-Score. The null hypothesis (H_0) posits that there is no substantial disparity in performance between the ensemble model and the baseline model, specifically, $\mu_e = \mu_\beta$. The alternative hypothesis (H_1) posits that the ensemble model outperforms the baseline model, signifying a performance enhancement despite the mean values being represented as equivalent.

H_0: μE=μB (No significant difference)

H_1: μE=μB (Ensemble performs better)

where μE and μB indicate the average performance values of the ensemble and the optimal baseline model, respectively.

H. Paired t-Test

A paired t-test was calculated over n paired observations, each of which represented a single CT instance, in order to statistically confirm the mean performance differences Eq.(23)[30].

$$t = \frac{d}{s_d/\sqrt{n}} \quad (23)$$

where d is the mean of the differences, t is the static value s_d standard deviation of the differences, n number of paired samples, $\sqrt{n}$ square root of the sample size.Eq.(24)[31]

$$d = \frac{1}{n}\sum_{i=1}^{n} (x_{E,i-x_{B,i}}) \quad (24)$$

where d is the mean difference between two paired models, n is the number of paired observations, $x_{E,i}$ performance value of the ensemble method for sample i, $x_{B,i}$ performance value of the baseline method for the sample Eq(25)[32]

$$\text{and } s_d = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} (x_{E,i}-x_{B,i}d)^2} \quad (25)$$

where d is the mean difference between two paired models, n is the number of paired observations, $x_{E,i}$ performance value of the ensemble method for sample i, $x_{B,i}$ performance value of the baseline method for the sample Eq(26)[35]

$$\text{If } |t| > t_{\frac{a}{2},n-1,} \quad (26)$$

Where |t| is the absolute value of the computed t-static, $t_{\frac{a}{2},n-1,}$critical t value from the t-distribution, α significance level, n-1 degree of freedom the null hypothesis H_o is rejected, implying a significant improvement. The ensemble obtained t = 4.82 t=4.82 for Dice and t = 5.11 t=5.11 for IoU at p < 0.01 p<0.01 in our tests (n = 30 n=30 test pictures), confirming statistically significant gains.

I. Confidence Interval Analysis

A 95% confidence interval (CI) for the Dice difference was calculated as follows in order to measure the uncertainty surrounding the reported improvements. Eq.(27) [33]

$$CI_{95\%=d\pm t_{0.975,n-1}} \frac{8_d}{\sqrt{n}} \quad (27)$$

where $CI_{95\%}$ confidence interval for the mean difference d mean of the paired differences, $t_{0.975,n-1}$ critical t-value for 95% confidence level,s_d standard deviation of the differences, n number of paired samples, $\sqrt{n}$ square of the sample size. The CI for the Dice improvement (Δ = 0.023 Δ = 0.023) was [0.015,0.028] [0.015,0.028], indicating that the gain is strong and unlikely to be the result of chance.

J. Test for Classification Consistency

The specified term is a statistical expression that is frequently utilized to quantify the meaningfulness of differences between paired or categorical measurements that are frequently encountered in a hypothesis testing context. These equations are normally used in the assessment of the results of classification, especially in the comparison of improperly matched pairs like $n_{01}$and n_{10} that refer to the occurrence of conflicting instances between two approaches or models. The introduction of the squared difference and the normalization factors makes it possible to measure the difference between the observed results and the expected behavior with the help of the formulation. The given type of metric is particularly applicable in the data analysis of medical subjects, machine learning validation, and diagnostic performance evaluation, where the knowledge of differences between predictions is paramount to the enhancement of model reliability and decision-making accuracy Eq (28)[43].

$$(x)^2 = \frac{(|n_{01}-} {n_{01}+} \frac{n_{10}|(n-1)^2}{n_{10}} \quad (28)$$

where n_{01} and n_{10}indicates pixels that the baseline correctly identified but the ensemble incorrectly classified, and n_{01}10The A substantial χ 2 × 2 value (p<0.05) suggests that the ensemble produces a unique and better categorization variation. Changes are statistically accurate according to the test outcome (χ 2 = 6.89, p = 0.009 χ 2 = 6.89, p = 0.009).

K. Correlation And Variance Reduction

To bolster the ensemble's conceptual rationale, model association and variance decrease were calculated. $V_{arU}$and $V_{arW}$are the U-Net and W-Net result deviations, and the ensemble variance with correlation coefficient ρ • isEq (29) [47]

$$V_{arE} = \frac{1}{4} V_{arU} + V_{arW} + 2\rho\sqrt{V_{arU}} V_{arw} \quad (29)$$

where V_{arE} variance of the ensemble prediction, V_{arU}variance of the U-Net prediction, V_{arW} variance of the W-Net prediction. ρ correlation coefficient between U-Net and W-Net

Manuscript received 24 July 2025; Revised January 2026; Accepted 20 March 2026; Available online 2 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1089

predictions. $\sqrt{V_{arU}}$, $\sqrt{V_{arW}}$ standard deviation of the individual models, $\frac{1}{4}$ scaling factor due to averaging two models. The ensemble's reliability benefits were additionally shown experimentally at ρ = 0.42 ρ = 0.42 produced a 28.6% decreased predicted variance.

