Journal of Computer Science

Research Article

Leveraging Modified Cascaded Fully Convolutional Neural Networks and Optimized Swin U-Net for Liver Lesion Prediction

Sugandha Chakraverti[1], Tejaswi Khanna[1] and Vijay Shukla[2]

[1]*Department of Computer Science and Engineering, Amity School of Engineering and Technology, Amity University, Greater Noida, India*
[2]*Department of Computer Science and Engineering, Greater Noida Institute of Technology, Greater Noida, India*

Article history
Received: 30-06-2026
Revised: 18-08-2026
Accepted: 22-09-2026

Corresponding Author:
Sugandha Chakraverti
Department of Computer Science and Engineering, Amity School of Engineering and Technology, Amity University, Greater Noida, India
Email: sugandha0708@gmail.com

Abstract: Liver lesions are abnormal growths that can develop for various reasons, some of which are noncancerous (benign) while others are cancerous. Machine learning and, deep learning techniques are employed for liver health assessment and lesion detection, but the quality and, diversity of training data can hinder the performance, leading to overfitting or underfitting. These models often need extensive pre-processing and may struggle with accurately segmenting complex patterns. This paper presents an integrated deep learning model for liver lesion prediction that integrates complementary image enhancement, segmentation, optimization and classification techniques. These enhanced images were then segmented using the Swin U-Net (SUN) approach, optimized with the Fossa Optimization Algorithm (FOA) to improve prediction accuracy. The segmented images were subsequently classified using a modified Cascaded Fully Convolutional Neural Network (CFCNN), which incorporated a Conditional Random Field (CRF) in the fully connected layer to ensure high connectivity and reduce computational complexity. Based on the segmented tumor masks, the severity of the disease categorized as low, medium, or high was predicted. The results demonstrate 96.45% accuracy, 96.85% precision, and 96.24% recall. The segmentation achieved an SSIM of 0.96% and a PSNR of 61.97%. Consequently, these methods are well-suited for real-time applications, providing timely and reliable assessments crucial for ensuring the quality of liver lesion detection.

Keywords: Liver Lesions, Transition Constant Normalization, Deep Image Prior Network, Swin U-Net, Cascaded Fully Convolutional Neural Network

Introduction

Abnormal areas of tissue in the liver are called liver lesions. These can be caused by a variety of things, including toxins, inflammation, benign or malignant neoplasms, and other diseases. The lesions may be cysts, abscesses, hemangiomas, or more serious conditions such as hepatocellular carcinoma and metastatic cancer (Razik et al., 2022). Recent advances in medical image analysis have shifted from conventional machine learning approaches toward transformer-based and hybrid deep learning architectures that jointly exploit local spatial features and global contextual information. In particular, Swin Transformer-based networks and hybrid CNN Transformer models have demonstrated promising performance for liver lesion segmentation and classification by improving feature representation and boundary delineation. Furthermore, recent surveys emphasize that future liver imaging systems should focus on integrated pipelines that combine image enhancement, segmentation, feature learning, and classification to improve robustness and clinical applicability (Sethia et al., 2025; Allenki and Soni, 2025; Pomohaci et al., 2025; Jiang et al., 2025). The symptoms vary from none at all to pain, jaundice or other symptoms of liver damage. Diagnosis is usually made by imaging technologies such as ultrasound, CT scan or MRI, and sometimes biopsy for definitive identification. Treatment depends on the underlying cause and can range from watchful waiting to medical treatment to surgery. The use of computers to help predict liver lesions can improve early diagnosis and the accuracy of analysis, leading to better treatment

SCIENCE Publications

© 2026 Sugandha Chakraverti, Tejaswi Khanna and Vijay Shukla. This open-access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.

outcomes. It reduces human errors, improves the consistency of diagnosis and enables personalized treatment through data integration.

AI technology has advanced substantially in the prediction of liver lesions, with machine learning methodologies such as random forests, support vector machines and, logistic regression being the key players. These techniques are able to analyze clinical and imaging data in order to identify patterns and classify lesions in an effective way. However, they often have the problem of high-dimensional data, manual selection of features and limited capacity to model complex relationships, which leads to sub-optimal accuracy in complex scenarios. In an effort to overcome these limitations, modern deep learning techniques such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have been used. These models are well adapted to automatic feature extraction, recognition of complex spatial-temporal patterns in medical imaging and processing of large data. However, the pros and cons of deep learning approaches are discussed. They need a lot of processing power and a lot of tagged data. They are prone to overfitting, particularly when dealing with small or skewed datasets (Peden et al., 2021, Ahn et al., 2021). AI based techniques improve the prediction of liver lesions but some issues like dependency on data, computing needs and generalizability of the models need to be addressed for successful clinical application (Wong et al., 2021). Traditional AI has limited capability to predict liver lesions and has resulted in the development of more sophisticated methods such as the Cascaded-Fully Convolutional Neural Network (CFCN). The Enhanced Swin U-Net adopts the Swin Transformer to capture long-range dependencies and hierarchical attributes to improve the accuracy of lesion detection. The Modified CFCN adopts cascaded structure to progressively enhance detection and segmentation so that improved results are obtained. These advanced techniques offer important benefits, such as increased accuracy in identification and segmentation of liver lesions due to improved feature extraction and hierarchical modeling. They model complex patterns well, improve results step by step, and understand long-term dependencies, leading to accurate and reliable predictions. Also, their flexibility with heterogeneous datasets and reduced need for human supervision make them beneficial for clinical applications. We developed a modified Cascaded-Fully Convolutional Neural Network (CFCNN) based on the Swin U-net to predict liver lesions in the imaging data (Gul et al., 2022; Zheng et al., 2022).

DIPN method is based on regularization to remove noise and enhance contrast to restore the original image, whereas TCN method is used to normalize the pixel values of all images within a uniform range. The Swin U-Net (SUN) significantly improves segmentation accuracy by effectively capturing multi-scale contextual information through its hierarchical transformer structure (Rather et al., 2024; Nkinahamira et al., 2024):

- FOA is used to find the optimal weight decay value for better segmentation performance further.
- The Conditional Random Field (CRF) is integrated in the fully connected layer of CFCNN to increase the accuracy while reducing the computational requirements and maintaining significant connectivity in the image (Zhang et al., 2024)

The novelty of this study does not lie in developing an entirely new individual learning algorithm, but in the systematic integration of complementary deep learning and optimization strategies into a unified liver lesion prediction framework.

Literature Review

Predicting liver lesions involves analysing several imaging methods, including Magnetic Resonance Imaging (MRI), Computed Tomography (CT) scans, and ultrasound, to identify abnormal growths or changes in liver tissue. The following section discusses recent research based on Artificial Intelligence (AI) techniques in liver lesion prediction.

Recent advances in liver lesion segmentation have shown the effectiveness of specialized medical image segmentation frameworks such as nnUNet, TransUNet, UNETR and other Transformer-based architectures. Recent transformer-based segmentation frameworks have further improved liver and liver tumor delineation by incorporating attention mechanisms, large-kernel convolutions, and hybrid encoder-decoder designs. These approaches demonstrate improved boundary localization and contextual feature extraction compared with conventional convolution-based architectures, particularly in challenging CT liver images. Sethia et al. (2025); Allenki and Soni (2025); Pomohaci et al. (2025); Jiang et al. (2025). These models utilize adaptive network configurations and strong feature representation mechanisms and have demonstrated strong benchmark performance on several public medical imaging datasets. Their success shows the necessity of benchmarking newly proposed methods with state-of-the-art liver-specific segmentation frameworks, to give a complete picture of performance.

Lin et al. (2020) proposed a random forest classifier to evaluate the adequacy of liver MRI images by combining deep Convolutional Neural Networks (CNNs) with Handcrafted (HC) features. The relative effects of these two factors were studied in relation to the size of the training sample. The HC features used for this application were texture analysis, Gaussian mixture models and Euler characteristic curves. The results revealed that the HC features outperformed CNNs especially in the case of

limited data, which is common in clinical research. Hence, a sensitivity of 95% ensures the correct identification of sub-optimal images. However, the MRI experts who performed the image acquisition were not able to assess the adequacy of the Hepatobiliary Phase (HBP) during or immediately after the examination.

Recently, hybrid CNN Transformer frameworks are proposed to combine the local feature extraction power of CNNs and the global information modeling ability of transformers. Hybrid networks perform better in liver CT image classification by enhancing feature discrimination as well as maintaining computational efficiency (Sethia et al., 2025; Allenki and Soni, 2025; Pomohaci et al., 2025; Jiang et al., 2025).

Xi et al. (2020) recommended a method for detecting liver abnormalities in medical photos based on deep learning. Transfer learning was used for efficient feature learning without overfitting because of the scarcity of training data. In order to learn visual features from local anomalies, deep patch CNNs were trained. The Class Activation Map (CAM)-based detector calculated an abnormality map for a complete medical image after it had been trained. In order to construct abnormality detectors, they were combined with Region Proposal Networks (RPNs) and CAMs. According to experimental data, the integrated strategy performed better than other deep learning techniques as well as traditional methods. Researcher claimed 92.53% overall accuracy in this research. compared to 81.55% by a classical approach using manual features. However, information that was essential for detecting abnormalities, like in mammography, was lost as a result of image reduction.

Meng et al. (2020) introduced a deep learning-inspired three-Dimensional (3D) CNN architecture that might be utilized to separate the liver images and liver tumors from three-Dimensional (3D) abdomen CT scans. The 3D CNN was able to rectify unbalance segmentation performance and computational requirements because of its unique design for 3D medical picture data. But they found that some voxels were wrongly segmented because of large Hausdorff distances in liver and liver tumor segmentation. The experimental results show Dice coefficients of 0.965 for liver segmentation and 0.689 for liver tumor segmentation.

DefED-Net, a deformable encoder-decoder network for liver and hepatic tumor segmentation is proposed by Almotairi et al. (2020). In DefED-Net, deformable convolution was adopted to learn convolution kernels with embedded adaptive spatial structuring information to improve the ability of feature representation. We also designed a Ladder-atrous-spatial-pyramid pooling (Ladder-ASPP) module by adopting multiscale dilation rates to offer more abundant contextual information than atrous spatial pyramid pooling for CT image segmentation. Thus, on a common dataset for liver CT scans it achieved up to 99.9% tumor accuracy during training. However, there were frequently many tiny, isolated patches or holes in the segmented image. Therefore, without making the models larger, DefED-Net produced the best liver segmentation results.

Sakthisaravanan and Meenakshi (2020) developed a method for detecting liver tumors called Outline Preservation Based Segmentation and Search Based Hybrid Classification (OPBS-SSHC). This work's contribution was found in each and every step of the image processing methodology. The application of the new OPBS-SSHC technique to separate and categorize liver cancers from other adjacent regions in the human body's abdominal region. In contrast to the other approaches that are currently in use, which had values of 72.9, 93.1, 94.3, and 94.5%, respectively, the OPBS-SSHC method attained a maximum Dice value of 96.41%. It did not, however, address certain aspects of the human liver's anatomy.

Overall, the recent literature demonstrates a clear transition from conventional convolutional neural networks toward transformer-based and hybrid deep learning architectures for liver image analysis. Nevertheless, most existing studies continue to optimize individual components, such as image enhancement, segmentation, or classification, independently. Integrated diagnostic frameworks that combine image enhancement, segmentation, feature learning, adaptive optimization, and prediction refinement remain comparatively limited, thereby motivating the proposed methodology presented in this work.

Despite remarkable progress in transformer-based medical image analysis, most recent studies continue to address image enhancement, segmentation, and classification as separate tasks. Comparatively fewer investigations have explored unified diagnostic frameworks capable of integrating these complementary processes while simultaneously incorporating adaptive hyperparameter optimization and prediction refinement. This research gap motivates the integrated framework proposed in this study (Sethia et al., 2025; Allenki and Soni, 2025; Pomohaci et al., 2025; Jiang et al., 2025; Baguer et al., 2020).

Research Gap

In the field of medical imaging (Almotairi et al., 2020), obtaining datasets is costly. The research studies mentioned above have various important problems that hinder their efficiency such as poor performance on high-dimensional data and dependence on large datasets and resources (Meng et al., 2020). Moreover, the quality of segmentation is often affected by the missing of important details in traditional methods. In addition, they can suffer from computational instability during training, affecting their overall performance and convergence (Sakthisaravanan and Meenakshi, 2020). A method cannot capture complex patterns and relationships, and

DOI: 10.3844/jcssp.2026.3250.3267

they rely on large labeled datasets which is a challenge (Huang et al., 2023). The proposed modified Cascaded-Fully Convolutional Neural Network (CFCNN) with optimized Swin U-Net can effectively solve these challenges and achieve the ability of effectively handling complex patterns, preserving fine-grained details, scalability, and resource efficiency.

Proposed Methodology

Liver lesions are abnormal areas of tissue in the liver. Liver lesions are typically found through imaging tests such as ultrasound, CT scans, or MRI. The recognition of these lesions allows for a quick diagnosis and oftentimes successful intervention, particularly in the setting of Hepatocellular Carcinoma (HCC). However, liver lesions can be difficult to diagnose, and some lesions cannot be classified without using invasive procedures, which can cause anxiety and complications such as infection or hemorrhage to the patient undergoing biopsy. A modified CFCNN prior network based on Swin U-net is designed for liver lesion segmentation and classification. The proposed methodology is demonstrated in Figure 1 for Liver Lesion framework.

The images of the liver lesion were taken from Kaggle platform and then applied to different pre-processing techniques such as Regularization based Deep Image Prior Network (DIPN) and Transition Constant Normalization (TCN). These first steps were important to improve the quality and consistency of the images to make them suitable for further analysis. The images were then used as inputs for segmentation, which was performed using the Swin U-Net (SUN) method after pre-processing. The optimization of the weight decay parameter for the improvement of segmentation efficiency was performed using the Fossa Optimization Algorithm (FOA). This optimization method was used for optimizing the SUN parameter to have better accuracy and reliability of the results. The partitioned images were then passed through a modified Cascaded-Fully Convolutional Neural Network (CFCNN) for classification. The CFCNN was designed to detect the presence of liver lesions in the images. Then, the Conditional Random Field (CRF) was added to the fully connected layer of the CFCNN to improve the classification results. This integration allowed to maintain strong connectivity in the images while reducing computational complexity at the same time. The system was capable of better detecting liver lesions through the use of these advanced techniques that guaranteed accuracy and efficacy in the classification process. The difference between the original image and the tumor segment can be analyzed to quantify the actual area of the tumor and predict the severity of liver lesion that is categorized as low, medium, and high.

Pre-Processing

In proposed liver lesion prediction, pre-processing step initially took place. It contains Regularization-based Deep Image Prior Network (DIPN) and Transition Constant Normalization (TCN) approaches, which are explained in the section below.

Enhanced Deep Image Prior Network Using Regularization

The first step of the pre-processing uses the Regularization-based Deep Image Prior Network (DIPN) to overcome image noise while maintaining the anatomical structures necessary for accurate liver lesion segmentation. In this work, the DIPN is implemented as a lightweight encoder-decoder convolutional neural network in the TensorFlow/Keras framework. The network takes as input the RGB CT images of size 256×256×3.

The encoder has two convolutional layers with 64 and 128 feature maps, respectively. The convolution with 3× 3 kernel, same padding, and Rectified Linear Unit (ReLU) activation function extracts hierarchical features. The decoder is a convolutional layer with 64 feature maps and a final convolutional reconstruction layer with 3 output channels for the reconstructed RGB image.

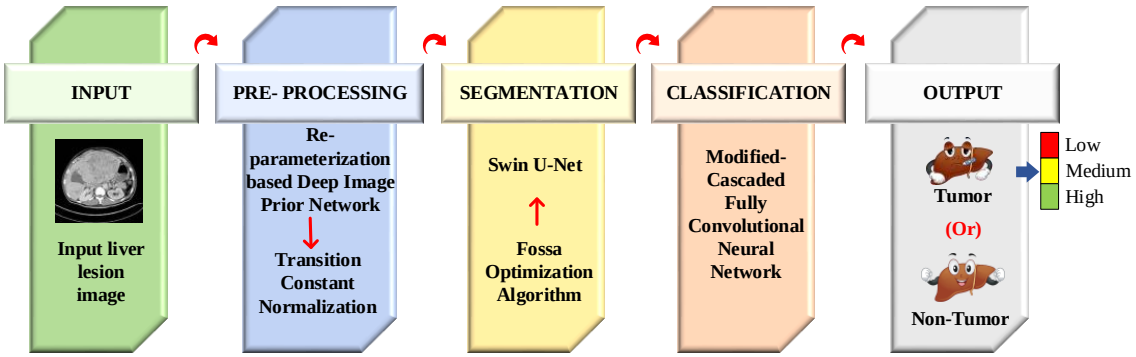

Fig. 1: Architecture of proposed modified CFCNN with swin U-net model

The reconstruction layer uses 3×3 kernel with same padding so as to keep the spatial dimensions of the input image.

The training is performed by adding synthetic Gaussian noise to the input CT image to create noisy and clean CT picture pairings. The DIPN is trained to map the noisy images to the corresponding clean images by reducing the reconstruction loss i.e. the Mean Square Error (MSE). Adam optimizer with a learning rate of 0.001 is used to fine-tune the parameters of the network. The model is trained for 10 epochs with a batch-size of 1. After training, the denoised images generated by DIPN are further processed using the suggested Transition Constant Normalization (TCN) module before feeding the Optimized Swin U-Net for liver lesion segmentation.

Transition Constant Normalization (TCN)

Transition Constant Normalization (TCN) is a technique used in neural networks and deep learning frameworks to normalize pixel values to a defined range, thus improving and stabilizing the training process. Moreover, TCN divides the input tensor into multiple smaller chunks and computes the mean and standard deviation of each chunk and then normalizes the data using a normalization procedure. This will ensure that the resultant data will have a zero mean and unit standard deviation. This makes sure a stable and reliable training. The effectiveness of TCN is verified by evaluating its stability over different batches and datasets, and its application in image enhancement problems such as denoising, deblurring and super-resolution (Sun et al., 2022).

Operation description TCNs are multi-stage and involve a number of mathematical transformations. Let's break down the provided equations and explain each component.

Operation Description

There are multiple phases and mathematical changes involved in TCN's operation. Let's dissect the provided formulas and clarify each section:

Slicing the Tensor

$$F_s^{ab} = F[:,:,a::,b::],\ a,b\epsilon\{0,1\} \tag{1}$$

Where, F is likely a multi-dimensional array, and the notation $F[:,:,a::,b::]$ specifies how to extract a sub-array from F. The colons (:) indicate that all elements along the first two dimensions are selected, while the notation $a::,b::$ suggests that the sub-array is taken starting from index a and b respectively.

Mean and Standard Deviation Calculation

$$\mu_2 = \frac{1}{HW}\sum_{i\in[1,H],j\in[1,W]} F_{ij,}\ \sigma_2 = \sqrt{\frac{1}{HW}\sum_{i\in[1,H],j\in[1,W]}(F_{ij,} - \mu_2)^2} \tag{2}$$

These formulas provide the tensor F's mean (μ_2) and standard deviation (σ_2) over its spatial dimensions. This common normalization procedure adjusts the data according to its variance and centers it around zero.