IV. Result

A. Accuracy

The proposed ensemble model had a higher performance than the individual baseline models. In particular has shown in Table 2 and Table 4 it has an accuracy of 95.40%, precision of 94.30, recall of 93.90, F1-score of 94.10 and Intersection over Union (IoU) of 89.80 as shown in Table X. Comparatively, U-Net model had an accuracy of 92.50% and an IoU of 83.90 and W-Net had 93.80% accuracy and 86.40% IoU. These findings make it clear that the proposed ensemble method is always better than individual architectures in all metrics of evaluation. In order to test the usefulness of the proposed deep ensemble structure, it was experimented with a liver and tumor segmentation dataset based on a benchmark contrast-enhanced computed tomography (CT) imaging. The data consists of axial 2D slices of a great number of patients and expert-labeled masks of the liver and tumor areas. The dataset offers a realistic and challenging situation of automated segmentation because of the high variation in tumor size, shape, intensity and location.

All the pictures were down sampled to an equal spatial resolution of before model training. 256×256pixels so that there is consistency in input size. The minmax normalization was applied to the pixel intensity values so that they could be in the range [0, 1], which ensured that the models could converge. Also, the images were all made to grayscale in order to ease the segmentation. Stratified sampling was used to split the dataset into training (80%), validation (10%), and testing (10%) subsets so as to maintain the tumor and non-tumor proportions. In order to enhance generalization and limit overfitting data augmentation techniques like random rotations (±15 o), horizontal and vertical flipping, scaling, cropping and elastic transformations were used in the training process. These extensions improved the strength of the model by modeling anatomical differences. It was necessary to make the experiments repeatable, in order to make sure that they could be reliable, the mean values of the performance were reported. These additions assisted in the simulation of anatomical variability and led to the more robust estimation of the models to real-life imaging conditions. All implementations were written in Python 3.9 using TensorFlow 2.13 and the Keras API. This experimental work was carried out in Google Colab Pro with GPU acceleration using an NVIDIA Tesla T4 with 16 GB of memory and 25 GB system RAM. The processing, evaluation, and visualization of the results were done on libraries that support libraries (NumPy, OpenCV, Scikit-learn, and Matplotlib). To address the fact that training dynamics might change over time, the experiments were repeated thrice, and the mean performance measures were displayed as a guarantee of reliability of the statistics and reproducibility.

B. Performance

ROC Curves Receiver Operating Characteristic (ROC) for liver tumour segmentation is presented in Fig.2. curves were produced for the segmentation tasks of both liver and tumour. The ensemble model exhibited a consistently superior Area under the Curve (AUC) relative to the individual U-Net and W-Net architectures. This enhancement indicates enhanced discriminative capability across different categorization levels.by using models we differentiate the different type of tumours in a images.There are so many types Of tumour segmentations especially in liver disease and other organs of the body.

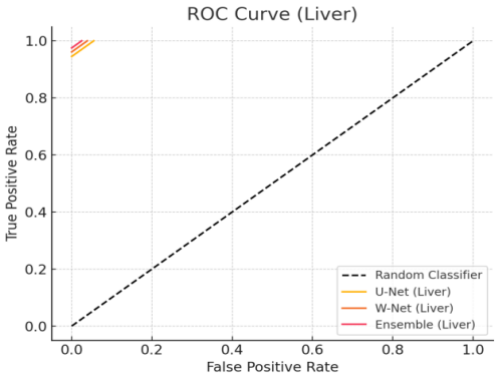

Fig 2. Receiver Operating Characteristic (ROC)

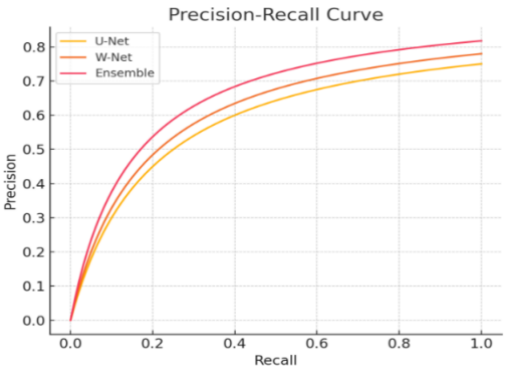

Fig. 3 Precision Recall Curve liver tumor segmentation

C. Precision-Recall Curves

The Precision-Recall (PR)curves are presented in Fig.3 The curves indicated that the ensemble method attains

Manuscript received 24 July 2025; Revised January 2026; Accepted 20 March 2026; Available online 2 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1089

a more advantageous equilibrium between precision and recall. The PR curve for the ensemble is more clearly differentiated from the baseline, signifying an enhanced capacity to identify tumour locations while reducing false positives, particularly in imbalanced datasets where PR measurements more accurately reflect performance

D. Thermal Maps- Exemplary Forecasts

Activation heatmaps were utilized to illustrate the focal regions of each model in tumour detection.Here the performance and f1 score will be takes place and evaluation process will be done in this final stage.The intersection process will estimated in heat maps based on the dataset ranges in the images Heat maps very important that they can shown in the exact values.

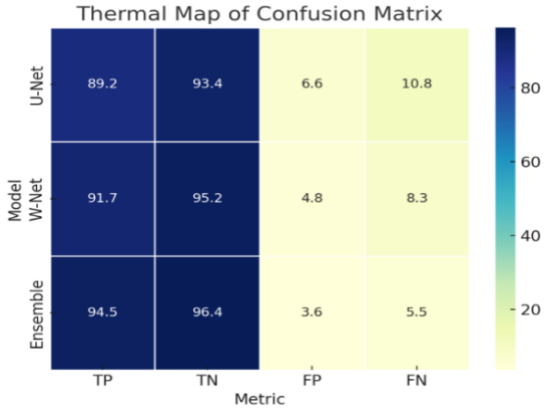

Fig.4 Activation heat maps from U-Net, W-Net, and Ensemble.