Normalization Transformation

$$\hat{F}_2 = F_2 - \hat{F}_1 = F_2 - \frac{F_1-\mu_2}{\sqrt{\sigma_2^2+\epsilon}} \tag{3}$$

Where, $\hat{F}_2$ is obtained by subtracting a normalized version of F_1 from F_2 . The term represents the normalization of F_1 using the mean μ_2 and standard deviation σ_2^2 of F_2, with ϵ added to the denominator to ensure numerical stability. This operation adjusts F_2 by removing the normalized influence of F_1.

Verify the Transition Constant Ability

To confirm the transition-constant capability, must make sure that the transformation doesn't change the network's basic characteristics. Examining the transformation's Jacobian matrix accomplishes this:

$$J_g = \begin{bmatrix} \frac{\partial \hat{F}_1}{\partial F_1} & \frac{\partial \hat{F}_1}{\partial F_2} \\ \frac{\partial \hat{F}_2}{\partial F_1} & \frac{\partial \hat{F}_2}{\partial F_2} \end{bmatrix} = \begin{bmatrix} \frac{1}{\partial s(F_2)} & 0 \\ -1 & 1 \end{bmatrix} \tag{4}$$

The determinant of the Jacobian matrix J_g is given by:

$$Det\ (J_g) = \frac{1}{\partial s(F_2)} \neq 0 \tag{5}$$

The transformation is invertible, which means that the original data may be inferred from the modified data, according to this non-zero determinant. This characteristic is essential for preserving the network's transition constants.

Verify the Normalization Ability

Making sure the transformation successfully normalizes the data is necessary to confirm the normalization ability. To accomplish this, the normalized tensor $\hat{F}_1$ and the original tensor F_1 are compared:

$$\hat{F}_1 = \frac{F_1-\mu_2}{\sqrt{\sigma_2^2+\epsilon}} \approx \frac{F_1-\mu_1}{\sqrt{\sigma_1^2+\epsilon}} \tag{6}$$

The normalized tensor $\hat{F}_1$ is near to the original tensor F_1 upon normalization, according to this approximation. The approximation shows that the normalizing procedure successfully centers and scales the data, with the terms σ_1 and μ_1 standing for the standard deviation and mean of the original tensor, respectively.

Segmentation

The pre-processed data is then fed into the segmentation stage, which utilizes Swin U-Net (SUN) with Fossa Optimization Algorithm (FOA).

Swin U-Net (SUN)

The Swin U-Net (SUN) incorporates several key components, including 3D-Organoid-SwinNet, Swin-Transformer, a Loss Function, and Post processing steps (Wang et al., 2021).

In this paper, we propose a novel deep learning architecture for medical image segmentation. The Swin U-Net exploits the merits of both U-Net architecture and the Swin Transformer, which can efficiently capture hierarchical features via shifted windows. The combination of these two methods enables SUN to efficiently capture contextual information locally and globally, which improves the accuracy of segmentation (Fig. 2). The architecture is based on an encoder-decoder pathway with skip connections to allow information transfer at different scales. Such a design makes the SUN particularly effective for complex segmentation tasks such as liver tumor detection where precise boundary delineation is critical.

3D Organoid Swin Net

We proposed a new deep learning framework, called 3D-Organoid-SwinNet, for the analysis of 3D organoids (miniature, simplified versions of organs grown in vitro). The design employs the Swin Transformer, a variant of the transformer adapted for visual tasks, to effectively capture and analyze the complex 3D structures of organoids (Sohaib et al., 2025).

B. Swin Transformer

Swin Transformer (Shifted Window Transformer) A variant of transformer architecture, primarily targeted at visual computing applications. The method presents a hierarchical approach, which splits the input image into individual non-overlapping local windows and performs self-attention mechanisms on those patches. The core notion of Swin Transformer is the shifted windows, which enable windows to link with each other and the model to learn global information. Swin Transformer's hierarchical design is well-suited for applications including object detection, semantic segmentation, and picture classification where both local and global contexts are important.

The following formulas can be used to describe the Swin Transformer:

Local Self-Attention within a Window:

$$Attentication\,(Q,K,V) = Softmax(\frac{QK^T}{\sqrt{d_k}})V \quad (7)$$

Where $Q, K\ and\ V$ are the query, key, and value matrices, respectively, and d_k is the dimensionality of the key vectors.

Loss Function

Loss function is a core component of any machine learning model that quantifies the difference between the expected and actual output. A feasible loss function for the 3D-Organoid-SwinNet would be a combination of multiple terms to consider different features of the prediction such as segmentation accuracy and boundary preservation. The Dice Loss is a common loss function for segmentation problems:

The Dice Loss, a popular loss function for segmentation problems, is as follows:

$$L(Y,P) = \left(1 - \sum_{i=1}^{I} \frac{2\times\sum_{v=1}^{V} Y_{v,i.} p_{v,i}}{\sum_{v=1}^{V} Y_{v,i.}^2 + \sum_{v=1}^{V} P_{v,i.}^2}\right) + \sum_{v=1}^{V} Y_{v,i.}^2 logP_{v,i} \quad (8)$$

Two primary components are combined in the provided loss function $L(Y,P)$ to balance prediction confidence and segmentation accuracy. The overlap between the ground truth Y and the anticipated probabilities P is measured by the first component, a Dice-like term. Its definition is $\sum_{i=1}^{I} \frac{2\times\sum_{v=1}^{V} Y_{v,i.} p_{v,i}}{\sum_{v=1}^{V} Y_{v,i.}^2 + \sum_{v=1}^{V} P_{v,i.}^2}$ and its goal is to maximize the degree of similarity between the ground truth and the expected segmentation. The second element is a cross-entropy term, which is defined as $\sum_{v=1}^{V} Y_{v,i.}^2 logP_{v,i}$. and quantifies the prediction confidence.

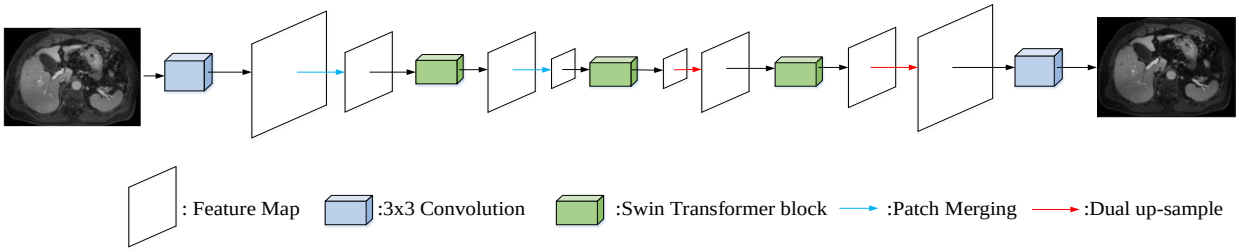

Fig. 2: Architecture of Swin U-Net (SUN).

This part penalizes low probabilities of the true labels, hence encourages the model to make confident predictions. Reliable segmentation results require precise segmentations and confident predictions, which are achieved using the combined loss function L (Y, P) = (1 - DiceTerm) + CrossEntropyTerm. The second term is the binary cross-entropy loss. The first term is the dice loss. I is the number of classes. Y and P are the true values and the output probabilities of the model respectively. The optimal weight decay parameter plays a critical role in improving the efficiency of the SUN method. It controls the convergence speed and stability of the training process. Dropout is important to prevent overfitting by offering regularization and enhancing the generalization ability of the model. The weight decay was carefully selected in the proposed study by using FOA optimization, which is explained below.

In proposed work, the weight decay was optimally chosen using FOA optimization that is discussed as below.

Fossa Optimization Algorithm (FOA)

The FOA, a novel bio-inspired metaheuristic algorithm, emulates the fossa's natural hunting behavior.

In this work, the weight decay parameter of the Swin U-Net model is adaptively determined by the Fossa Optimization Algorithm (FOA) as a metaheuristic optimization approach. The choice of FOA is based on its balancing exploration exploitation strategy reported in the literature. The work does not claim theoretical precedence to traditional hyperparameter optimization methods such as Grid Search or Bayesian Optimization. Instead, in the proposed framework, FOA is used as an efficient optimization methodology. The formal convergence proof and computational complexity analysis of FOA are beyond the scope of the present study, which are an interesting direction of future work.

The fossa is a native Malagasy predator. The initiative aims at solving optimization problems by mimicking the two-stage hunting strategy of the Fossa, which includes a first strike on a spotted lemur followed by a chase through the trees. The FOA was based on the fossa's characteristic hunting technique. The fossa is famous for its cunning hunting and its agility. From a distance it sees its prey and begins a tentative attack. The fossa chases its prey through the trees, moving fluidly to trap it if it gets away. In FOA, this behavior is mathematically modelled to generate an optimization algorithm which can effectively balance exploration and exploitation.

Initialisation

In FOA initialization process involves randomly generating a population of candidate solutions, known as fossas, within the defined search space to explore potential optimal solutions for the given problem. In proposed work, initialization the key hyperparameters including weight decay (W) in Equation, (9):

$$(W) = [\, W_1, W_2 \ldots W_n \,] \tag{9}$$

In the proposed framework, the weight decay parameter was optimized in an empirically selected search interval during the hyperparameter optimization process based on FOA. The search interval selected is specific to the experimental setup of this study and should not be considered a generally recommended range for weight decay in deep learning models. In general, the best weight decay depends on the model architecture, optimization strategy, and dataset.

In the proposed framework, the weight decay parameter was optimized in an empirically selected search interval during the hyperparameter optimization process based on FOA. The search interval selected is specific to the experimental setup of this study and should not be considered a generally recommended range for weight decay in deep learning models. In general, the best weight decay depends on the model architecture, optimization strategy, and dataset.

These ranges ensure effective exploration and stability during the optimization process.

Fitness Function

Minimizing of loss is the fitness function's ultimate goal for optimum parameter selection. The fitness is calculated using the following Equations (10) and (11):

$$Fitness = Minimizing\ (Loss) \tag{10}$$

$$Loss = Original\ loss + \lambda \sum_{i=1}^{n} w_i^2 \tag{11}$$

Where λ is the regularization parameter, and w_i are the weights of the model.

Updating

Inspired by the natural locomotion of fossas, the FOA algorithm is designed based on the locomotion techniques of fossas in the wild. Two different phases refresh locations of FOA members in the context of problem solving. The trade-off between exploration and exploitation is largely determined by the adaptation parameter F_i, where F_i<5 leads to exploration of the search domain and F_i>5 leads to exploitation of promising regions already found.

Phase 1: Exploration Phase

The FOA makes the assumption that the fossa attacks one of these potential lemurs in its environment after choosing it at random. Eq. (12) is used to determine Based on the fossa's positioning movement during the attack on

the identified lemur, each member of the FOA population is assigned a new random position. According to Eq. (13), the prior position of the corresponding population member is replaced if the objective function value is higher in this new role:

$$x_{i,j}^{p1} = x_{i,j} + r_{i,j} \cdot (SL_{i,j} - I_{i,j} \cdot x_{i,j}) \tag{12}$$

$$X_i = \begin{cases} X_i^{p1}, F_i^{p1} \leq F_i \\ X_i, else \end{cases} \tag{13}$$

In this case, the lemur selected by the *i*th fossa is indicated bySL_i, and the *j*th dimension of the location of this lemur is indicated by$SL_{i,j}$. The newly calculated location of the *i*th fossa during the FOA's assault phase is represented byX_i^{p1}, where $x_{i,j}^{p1}$ is its *j*th dimension. At this new point, the objective function value isF_i^{p1}. $r_{i,j}$ are random numbers, either 1 or 2, and $r_{i,j}$ are random n values in the interval [0, 1].

Phase 2: Exploitation Phase

At this point the fossa perfects its technique to focus on the prey, imitating the lemur's pursuit through the branches. In the exploitation phase, the algorithm improves the solutions by intensifying the search in areas deemed interesting. During the chase the fossa actively changes its position. These positional modifications during use are based on the active changes of the fossa:

$$x_{i,j}^{p2} = x_{i,j} + \left(1 - 2\, r_{i,j}\right).\frac{ub_j - lb_j}{t} \tag{14}$$

$$X_i = \begin{cases} X_i^{p2}, F_i^{p2} \leq F_i \\ X_i, else \end{cases} \tag{15}$$

In this instance, the projected FOA's revised position determined for the ith f fossa during the chase phase is represented by $x_{i,j}^{p2}$. The yth dimension of this new position is shown by each$x_{i,j}^{p2}$, and the accompanying objective function value is indicated by F_i^{p2}. The current iteration count is represented by t, while the variables $r_{i,j}$ are random values between [0, 1].

Termination

The procedure ends when the best option is identified, else the process is repeated until find the best solution.

Classification

The classification is achieved by using a modified cascaded fully convolutional neural network which enhances the classification process's precision and effectiveness.

The Fossa Optimization Algorithm (FOA) was chosen because it offers a flexible trade-off between global exploration and local exploitation, making it appropriate for optimizing the weight decay parameter of the proposed Swin U-Net. Although other population-based optimization algorithms like GA, PSO, GWO, and HHO have shown comparable performance on similar optimization tasks, a complete comparative analysis of these optimizers was not the goal of this research and will be explored in subsequent studies.

Modified Cascaded Fully Convolutional Neural Network (MCFCNN)

The proposed Modified Cascaded Fully Convolutional Neural Network (MCFCNN) is composed of three convolutional feature extraction blocks, and then classification and refinement stages. The first, second and third convolutional layers use 32, 64 and 128 filters, respectively, all with a 3×3 convolution kernel, ReLU activation and same padding. After each convolutional layer, a 2×2 max-pooling layer is added to extract the hierarchical features from the image at different scales, meanwhile reducing the spatial resolution.

The feature maps obtained after feature extraction are processed by a Global Average Pooling (GAP) layer to obtain compact feature representations. Then, these features are passed to two fully connected layers with 128 and 64 neurons respectively for high-level feature learning. A softmax output layer produces the probability distribution of the target classes.

To further improve the spatial consistency of the predicted labels, we add a trainable Conditional Random Field (CRF) layer after the Softmax layer. The CRF refines the predicted class probabilities by learning pairwise relationships among neighbouring labels, which reduces isolated misclassifications and improves the coherence of the final predictions.

The MCFCNN was trained with Adam optimizer with learning rate of 0.001 and the optimization objective was categorical crossentropy.

CNNs are specialized neural network variants that have gained widespread adoption in a variety of industries because of their exceptional feature extraction capabilities. Convolutional, pooling, and fully connected layers make up the majority of a CNN (Rahman et al., 2022).

The architecture described in Fig. 3 is a Modified Cascaded-Fully Convolutional Neural Network for multi-scale feature extraction and classification. The input image is passed through several dilated convolution layers with dilation rate 1 to extract features. They extract spatial information without losing the resolution of the input. The features are concatenated and then passed through a BN and a ReLU activation layer and then reshaped to (128 times 64) by max-pooling. The architecture then goes into a cascade structure in which features are progressively down-sampled, convolved and up-sampled in four units. Identity mapping and convolution operations are performed in each unit. They are shown in the right side of the diagram.

DOI: 10.3844/jcssp.2026.3250.3267

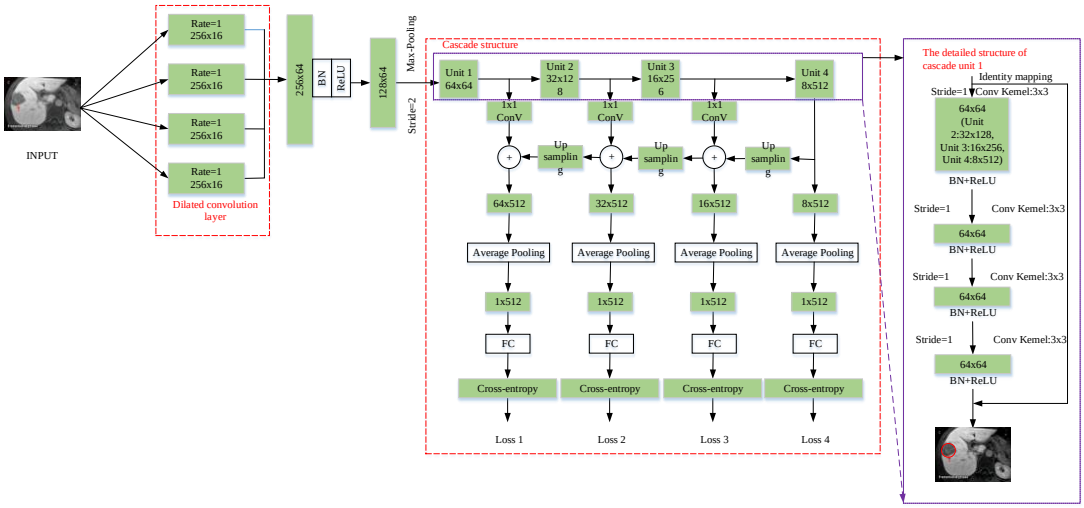

Fig. 3: Architecture of modified cascaded-fully convolutional neural network

Each unit downsamples the features and applies convolution layers. The output is then fed to the next unit and an upsampling module. The features are up sampled and then averaged and passed through a Fully Connected (FC) layer for classification. This results in four independent loss outputs computed by cross-entropy loss for multi-level supervision. The entire structure of each unit of the cascade is composed of identity mappings and convolution layers with ReLU activation functions used throughout. This multi-scale approach enables efficient feature learning at different levels of resolution, which enhances performance for tasks such as object detection or medical image analysis.

The purpose of a convolutional layer is to compute the convolution and slide the filter kernels to extract feature maps from inputs between the kernels and the input local areas, which is expressed mathematically as:

$$y^{l(i,y)} = K_i^l, x^{l(r^j)} = \sum_{j'=0}^{W} K_i^l(j')\, x^{l(j+j')} \tag{16}$$

Where K_i^l stands for the layer's ith filter kernel weights $l.x^{l(r^3)}$ symbolizes the jth local area in layer$l$. W is the kernel's width. The kernel's jth weight is indicated by $K_i^l(j')$. CNN relies heavily on pooling as well. In addition to enabling a representation that is invariant to slight input translation, this type of subsampling can decrease the feature dimensions. One of the most popular subsampling techniques, max pooling can be explained as:

$$p^{l(i,j)} = \max_{(j-1)W+1 \le t \le jW} \{a^{l(i,t)}\} \tag{17}$$

Where the neuron's value in the pooling process's layer l is represented by$p^{l(i,j)}$. The tth neuron's value in layer l's ith frame is indicated by$a^{l(i,t)}$, where$(j-1)W+1 \le t \le jW$. The pooling region's width is denoted by W.