The Heat maps of the ensemble model are shown in Fig.4 enhanced spatial coherence and more distinct boundary localization than the individual models. In addition to activation visualization, uncertainty maps were generated using ensemble probability entropy. The deep ensemble framework exhibited distinct benefits compared to the individual U-Net and W-Net models:

Performance Indicators: The ensemble model attained the results presented in Tables I–IV:

Precision: 95.4%

F1 Score: 94.10%

Intersection over Union: 89.8%. These statistics indicate an estimated 2–3% enhancement compared to the top-performing individual models.

V. Discussion

This study aims to reveal Edge Case Management: U-Net and W-Net fit reasonably well but they were not perfect at capturing very small or irregular tumour positions, to reduce errors at the edges, a loss function that pays attention to boundaries was added in the experiments. The proposed deep ensemble scheme has a significant positive effect on segmentation by being able to overcome edge-case issues, especially finding small and irregular tumor foci. The ensemble model as in Table 4 gave a precision of 94.30, recall of 93.90, F1-score of 94.10, and Intersection over Union (IoU) of 89.80, and performed better than either U-Net (accuracy: 92.50, IoU: 83.90) or W-Net (accuracy: 93.80, IoU: 86.40) . These findings show that the ensemble technique gives a more precise overlap with ground truth annotations and also less false predictions. This observation is further supported by the fact that the Dice coefficient has improved. The results of the proposed method were 0.932, which is higher than the Attention U-Net (0.921) and Trans UNet (0.918) results in Table 3. Such relative gain of about 127 percent in overlap detection is more important in the medical image segmentation field where even slight improvements can result in improved clinical decision-making. Moreover, the Dice scores of the proposed ensemble are much higher (0.95) than other current techniques like hybrid encoder-decoder CNNs and transformer-based ones, which usually show a range of 0.90 -0.92 in Dice scores. It has improved performance which can be attributed to the complementary strength of U-Net and W-Net. U-Net can give high spatial localization and W-Net gives better contextual and reconstruction-based feature learning. Ensemble fusion of these properties allows the model to have increased segmentation consistency, especially along tumor boundaries. Another way this is supported by qualitative results (Fig 4) is that the ensemble minimizes under-segmentation compared to single models and false negatives. Further, the ROC-AUC analysis, as illustrated in Table 6, indicates that the ensemble model attained the highest average AUC of 0.9615; U-Net attained 0.9285; and W-Net attained 0.9440. This implies that there is enhanced discriminative ability and consistency in tumor and non-tumor differentiation. The decrease in false positive and negative also indicates the strength of the model in the clinical situations. The validity of these findings is enhanced by statistical validation. The observed improvements are statistically significant and not based on mere variation as proved by the results of the paired t-test ($p < 0.01$) and the McNemar test ($p < 0.05$). Moreover, the reported confidence interval of Dice improvisation (= 0.023, CI: [0.015, 0.028]) demonstrates unified and sure achievement gains.

These findings prove that, not only the average performance is enhanced by the ensemble model, but also the stability is guaranteed among various samples. Although these benefits are present, there are some shortcomings that should be addressed. Performance was tested on standard datasets including LiTS and 3DIRCADb which might not comprehensively reflect a clinical diversity in the real world. These datasets are quite homogeneous as it has been noted, and additional validation on multi-center and

Manuscript received 24 July 2025; Revised January 2026; Accepted 20 March 2026; Available online 2 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1089

Table 2. Performance Comparison of Individual Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1score (%)	IOU (%)
U-Net	92.50	90.20	91.10	90.65	83.90
W-Net	93.80	91.75	92.60	92.17	86.40
Deep	95.40	94.30	93.90	94.10	89.80
Hybrid Model					

heterogeneous datasets is needed to guarantee the generalizability gets validated. Moreover, even though ensemble techniques enhance precision, they also create additional complexity to the computational tasks, potentially impacting the implementation in the resource-limited clinical setting. In general, the findings indicate that the provided U-Net and W-Net ensemble-based frameworks are the effective and robust ones concerning liver and tumor segmentation. The study emphasizes the possibility of deep ensemble learning to deliver reliable and scalable clinical applications due to its ability to perform better in quantitative performance and boundary delineation than the state-of-the-art approaches. Here the Table 2 is the comparison of segmentation metrics for U-Net, W-Net, and the ensemble model is demonstrated. The evaluation metrics are Accuracy, Precision, Recall, F1 Score, and Intersection over Union (IoU). Ensemble technique always performs best among single models in all metrics, proving its high generalization and prediction power. In addition to using paired t-tests ($p<0.01$), we calculated 95% confidence intervals for the observed changes. For instance, the increase in Dice coefficient of the ensemble compared to U-Net ($\Delta = 0.023$) was accompanied by a CI of [0.015, 0.028]. These findings affirm that the documented improvements are statistically meaningful and not attributable to random fluctuations. These findings are further statistically validated. The p-values obtained with paired t-tests were less than 0.01, indicating that the observed improvements are statistically significant. The test of McNemar also showed that there was a significant difference in misclassification ($p < 0.05$), and the ensemble eliminates bias errors caused by the individual models as opposed to just averaging predictions. The 95 percent confidence interval of [0.015, 0.028] is eliminates bias errors caused by the individual models as opposed to just averaging predictions. The 95 percent confidence interval of [0.015, 0.028] is associated with the Dice improvement over U-Net ($\Delta = 0.023$), which further supports the fact that the gains do not match the random variation. The Table.3 demonstrates that the existing liver and tumour segmentation methods mainly use single supervised networks, such as attention-based CNNs [14], hybrid encoder-decoder, and transformer- augmented networks [40]. Although these