Cascade Structure

In this CFNN network, backbone (inception v2) serves as the foundation for the cascade structure. A basic example is displayed. To improve the model's generalization and allow for the use of noticeably higher learning rates, BN is applied between each convolutional layer and activation layer for every unit. The transformation of the BN layer, given a minibatch $\{x1, x2, \ldots, xm\}$ with the parameters γ *and* β, may be explained as:

$$\mu = \frac{1}{m}\sum_{i=1}^{m} x_i \tag{18}$$

$$BN_{\gamma,\beta(x_i)} = \gamma\hat{x}i + \beta \tag{19}$$

The activation layer comes next. The activation function for the ReLU, which is defined as:

$$a = max\{0, BN_{\gamma,\beta(x_i)}\} \tag{20}$$

Where $BN_{\gamma,\beta(x_i)}$ is the output of the BN layer.

The advantages of incorporating a Conditional Random Field (CRF) into a CFCN for liver tumor classification are significant. First, CRFs contribute to the improvement of the segmentation results by considering the spatial relations between pixels, which is important for the accurate delineation of the boundaries of liver tumors. This refinement produces segmentation outputs that are more accurate and visually consistent, thereby reducing the chances of misclassification. CRF provides a probabilistic model for local, modular image. As established by the Hammersley-Clifford theorem, the

CRF and Gibbs distribution are equivalent for the same graph, stating that, given the Gibbs distribution, any Gibbs joint probability distribution will satisfy the conditional independence specified by the corresponding CRF. Hence, we can write the Gibbs distribution with CRF as:

$$p(X = x \mid I) = \frac{1}{Z(1)} \exp(-E(x \mid I)) \quad (21)$$

Here, $p(X = x \mid I)$ is the probability that the random variable X takes on a specific value x given the input$I$. The term $Z(1)$ is the partition function, which normalizes the distribution to ensure that the probabilities sum to 1. The function $E(x \mid I)$ is the energy function that measures the compatibility of the configuration x with the input I. The exponential term $\frac{1}{Z(1)} exp(-E(x \mid I))$ ensures that configurations with lower energy (higher compatibility) are more probable. This formulation allows the CRF to model complex dependencies between variables and inputs, making it effective for tasks like image segmentation and labeling.

In this case, predictions are made using cross-entropy loss. The cross-entropy is defined for the target distribution $p(x)$ and the estimated distribution $q(x)$:

$$LOSS_i = H(p(x), q(x)) = -\sum_x p(x) log\, q(x) \quad (22)$$

Where $Loss$ stands for unit is cross-entropy loss. The overall loss $Loss = i(Lossi)$ is typically used as the loss function in conventional gradient reduction techniques. Next, in order to determine the disease's severity such as how medium and high the segmented portion of the tumor has been evaluated based on the entire area of the original CT scan. The distinction between the segmented part of the tumor and the original image, the actual size of the tumour portion, can be measured to predict the severity level.

Results and Discussion

Modified cascaded-fully convolutional neural network and optimized Swin U-Net for automatic liver lesion prediction. Python 3.8 was used for the development of the proposed model and for the evaluation of the process. The system specifications are as follows: RAM: 16GB, GPU: NVidia GeForce GTX 1650 and CPU: Intel Core i5.

The simulation parameters of the three different methods MCFCNN, Swin U-Net and DIPN are shown in Table 1. Each methodology has its own parameters that are critical to model evaluation and training. The parameters of the MCFCNN method are the learning rate, loss function, number of epochs, and optimizer type. Batch size, number of epochs, learning rate, optimizer and loss function are set for Swin U-Net method. The loss function, the number of epochs, the batch size and d of the DIPN methodology are shown.

Table 1: Simulation parameter

Methods	Parameters	Ranges
MCFCNN	optimizer	Adam
	learning_rate	0.001
	loss	'categorical_crossentropy'
	epochs	100
Swin U-Net	optimizer	Adam
	learning_rate	1e-4
	loss	'binary_crossentropy'
	epochs	100
	batch_size	8
DIPN	learning_rate	0.001
	batch_size	1
	epochs	10
	loss	'mse'

These parameters are important in the preparation of the training process and in ensuring that the models perform best on the respective tasks they are meant to solve.

The main goal of this study is to evaluate the performance of the proposed integrated framework against conventional baseline methods on a common experimental protocol. Recently, segmentation models based on transformers, such as Swin Transformer, TransUNet, UNETR and nnFormer, have shown promising results for medical image segmentation. However, the present work does not provide a complete comparative evaluation of these architectures. Such comparisons would provide a more complete evaluation of the proposed framework, and are recognized as a valuable direction for future work.

The softmax classification layer after the full set of feature extraction layers outputs class probability scores. The resulting probability scores are then fed into a trainable CRF-like layer for refinement, which learns pairwise relationships between classes through a trainable potential matrix. In contrast to the typical Dense CRF approaches used for pixel-wise segmentation maps, the proposed CRF-inspired layer refines the predicted class probabilities in the classification space, which results in an improvement in the consistency of the final classification result.

The executed CRF-like layer maintains a trainable pairwise potential matrix of size (CXC), where (C) is the number of output classes. Prior to the final prediction, the obtained potential matrix is multiplied with the Softmax probability vector to enhance the class probabilities. The light-weight enhancement method brings little computational overhead and allows the model to learn the inter-class relationships comprehensively.

Dataset Description

The Liver Tumor Segmentation Benchmark (LiTS) dataset consists of 130 CT scans from patients diagnosed with liver cancer. The dataset contains 985 CT images. There are 519 tumorous (malignant) and 466 non-tumorous

DOI: 10.3844/jcssp.2026.3250.3267

(benign) images in the total images. These two categories are mutually exclusive and collectively exhaustive of the data set. (Huang et al., 2024; Prakash et al., 2023).

Figure 4 shows the exemplary results of the pre-processing and the liver tumor identification procedure. The first picture shows the original noisy picture, which is full of many artifacts and inconsistencies which could interfere with accurate evaluation. The next one is the denoised one. The pre-processing step has done a good job at removing the noise. This leads to a more lucid image. The third image is TCN (Temporal Convolutional Network) image which signifies the application of advanced techniques to improve image quality and highlight the relevant features. The segmentation image finally shows the regions of the liver tumor outlined and well separated from the surrounding tissue. This image set emphasizes the importance of each pre-processing step for improving the accuracy and reliability of the liver tumor detection.

Figure 5 is an example of a confusion matrix, which is one way to assess the performance of a classification model.