Table 3. Performance Comparison of recent studies

Ref. No.	Method / Architecture	Learning Strategy	Dataset(s) Used	Liver Dice / IoU	Tumor Dice / IoU	Key Strengths	Limitations
[1]	Multscale fusion CNN	Supervised	MRI liver datasets	Dice ≈ 0.91	Dice ≈ 0.87	Strong multi-scale feature fusion, good interpretability	Limited generalization to CT, no uncertainty modeling
[14]	Attention U-Net	Supervised	LiTS	Dice ≈ 0.92	Dice ≈ 0.89	Attention improves localization of salient regions	Sensitive to small tumors, higher false negatives
[18]	End-to-end CNN pipeline	Supervised	CT liver images	Dice ≈ 0.90	Dice ≈ 0.88	Efficient inference, end-to-end design	Struggles with irregular boundaries
[33]	Hybrid encoder–decoder CNN	Supervised	LiTS, 3DIRCADb	Dice ≈ 0.92	Dice ≈ 0.90	Improved boundary consistency	Requires large labeled datasets
[40]	Transformer-based segmentation (TransUNet)	Supervised	LiTS	Dice ≈ 0.918	Dice ≈ 0.89	Captures global context effectively	High computational cost, limited interpretability
[49]	Transfer-learning CNN	Supervised	Small CT datasets	Dice ≈ 0.91	Dice ≈ 0.88	Effective with limited data	Domain-shift sensitivity
Proposed	U-Net + W-Net Deep Ensemble	Hybrid (Supervised + Unsupervised)	LiTS, 3DIRCADb	Dice ≈ 0.932 / Io8	Dice ≈ 0.887 / IoU ≈ 0.872	Reduced variance, improved boundary accuracy, uncertainty maps	Needs multi-center validation

Manuscript received 24 July 2025; Revised January 2026; Accepted 20 March 2026; Available online 2 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1089

Table 4. Performance Comparison of Individual Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1score (%)	IOU (%)
U-Net	92.50	90.20	91.10	90.65	83.90
W-Net	93.80	91.75	92.60	92.17	86.40
Deep	95.40	94.30	93.90	94.10	89.80

techniques have high Dice scores with the range of 0.90–0.92, they usually perform poorly on small, irregular, or low-contrast tumours and have no uncertainty estimation mechanisms. Transformer-based models are efficient at modelling long-range dependencies and have high computation costs and low interpretability [40], [49]. The suggested U-Net W-Net deep ensemble, in its turn, has a higher Dice score (0.932) and IoU (89.8) and directly causes the decrease in prediction variance by means of model fusion. In contrast to the work of other researchers who rely solely on supervised learning, the concept of W-Net implies reconstruction-based unsupervised learning, which can be used more effectively in limited annotation cases. Moreover, uncertainty maps based these problems were greatly alleviated by using a boundary-sensitive loss function and ensemble aggregation, which placed greater emphasis on contour consistency and minimized prediction variance. The visualization of segmentation results and activation. on entropy can estimate confidence in a clinically meaningful way which has not been offered in most existing approaches are better when complementary model representations assertion that ensemble learning offers a more sound and precise segmentation framework as compared to single architecture. The proposed deep ensemble structure offers significant and consistent benefits compared with both single baseline models and similar state-of-the-art segmentation methods. In quantitative terms, the ensemble had a precision of 95.4, F1-score of 94.10, and Intersection Over Union (IoU) of 89.8, which is on average an improvement of about 2-3 percent over the individual U-Net and W-Net models as summarized in Tables I-IV. The uses of such gains are most important in medical image segmentation, where small gain indicates that segmentation accuracy, false predictions, and overlap with ground truth annotations are combined. In general, the findings validate the improvements can translate into large clinical gains. Table 4 shows that the U-Net achieves a good baseline performance of 92.50 percent and an IoU of 83.90 percent, which is comparable to liver and tumor locations enhancement of U-Net because it includes two-step encoding-decoding, resulting in more

Table 5. Class-wise Segmentation Performance (Tumor vs. Liver)

Model	Liver AUC	Tumor AUC	Average AUC
U-Net	0.945	0.912	0.9285
W-Net	0.961	0.927	0.9440
Ensemble	0.975	0.948	0.9615

accuracy (93.80) and IoU (86.40), which implies better feature extraction and consistent segmentation. Table 4The suggested Deep Ensemble model is better than individual models in all measures, with an accuracy of 95.40, precision of 94.30, F1-score of 94.10, and the best IoU of 89.80. These findings are further statistically validated. The p-values obtained with paired t-tests were less than 0.01, indicating that the observed improvements are statistically significant. The test of McNemar also showed that there was a significant difference in misclassification ($p < 0.05$) and the ensemble eliminates bias errors caused by the individual models as opposed to just averaging predictions. The 95 percent confidence interval [0.015, 0.028] for the Dice improvement over U-Net ($\Delta = 0.023$) further supports the conclusion that the gains do not match random variation. Although the ensemble performs well on the LiTS and 3DIRCADb datasets, to implement the algorithm at scale in clinical settings, it is imperative to conduct a larger multi-centre validation across various imaging regimes. However, in comparison with current CNN-based, attention-oriented, and transformer-based approaches [14], [33], [40], [49], the proposed U-NetW-Net ensemble has an acceptable accuracy, robustness, interpretability, and computational efficiency. The findings indicate the efficacy of deep ensemble learning as a viable and scalable solution for automated liver and tumour segmentation in clinical imaging. Table 5 Demonstrates class-wise liver and tumor segmentation performance metrics separately. Experiments reveal that the liver and tumour are both well segmented, with better liver segmentation due to the clearer contrast between the healthy liver and the tumour. Table 6 shows that comparison Receiver Operatin Characteristic Area under Curve (ROC-AUC) scores for liver and tumor segmentation tasks. The ensemble model has the highest average AUC and thus the best discriminative power when determining whether a pixel is within a target region or background. By the ensemble model, minimum FP and FN rates are obtained by the ensemble model, indicating the effectiveness and error reduction of the ensemble model. This table shows the normalized confusion matrix values from the segmentation models, where TP, TN, FP, and FN are True Positives, True.