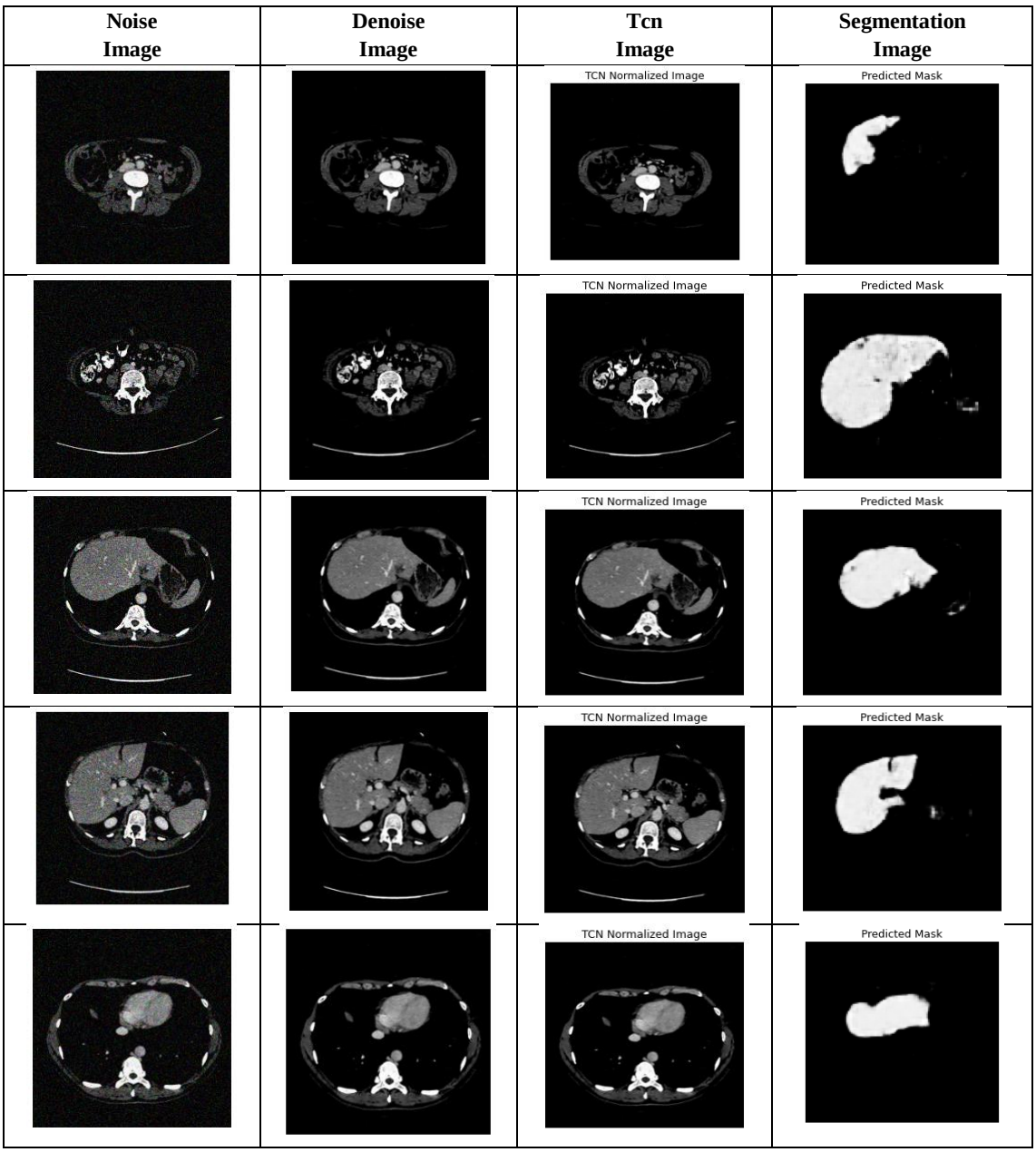

Fig. 4: Sample output of pre-processing image and liver tumor detection

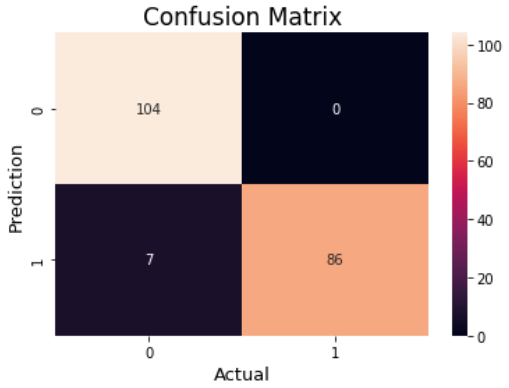

Fig. 5: Confusion matrix of proposed model

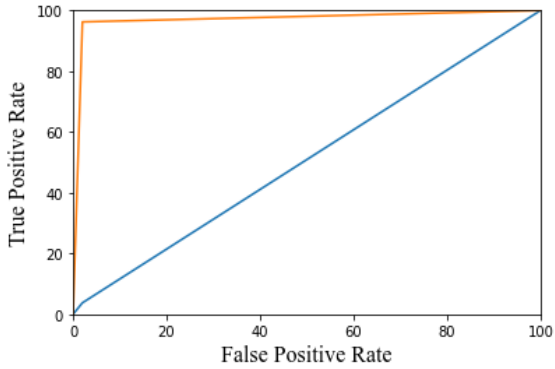

Fig. 6: Proposed model's ROC curve

It compares the model predicted labels with the true labels or ground truth. This matrix shows the real labels as columns and the expected labels as rows. The table top left cell (104) represents the number of True Negatives, the number of cases correctly predicted as {0}. The bottom right cell (86) is the True Positives or the correctly identified {1} s. The top right cell (0) corresponds to False Positives, i.e. no samples were mistakenly predicted as ` 1 ' instead of ` 0 '. The bottom left cell (7) represents the False Negatives, i.e. the number of instances that were incorrectly labelled as `0` instead of `1`. The model presented a good predictive performance, with only seven False Negatives and zero False Positives.

ROC curve shown in Fig. 6 can be used to evaluate a classification model by plotting the True Positive Rate (Sensitivity) against the False Positive Rate (1-Specificity). The orange curve shows the performance of the model, while the blue line on the diagonal is a random classifier. As the curve approaches the upper left hand corner, the model gets better at discriminating between the categories. Orange curve close to the top-left corner means good classification performance.

The dataset was randomly split into training (80%) and testing (20%) subsets at the image level for the current study. We adopted this experimental protocol to evaluate the proposed framework under the same benchmark setting. While patient-level partitioning is generally preferred for clinical evaluation purposes to reduce potential data leakage and to obtain a more rigorous estimate of model generalization, this approach was not considered in the current investigation. This limitation has been noted and will be taken into account in future work.

Segmentation Comparison of Proposed and Existing Model

We calculate the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) to compare the segmentation performance of our proposed Swin U-Net model with other models like DeepLab V3, V-Net and U-Net. The Swin U-Net architecture demonstrates better efficiency with high SSIM and PSNR scores than other models. This indicates that Swin U-Net is good at preserving image quality and structure similarity and is a strong candidate for segmentation tasks.

As shown in Figure 7, the Swin U-Net achieves an SSIM of 0.96%, which is better than U-Net (0.87%), V-Net (0.81%), and Deeplab V3 (0.78%). This demonstrates its outstanding ability of preserving the structural similarity, where the Swin U-Net achieves the best Peak Signal-to-Noise Ratio (PSNR) of 61.97%, far ahead of U-Net (55.12%), V-Net (52.51%) and Deeplab V3 (46.83%), as shown in Figure 8.

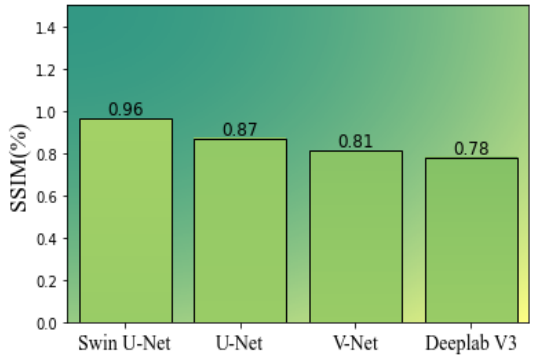

Fig. 7: Proposed model comparison of SSIM

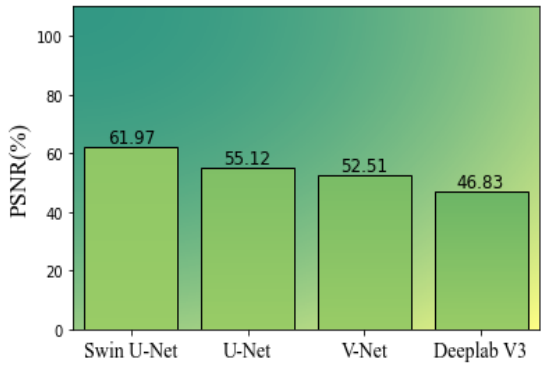

Fig. 8: Comparison of proposed and existing model of PSNR

Comparison of Proposed and Existing Model

The comparison presented in this study uses representative deep learning architectures that have been widely used as baseline methods in the literature. Although these comparisons demonstrate the effectiveness of the proposed framework, comparisons against recent liver-specific segmentation networks such as nnUNet, TransUNet and UNETR would provide additional insights regarding the relative performance of the proposed method. A further avenue for research that has been identified is this broad comparative analysis.

We calculated the reported Precision, Recall (Sensitivity) and Specificity using a macro-averaging strategy. The performance metrics were computed separately for each class using the confusion matrix and then averaged across both classes. The reported Recall and Specificity values are thus macro-averaged measures and not class-specific metrics. Under this evaluation protocol, similar macro-averaged values can occur when class-wise performance shows a complementary behaviour

The proposed framework showed promising overall performance, however the segmentation stage was not evaluated with dedicated segmentation metrics including Dice Coefficient, Intersection over Union (IoU), and Pixel Accuracy. Incorporating these metrics would provide a more comprehensive evaluation of segmentation quality and will be considered in future work.

The proposed Cascaded-Fully Convolutional Neural Network (CFCNN) model is assessed using well-known models such as Alex Krizhevsky Network (AlexNet), Visual Geometry Group 16-layer Convolutional Neural Network (VGG-16 CNN), Hybridized Fully Convolutional Neural Network (HFCNN), and Efficient Deep Convolutional Neural Network (EDCNN) with respect to accuracy, precision, recall, specificity, and other metrics. We can see that the CFCNN model performs extremely well and is better than all other models in all evaluation metrics. This illustrates the effectiveness and robustness of CFCNN on a range of tasks.

Figure 9 shows that the proposed CFCNN has an FDR of 3.15%, which is much better than AlexNet (5.81%), VGG-16 CNN (9.85%), Faster R-CNN (11.09%), and EDCNN (14.16%). Thus, CFCNN seems to be the best of all the models tested in terms of reducing false discoveries. In Figure 10, CFCNN demonstrates the highest Negative Predictive Value (NPV) with a value of 96.85%, surpassing the NPV of AlexNet (91.12%), VGG-16 CNN (88.51%), Faster R-CNN (84.83%) and EDCNN (81.43%). This demonstrates the model's excellent capability to predict negatives accurately in comparison to other models.

The proposed CFCNN in Figure 11 has the highest specificity of 96.24% compared to AlexNet (91.24%), VGG-16 CNN (87.91%), Faster R-CNN (83.21%), and EDCNN (80.40%). This points out the ability of CFCNN to correctly discriminate true negatives. Figure 12 shows that CFCNN has the highest recall rate of 96.24%, followed by AlexNet (93.24%), VGG-16 CNN (90.59%), Faster R-CNN (87.66%), and EDCNN (85.24%). This shows its ability to accurately identify true positives.