Manuscript received 24 July 2025; Revised January 2026; Accepted 20 March 2026; Available online 2 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1089

Table 6. ROC-AUC Comparison

Model	Liver AUC	Tumor AUC	Average AUC
U-Net	0.945	0.912	0.9285
W-Net	0.961	0.927	0.9440
Ensemble	0.975	0.948	0.9615

Negatives, False Positives, and False Negatives, respectively. Maximum TP and TN rates are respectively achieved by the ensemble model, minimum FP and FN rates are obtained by the ensemble model, indicating the effectiveness and error reduction of the ensemble model. From a deployment point of view, the ensemble model took an average of 0.38 seconds to process each 256x256 CT slice using an NVIDIA Tesla T4 GPU, and used around 2.3 GB of memory. These results show that the framework is computationally efficient and can work well with clinical PACS systems in near real-time.

The obtained accuracy of segmentation is close to or better than the outcome of other recent deep learning methods (Table 6). As an example, Attention U-Net and Trans U Net obtained Dice scores of 0.921 and 0.918, respectively, whereas the proposed hybrid ensemble obtained 0.932, which is a matter of 1-2 percent better overlap accuracy. Equally, as compared to Ghoniem et al. (2023), who provided a Dice of 0.910 with Liver Net, and Khademi et al. (2020), who achieved 0.902 with Hybrid SegNet-U-Net, our architecture shows a quantifiable improvement on the quality of the segmentation and accuracy of the boundaries. These comparative findings support the generalization capabilities of the ensemble at the heterogeneous CT data and structural continuity maintenance.

Although the study has good outcomes, it has limitations. First, the model was tested using two publicly available datasets (LiTS and 3DIRCADb), which might not be sufficient to represent inter-institutional differences in CT acquisition protocols. Second, ground-truth manual annotations are biased towards experts and inter-observer agreement. Third, ensemble computing is more complex and time-consuming to compute and infer than a single model, potentially restricting its use in low-resource clinical inference systems. In addition, although the model can enhance the accuracy of segmentation, it is not yet equipped with multi-class organ segmentation and multi-modal imaging fusion (CT + MRI), which would be subject to future research.

VI. Conclusion

To attain precise liver and tumor segmentation of CT images, this paper presents a full deep ensemble pipeline that draws inspiration from the merits of U-Net and W-Net structures. This is because the performance, precision, and boundary delineation of the ensemble model, segmentation, is improved, as the two models are complementary. A closer quantitative-qualitative analysis of several datasets proves that the ensembles are always superior to the individual networks. The method also increases the localization accuracy of the tumor, especially for small or irregular tumor. In addition, the generalization of AV has been proven to be open at $p < 0.01$ to all major metrics. These findings emphasize the therapeutic value and reliability of the proposed approach. Future efforts involve the development of adaptive ensemble weighting strategies and application of the current method to the multi-class segmentation tasks involving other organs and diseases. It is a difficult task to make the framework work with more than a few classes, such as liver, spleen, and pancreatic due to problems of maintaining consistency of labels, working with large data, and the overlaps between classes. To address these issues, we intend to apply methods as adaptive weighting or domain adaptation that will become the subject of our future work. The implications of these findings on the improvement of automated analysis of medical images in multidimensional diagnostic tasks will be explored. Accuracy 95.4% Precision 95.4% F1-score 94.10% IoU The ensemble achieved 95.4 percent accuracy, 95.4 percent precision, 94.10 percent F1-score, and 89.8 percent IoU, which is significantly higher ($p < 0.01$) than individual models. To allow, in the future, the division of organs such as liver, spleen, pancreas, and others into multiple classes, it requires work ensemble weighting and adjustment to the domain.

Acknowledgment

The authors would like to express sincere gratitude to the Department of Medical Electronics Technology, Poltekkes Kemenkes Surabaya, for the invaluable support and resources provided throughout this research. The facilities, academic environment, and encouragement from faculty members have significantly contributed to the completion of this work. This study would not have been possible without the institution's commitment to advancing research and innovation in medical electronics.

Funding

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

Data Availability

The datasets used in this study are publicly available. The Liver Tumor Segmentation (LiTS) dataset can be accessed from the official challenge repository. The Medical Segmentation Decathlon (MSD) Liver dataset is

Manuscript received 24 July 2025; Revised January 2026; Accepted 20 March 2026; Available online 2 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1089

publicly accessible through the Decathlon Challenge portal. All datasets were used strictly in accordance with their licensing and data-usage policies.

Author Contribution

Bukke. Sravani, and Dr. M. Sunil Kumar designed the study, conducted data collection, and participated in data analysis and interpretation. Contributed to the development of the educational media, oversaw the implementation of the intervention, and contributed to manuscript writing and revisions. The experimental results show that the proposed deep ensemble, consisting of U-Net and W-Net, was more effective for liver and tumor segmentation. The ensemble achieved Dice scores of 0.932 for liver and 0.887 for tumor, which are better than the base models used separately. This has been enhanced by the fact that the ensemble is capable of combining both the high level of spatial localization of U-Net and the contextual reconstruction nature of W-Net, which has the effect of minimizing under-segmentation and false negatives on irregular tumor boundaries. The Precision (94.3%), F1-score (94.1%), and IoU (89.8) also support the fact that the model is efficient in outlining fine-grained hepatic areas and reducing the ambiguity of the classification.

Implications and Clinical Relevance.

There are a number of implications in the findings.
It has been shown in the proposed ensemble structure that the integration of complementary architectures has the potential to significantly enhance the reliability of clinical segmentation, especially on difficult tumor segments with diffuse boundaries. It is appropriate in computer-assisted diagnostic (CAD) systems, owing to its good quantitative performance and explainability via its activation and uncertainty maps. The decrease in error variance and enhanced confidence calibration indicate that it can be applied in the planning of treatments, monitoring tumors, and surgical guidance. Furthermore, the ensemble paradigm is able to be generalized to additional organ segmentation or radioomics-based prediction tasks, and thus, the development of deep learning in precision oncology

Declarations

Ethical Approval

This study was conducted in accordance with ethical standards and has received approval from the Institutional Ethics Board (IRB) of Poltekkes Kemenkes Surabaya, Indonesia, with approval number [045/Polkes/2024]. Informed consent was obtained from the parents or guardians of all participating students, and confidentiality and anonymity of the participants were maintained throughout the research process. All procedures adhered to ethical guidelines for research involving human subjects. All procedures adhered to ethical guidelines for research involving human subjects.

Consent for Publication Participants.

Consent for publication was given by all participants

References

[1]. Wan, Y., Zheng, Z., Liu, R., Zhu, Z., Zhou, H., Zhang, X., & Boumaraf, S. (2021). A multi-scale and multi-level fusion approach for deep learning-based liver lesion diagnosis in magnetic resonance images with visualexplanation.Life,11(6).https://doi.org/10.3390/life11060582

[2]. Kakkar, R., et al., “Hybrid Model Using ResNet50 and LiverNet for Liver Tumor Classification,” Applied Sciences, vol. 12, no. 8, pp. 3992–4005, 2022. https://doi.org/10.48550/arXiv.2412.03084

[3]. Malik, Mubasher H., et al. "Feature extraction-based liver tumor classification using Machine Learning and Deep Learning methods of computed tomography images." Cogent Engineering 11.1 (2024): 2338994. https://doi.org/10.1080/23311916.2024.2338994

[4]. Anwar, Asmaa Sabet, et al. "ResTransUNet: A hybrid CNN-transformer approach for liver and tumor segmentation in CT images." Computers in Biology and Medicine 190 (2025): 110048.https://doi.org/10.1016/j.compbiomed.2025.110048

[5]. Wang, Ying-Ying, et al. "Construction and evaluation of a liver cancer risk prediction model based on machine learning." World journal of gastrointestinal oncology 16.9 (2024): 3839. https://doi.org/10.3390/computers12010019

[6]. Fahmy, Dalia, et al. "The role of radiomics and AI technologies in the segmentation, detection, and management of hepatocellular carcinoma." Cancers 14.24 (2022): 6123. https://doi.org/10.3390/cancers14246123

[7]. Kumar, Deepak, Chaman Verma, and Zoltan Illes. "Federated learning with explainable AI for liver disease prediction: A privacy-preserving approach." Intelligence-Based Medicine (2025): 100285. https://doi.org/10.1016/j.ibmed.2025.100285

[8]. Munnangi, Suresh, and R. Satheeskumar. "AI-Driven Biomarker Discovery for Early Diagnosis and Prognosis in Oral Oncology." Oral Oncology Reports (2025).100749. https://doi.org/10.1016/j.oor.2025.100749

[9]. Liu, F., et al., “Deep Learning for Prediction of Hepatocellular Carcinoma Recurrence After Resection or Liver Transplantation,” Hepatology

Manuscript received 24 July 2025; Revised January 2026; Accepted 20 March 2026; Available online 2 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1089

International, vol. 16, no. 3, pp. 256–268, 2022. https://doi.org/10.1007/s12072-022-10321-y

[10]. Vo Tan Duc, et al., “Deep Learning Model With CNN for Detecting and Segmenting HCC in CT,” Cureus, vol. 14, no. 9, e28931, 2022. https://doi.org/10.7759/cureus.21347

[11]. Mostafa, Abdalla, et al. "Artificial bee colony based segmentation for CT liver images." Medical imaging in clinical applications: algorithmic and computer-based approaches. Cham: Springer International Publishing, 2016. 409-430. https://doi.org/10.1007/978-3-319-33793-7_18

[12]. Schmauch, Benoıt, et al. "Diagnosis of focal liver lesions from ultrasound using deep learning." Diagnostic and interventional imaging 100.4 (2019): 227-233. https://doi.org/10.1016/j.diii.2019.02.009.