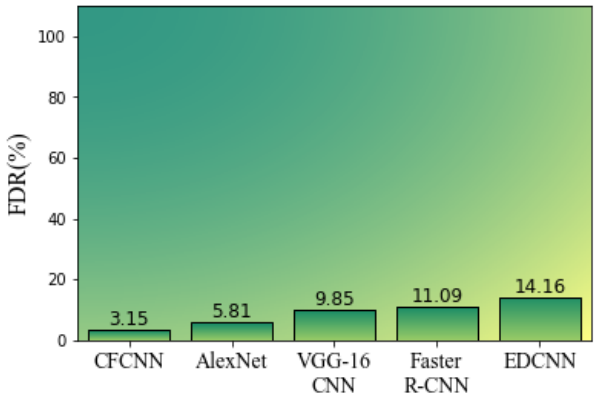

Fig. 9: Comparison between the proposed and current FDR models

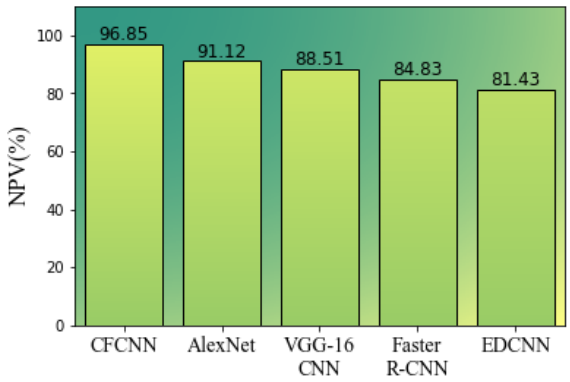

Fig. 10: NPV comparison for the suggested model

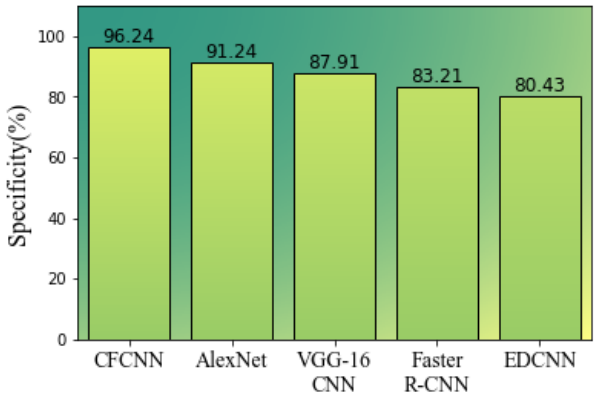

Fig. 11: Comparison of Specificity for proposed model

DOI: 10.3844/jcssp.2026.3250.3267

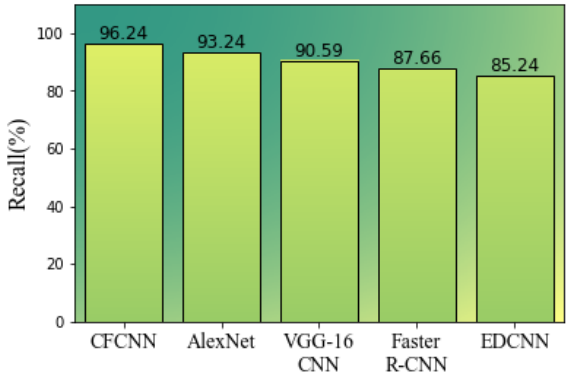

Fig. 12: Comparison of recall

The comparison of CFCNN's precision against alternative methods is presented in Figure 13. The proposed model has a 96.85% accuracy which is higher than the AlexNet (94.19%), VGG-16 CNN (90.15%), Faster R-CNN (88.91%) and EDCNN (85.84%). The proposed CFCNN achieved the lowest error rate 3.55% with significant improvement over AlexNet (5.90%), VGG-16 CNN (8.50%), Faster R-CNN (13.00%) and EDCNN (17.00%). This shows its dependability in decreasing the prediction errors as illustrated in Figure 14.

Figure 15 shows that CFCNN has the highest accuracy rate of 96.45% which is better than AlexNet (94.10%), VGG-16 CNN (91.50%), Faster R-CNN (87.00%) and EDCNN (83.00%). This emphasizes the improved overall performance of the proposed model. Figure 16 shows the Jaccard Index of CFCNN is 93.08%, higher than AlexNet (89.12%), VGG-16 CNN (85.51%), Faster R-CNN (82.83%), and EDCNN (78.43%). This indicates the improved segmentation accuracy of CFCNN.

As shown in Figure 17, the proposed CFCNN achieves the lowest False Omission Rate (FOR) of 3.15% compared to AlexNet (8.88%), VGG-16 CNN (11.49%), Faster R-CNN (15.17%), and EDCNN (18.57%).

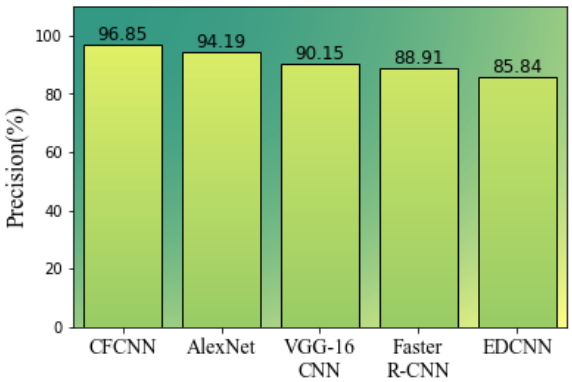

Fig. 13: Precision comparison between the suggested and current models

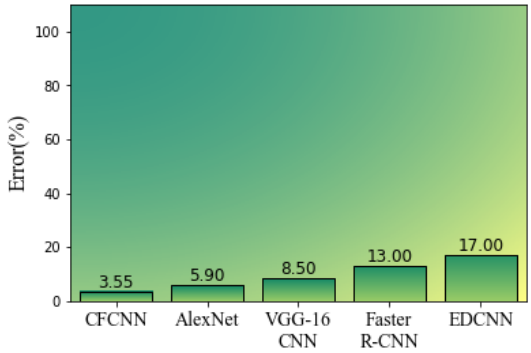

Fig. 14: Error comparison

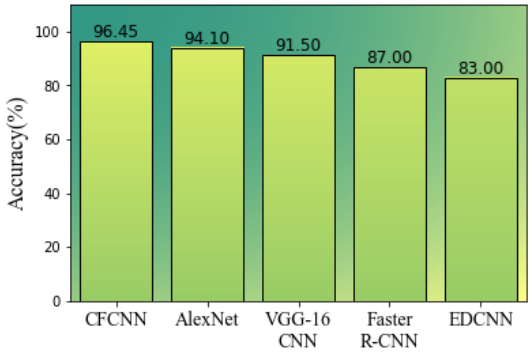

Fig. 15: Accuracy comparison between the suggested and current models

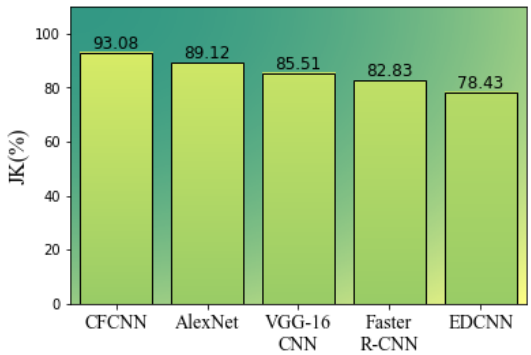

Fig. 16: Suggested model comparison for JK

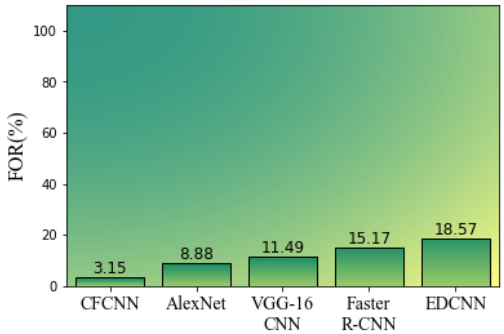

Fig. 17: FOR comparison of proposed and existing model

This underscores its efficacy in avoiding missed positive cases. The proposed CFCNN achieves a False Positive Rate (FPR) of 3.76% which outperforms AlexNet (8.76%), VGG-16 CNN (12.09%), Faster R-CNN (16.79%) and EDCNN (19.57%) and demonstrates its improved capability in minimizing false alarms as shown in Figure 18. As can be seen in Figure 19, CFCNN has the highest value of the Matthews Correlation Coefficient (MCC) of 96.45% than AlexNet (92.00%), VGG-16 CNN (89.00%), Faster R-CNN (84.00%), and EDCNN (81.00%).

The Receiver Operating Characteristic (ROC) curves of the proposed framework and some representative baseline methods are shown in figure. The methods were selected to allow a clear visual comparison while keeping the readability of the ROC plot. The ROC curves of all the evaluated methods would overlap a lot and the figure would be less interpretable. The corresponding tables and performance analyses include the quantitative performance comparison of all evaluated methods.

State of the Art Comparison

This section compares the suggested approach with the current model in terms of liver tumor, as shown in Table 2.

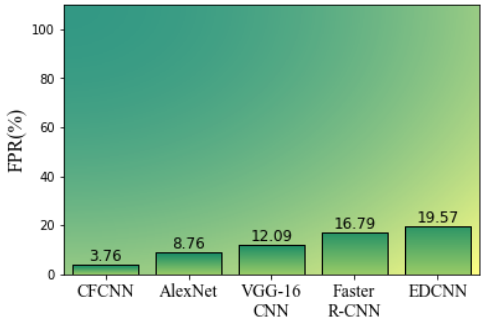

Fig. 18: Proposed model comparison of FPR

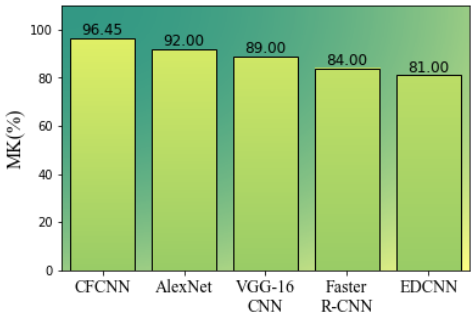

Fig. 19: MK comparison of proposed and existing model

Table presents the comparative results from the respective published studies to provide a literature-based perspective on performance. It is to be noted that there may exist differences in datasets, pre-processing, training/testing splits and experimental settings across these studies and hence the comparison should be considered indicative rather than a direct experimental benchmark. A comprehensive evaluation of all competing methods on a common dataset and the same experimental protocol is an important future direction.

Table presents the comparison of state-of-the-art models for liver tumor detection. Here the proposed model shows competitive performance in different metrics. The proposed model achieves better accuracy and precision, outperforming D-ResNet and CNN-HCC in these metrics. However, D-ResNet and CNN-HCC show considerable specificity, with CNN-HCC showing impressive ability in this regard. The proposed model also has higher recall, which means the ability to identify true positive. Even though there are some differences in some metrics, the proposed model is still a valid option for identifying liver tumors, achieving an effective balance between accuracy, precision, and recall.