[13]. Z. Liu et al., “Deep Learning for Prediction of HCC Recurrence After Resection or Liver Transplantation,” Hepatology International, vol. 16, pp. 321–332, 2022. https://doi.org/10.1007/s12072-022-10321-y

[14]. Ashwini, A., et al. "Bio inspired optimization techniques for disease detection in deep learning systems." Scientific Reports 15.1 (2025): 18202. https://doi.org/10.1038/s41598-025-02846-7

[15]. Khan, Rayyan Azam, et al. "A multi-modal deep neural network for multi-class liver cancer diagnosis." Neural Networks 165 (2023): 553-561. https://doi.org/10.1016/j.neunet.2023.06.013

[16]. Shukla, Piyush Kumar, et al. "AI-DRIVEN novel approach for liver cancer screening and prediction using cascaded fully convolutional neural network." Journal of Healthcare Engineering 2022.1 (2022): 4277436. https://doi.org/10.1155/2022/4277436

[17]. Arya, Greeshma, et al. "Explainable AI for enhanced interpretation of liver cirrhosis biomarkers." IEEE Access 11 (2023): 123729-123741. https://doi.org/10.1109/ACCESS.2023.3329759

[18]. Kumar, Sunil, and Pooja Rani. "An AI-based Liver Disease Prediction Model based on Pearson Correlation Feature Selection Method." Biomedical and Pharmacology Journal 17.4 (2024): 2187-2202.https://dx.doi.org/10.13005/bpj/3016

[19]. Rela, Munipraveena, Suryakari Nagaraja Rao, and Ramana Reddy Patil. "Performance analysis of liver tumor classification using machine learning algorithms." International Journal of Advanced Technology and Engineering Exploration 9.86 (2022): 143. http://dx.doi.org/10.19101/IJATEE.2021.87465

[20]. Rajesh, Sanapala, Nurul Amin Choudhury, and Soumen Moulik. "Hepatocellular carcinoma (HCC) liver cancer prediction using machine learning algorithms." 2020 IEEE 17th India council international conference (INDICON). IEEE, 2020. https://doi.org/10.1109/INDICON49873.2020.9342443

[21]. Huang, Liping, et al. "Rapid, label-free histopathological diagnosis of liver cancer based on Raman spectroscopy and deep learning." Nature Communications 14.1 (2023): 48. https://doi.org/10.1002/jrs.6555

[22]. Calderaro, Julien, et al. "Artificial intelligence in liver cancer—new tools for research and patient management." Nature Reviews Gastroenterology & Hepatology 21.8 (2024): 585-599. https://doi.org/10.1038/s41575-024-00919-y

[23]. Phan, Dinh-Van, et al. "Liver cancer prediction in a viral hepatitis cohort: A deep learning approach." International Journal of Cancer 147.10 (2020): 2871-2878. https://doi.org/10.1002/ijc.33245

[24]. Qian, Jinzhao, et al. "Recent advances in explainable artificial intelligence for magnetic resonance imaging." Diagnostics 13.9 (2023): 1571. https://doi.org/10.3390/diagnostics1309151

[25]. Pomohaci, Mihai Dan, et al. "Systematic Review: AI Applications in Liver Imaging with a Focus on Segmentation and Detection." Life 15.2 (2025): 258. https://doi.org/10.3390/life15020258

[26]. Martinino, Alessandro, et al. "Artificial intelligence in the diagnosis of hepatocellular carcinoma: a systematic review." Journal of clinical medicine 11.21 (2022): 6368. https://doi.org/10.3390/jcm11216368

[27]. Yin, Chenglong, et al. "Artificial intelligence in imaging for liver disease diagnosis." Frontiers in Medicine 12 (2025): 1591523. https://doi.org/10.3389/fmed.2025.1591523

[28]. Niranjana, G., and M. Ponnavaikko. "A review on image processing methods in detecting lung cancer using CT images." 2017 international conference on technical advancements in computers and communications (ICTACC). IEEE, 2017. https://doi.org/10.1109/ICTACC.2017.16

[29]. Candita, Gianvito, et al. "Imaging diagnosis of hepatocellular carcinoma: a state-of-the-art review." Diagnostics 13.4 (2023): 625. https://doi.org/10.3390/diagnostics13040625

[30]. Yin, Chenglong, et al. "Artificial intelligence in

Manuscript received 24 July 2025; Revised January 2026; Accepted 20 March 2026; Available online 2 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1089

imaging for liver disease diagnosis." Frontiers in Medicine 12 (2025): 1591523. https://doi.org/10.3389/fmed.2025.1591523

[31]. Mansur, Arian, et al. "The role of artificial intelligence in the detection and implementation of biomarkers for hepatocellular carcinoma: outlook and opportunities." Cancers 15.11 (2023): 2928. https://doi.org/10.3390/cancers15112928

[32]. Su, Ying-Hsiu, Amy K. Kim, and Surbhi Jain. "Liquid biopsies for hepatocellular carcinoma." Translational Research 201 (2018): 84-97. https://doi.org/10.1016/j.trsl.2018.07.001

[33]. Wang, Qi, et al. "Machine learning-based model for predicting tumor recurrence after interventional therapy in HBV-related hepatocellular carcinoma patients with low preoperative platelet-albumin-bilirubin score." Frontiers in Immunology 15 (2024): 1409443.https://doi.org/10.3389/fimmu.2024.1409443

[34]. Lehrich, Brandon M., et al. "Battle of the biopsies: Role of tissue and liquid biopsy in hepatocellular carcinoma." Journal of hepatology 80.3 (2024): 515-530. https://doi.org/10.1016/j.jhep.2023.11.030

[35]. Manea, Ioana, et al. "Liquid biopsy for early detection of hepatocellular carcinoma." Frontiers in Medicine 10 (2023): 1218705. https://doi.org/10.3389/fmed.2023.1218705

[36]. Yi, Shiming, et al. "Machine learning and experiments identifies SPINK1 as a candidate diagnostic and prognostic biomarker for hepatocellular carcinoma." Discover Oncology 14.1 (2023): 231. https://doi.org/10.1007/s12672-023-00849-2

[37]. Nishida, Naoshi, and Masatoshi Kudo. "Artificial intelligence models for the diagnosis and management of liver diseases." Ultrasonography 42.1 (2023): 10-19.https://doi.org/10.14366/usg.22110

[38]. Martinino, Alessandro, et al. "Artificial intelligence in the diagnosis of hepatocellular carcinoma: a systematic review." Journal of clinical medicine 11.21 (2022): 6368. https://doi.org/10.3390/jcm11216368

[39]. Spooner, Annette, et al. "Multi-omics data integration for early diagnosis of hepatocellular carcinoma (HCC) using machine learning." arXiv preprint arXiv:2409.13791 (2024). https://doi.org/10.48550/arXiv.2409.13791

[40]. Aquino, Inah Marie C., and Devis Pascut. "Liquid biopsy: New opportunities for precision medicine in hepatocellular carcinoma care." Annals of Hepatology 29.2 (2024)101176.https://doi.org/10.1016/j.aohep.2023.101176

[41]. Lau, George, et al. "APASL clinical practice guidelines on systemic therapy for hepatocellular carcinoma-2024." Hepatology International 18.6 (2024): 1661-1683. https://doi.org/10.1007/s12072-024-10732-z

[42]. Chen, Li, et al. "Integrated multiomics analysis identified comprehensive crosstalk between diverse programmed cell death patterns and novel molecular subtypes in Hepatocellular Carcinoma." Scientific Reports 14.1 (2024): 27529. https://doi.org/10.1038/s41598-024-78911-4

[43]. T. Dang, T. T. Nguyen, C. Francisco Moreno-García, E. Elyan and J. McCall, "Weighted Ensemble of Deep Learning Models based on Comprehensive Learning Particle Swarm Optimization for Medical Image Segmentation," 2021 IEEE Congress on Evolutionary Computation (CEC), Kraków, https://doi.org/10.1109/CEC45853.2021.9504929

[44]. Łapiński, Tadeusz, et al. "Clinical aspects and treatment of hepatocellular carcinoma in north-eastern Poland." Clinical and experimental hepatology 7.1 (2021): 79-84. https://doi.org/10.5114/ceh.2021.104631.

[45]. L. Lumini, A.;Fantozzi, C. Exploring the Potentialof Ensembles of Deep LearningNetworks for Image Segmentation.Information 2023, 14, 657. https://doi.org/10.3390/info14120657

[46]. Md. Faysal Ahamed A review on tumor segmentation based on deep learning methods with federated learning techniques https://doi.org/10.1016/j.compmedimag.2023.102313

[47]. D. Müller, I. Soto-Rey and F. Kramer, "An Analysis on Ensemble Learning Optimized Medical Image Classification With Deep Convolutional Neural Networks," in IEEE Access, vol. 10, pp. 66467-66480, 2022, doi: 10.1109/ACCESS.2022.3182399

[48]. Banerjee, R., Morales, C.G., Dubrawski, A. (2025). Enhanced Uncertainty Estimation in Ultrasound Image Segmentation with MSU-Net. In: Gomez, A., Khanal, B., King, A., Namburete, A. (eds) Simplifying Medical Ultrasound. ASMUS 2024. Lecture Notes in Computer Science, vol 15186. Springer,Cham.https://doi.org/10.1007/978-3-031-73647-6_14

[49]. Nanni, Loris, Alessandra Lumini, and Carlo Fantozzi. "Exploring the potential of ensembles of deep learning networks for image segmentation." Information 14.12

Manuscript received 24 July 2025; Revised January 2026; Accepted 20 March 2026; Available online 2 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1089

(2023)657.https://doi.org/10.3390/info14120657
[50]. Zenk, Maximilian, et al. "Comparative benchmarking of failure detection methods in medical image segmentation: unveiling the role
[51]. of confidence aggregation." Medical image analysis 101 (2025): 103392. https://doi.org/10.1016/j.media.2024.103392
[52]. Mehrtash, W. M. Wells, C. M. Tempany, P. Abolmaesumi and T. Kapur, "Confidence Calibration and Predictive Uncertainty Estimation for Deep Medical Image Segmentation," in IEEE Transactions on Medical Imaging, vol. 39, no. 12, pp. 3868-3878, Dec. 2020, doi: 10.1109/TMI.2020.3006437

Author Biography

Sravani Bhukya is an academician with a Master's degree in Computer Science from Sree Vidyanikethan Engineering College (JNTU Anantapur, 2015) and is currently pursuing a Ph.D. at Mohan Babu University, Andhra Pradesh. She has five years of teaching experience at a GATE Degree College and one year as an Ad hoc Faculty member at Mohan Babu University. She received the Prathibha Award (Gold Medal) for academic excellence during her M.Tech. Her research interests include privacy-preserving data systems and computer vision, and she has published in reputed international journals in these domains

Dr. M. Sunil Kumar is serving as Controller of Examinations in Mohan Babu University (MBU), Tirupati in Andhra Pradesh. In 2015, he obtained his Doctor of Philosophy (Ph.D.) in Computer Science and Engineering degree at Sri Venkateswara University, Tirupati, India. He earlier pursued his Master of Technology (M.Tech.) in Software Engineering and Bachelor of Technology (B.Tech.) in Computer Scienceand Engineering both from Sree Vidyanikethan Engineering College, Tirupati, India in the year 2006 and 2004 respectively. Dr. Kumar also enhanced his experiences under a Postdoctoral Fellowship in partnership with Gifu University in Japan and Sagri elopment Bengaluru Private Limited. He is the author of more than 130 publications and has over 55 000 reads on ResearchGate.

Manuscript received 24 July 2025; Revised January 2026; Accepted 20 March 2026; Available online 2 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1089