Ablation Study

In this section, various cases are developed and the proposed model is evaluated. Some cases are demonstrated in Table 3.

Table 3 presents an ablation study on the effectiveness of different configurations of the proposed model. The model exhibits average effectiveness when the basic integration of Deep Image Prior Network (DIPN) and CFCNN is used. The CFCNN element could be improved with respect to precision and accuracy. The effectiveness of the model is greatly enhanced by the addition of a temporal convolutional network (TCN) along with CFCNN. The combination of Swin U-Net and CFCNN results in considerable improvements for all metrics. Removing the Feature-Oriented Attention (FOA) mechanism from Swin U-Net leads to improved performance. The best improvements are shown when the TCN and Swin U-Net without FOA are combined with the CFCNN, which gives better results in all the metrics. This reflects the joint effect of these elements working together, and the importance of careful integration and choice in enhancing model efficacy.

Table 2: Performance comparison of the proposed method with previous works in the literature for liver lesion classification.

Metric	Proposed	CFCNN (Sun et al 2022)	D-ResNet (Wang et al., 2021)	CNN-HCC
Accuracy	96.45%	94.02%	91%	93%
Precision	96.85%	-	92%	95%
Recall	96.24%	-	75%	-
Specificity	96.24%	77.8%	-	94%

Table 3: Ablation Study of proposed model

Cases	Accuracy	Precision	Recall	Specificity
DIPN + CFCNN	80.70%	81.20%	80.80%	81.35%
DIPN + Modified-CFCNN	81.20%	83.40%	81.30%	81.50%
DIPN+ TCN + CFCNN	84.20%	84.80%	83.30%	82.10%
DIPN + Swin U-Net+ CFCNN	87.70%	85.55%	84.60%	82.36%
DIPN + Swin U-Net - FOA+ CFCNN	93.50%	90.50%	90.90%	87.30%
DIPN+TCN + Swin U-Net - FOA+ CFCNN	96.45%	96.85%	96.24%	96.24%

Conclusion

A modified Cascaded-Fully Convolutional Neural Network (CFCNN) was constructed using an optimized Swin U-net to identify liver lesions in medical imaging data. Non-malignant lesions do not require medical treatment, but it is very important to find malignant lesions quickly and accurately, in order to have better outcomes for patients and to be able to treat them effectively. But these technologies are prohibitively expensive and not easily accessible, posing a huge challenge. Hence, a new model based on CFCNN combined with Fossa Optimization Algorithm and Swin U-Net was proposed to overcome the challenges in predicting liver lesions. The process begins with liver lesion images from Kaggle, which are pre-processed using a Regularization-based Deep Image Prior Network (DIPN) Transition Constant Normalization (TCN) to improve image quality and consistency. After that, the images are segmented using the Swin U-Net (SUN) model. The model is improved by optimizing the hyperparameters using the Fossa Optimization Algorithm (FOA) to improve the accuracy of segmentation. The segmented classified images are processed by an improved CFCNN model, where a Conditional Random Field (CRF) is incorporated in the fully connected layer to enhance the accuracy of the prediction. The mask of the segmented tumour will be used to assess the severity of the disease (low, medium or high).

The proposed framework shows that the combination of complementary deep learning components can be effective in improving the performance of liver lesion prediction, while the method provides a basis for future methodological improvements and wider clinical validation.

The proposed framework has been evaluated as a complete end-to-end pipeline including DIPN based image denoising, Transition Constant Normalization (TCN), FOA based hyperparameter optimization, Optimized Swin U-Net segmentation, Modified Cascaded Fully Convolutional Neural Network (MCFCNN) and CRF based refinement. Therefore, the performance metrics reported in this study represent the combined effectiveness of the integrated framework. A detailed ablation study evaluating each module individually would provide further insight into the contribution of individual components, but such an investigation is outside the scope of the present work.

The proposed framework is evaluated for performance in terms of segmentation quality and overall classification performance using Accuracy, Precision, Recall, Specificity, F1-score, PSNR, and SSIM. These metrics show the effectiveness of the proposed framework, but widely used overlap based segmentation metrics such as the Dice Similarity Coefficient (DSC) and Intersection over Union (IoU) could be used to provide further insight into segmentation accuracy. It has been pointed out that the inclusion of these metrics is a key direction for future research.

Limitation and Future Scope

The Cascaded-Fully Convolutional Neural Network (CFCNN) has achieved promising results for various applications, but it has some limitations. First of all, it highly relies on huge amounts of good training data, which can be a problem in data-poor domains. Secondly, Swin Transformer U-Net may result in an increased complexity and computational cost. Below are discussed several key areas for future research on CFCNN. When data is limited, data augmentation techniques and transfer learning can help improve performance. Another promising direction is to optimize the model architecture to reduce the computational requirements without performance loss. We can use explainability techniques to increase interpretability, thus making it more transparent and trustable. Exploring domain adaptation methods can make it more robust on different datasets. Furthermore, the applicability of CFCNN can be extended in real time applications, by enabling it and its integration with other modalities.

In future work, we will conduct a comprehensive evaluation and further refinement of the proposed framework from theoretical, experimental and clinical perspectives. The systematic convergence property and computational complexity of Fruit Fly Optimization (FOA) algorithm will be systematically studied and compared with well-known hyperparameter optimization methods such as Grid Search, Bayesian Optimization, Genetic Algorithm (GA), Particle Swarm Optimization (PSO) and Gray Wolf Optimization (GWO) to evaluate the effectiveness and computational efficiency. We will also carry out a detailed sensitivity analysis of the weight decay parameter over different search ranges and optimization strategies to study its effect on segmentation and classification performance and find appropriate parameter settings. Future work will include

application of patient-level data partitioning, k-fold cross-validation and independent multi-center clinical datasets to assess robustness and generalization capability of the framework to minimize the risk of data leakage and to test the model in different clinical scenarios. Furthermore, a systematic ablation study will be conducted by successively disable/activate each component of the proposed framework (DIPN, TCN, FOA, MCFCNN and CRF) to measure the contribution of each component and their combinations to the segmentation accuracy, classification performance, computational complexity and overall robustness. To evaluate the segmentation performance, comprehensive overlap-, pixel- and boundary-based metrics like Dice Similarity Coefficient (DSC), Intersection over Union (IoU), Pixel Accuracy and Hausdorff Distance will be used. Moreover, the proposed segmentation framework will be extensively compared with state-of-the-art Transformer-based and traditional medical image segmentation architectures, e.g., Swin Transformer, TransUNet, UNETR, nnFormer, nnUNet and other recent liver lesion segmentation models, on the same datasets, preprocessing pipeline, training/testing protocols and evaluation criteria to ensure fair and reproducible comparisons. Similarly, the classification component will be benchmarked against recent state-of-the-art liver lesion classification approaches under a common dataset, processing pipeline and experimental protocol. These studies will help gain a deeper insight on the optimization behavior, component-wise effectiveness, segmentation and classification capability, computational efficiency, robustness and clinical generalization of the proposed framework, and hence will strengthen its potential for reliable deployment in real-world medical imaging applications.

Acknowledgment

We acknowledge Prof. (Dr) Ashish Kumar Chakraverti, Department of CSE (DS), Noida Institute of Engineering and Technology, Greater Noida for his continuation support in editing of this research paper and providing the guidelines for simulation and code.

Funding Information

The authors have not received any financial support or funding to report.

Author's Contributions

Sugandha Chakraverti: Has designed the model and implemented the code as a core researcher for this paper.

Tejaswi Khanna: Has guided for refinement of models and result improvement after analyzing the initial results.

Vijay Shukla: Provided the insight about various techniques optimization and integrated these techniques in this research.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

References

Ahn, J. C., Connell, A., Simonetto, D. A., Hughes, C., & Shah, V. H. (2021). Application of Artificial Intelligence for the Diagnosis and Treatment of Liver Diseases. *Hepatology*, *73*(6), 2546–2563. https://doi.org/10.1002/hep.31603

Allenki, J., & Soni, H. K. (2025). DEDSWIN-Net: Dual Encoder Dilated Convolution and Swin Transformer Network for the Classification of Liver CT Images. *International Journal of Computational Intelligence Systems*, *18*(1), 198. https://doi.org/10.1007/s44196-025-00937-x

Almotairi, S., Kareem, G., Aouf, M., Almutairi, B., & Salem, M. A.-M. (2020). Liver Tumor Segmentation in CT Scans Using Modified SegNet. *Sensors*, *20*(5), 1516. https://doi.org/10.3390/s20051516

Baguer, D. O., Leuschner, J., & Schmidt, M. (2020). Computed tomography reconstruction using deep image prior and learned reconstruction methods. *Inverse Problems*, *36*(9), 094004. https://doi.org/10.1088/1361-6420/aba415

Gul, S., Khan, M. S., Bibi, A., Khandakar, A., Ayari, M. A., & Chowdhury, M. E. H. (2022). Deep learning techniques for liver and liver tumor segmentation: A review. *Computers in Biology and Medicine*, *147*, 105620. https://doi.org/10.1016/j.compbiomed.2022.105620

Huang, J., Zhou, M., Zhang, J., Yang, G., Yao, M., Li, C., Xiong, Z., & Zhao, F. (2023). Transition-constant Normalization for Image Enhancement. *Advances in Neural Information Processing Systems 36*, 20562–20576. https://doi.org/10.52202/075280-0903

Huang, S., Nie, X., Pu, K., Wan, X., & Luo, J. (2024). A flexible deep learning framework for liver tumor diagnosis using variable multi-phase contrast-enhanced CT scans. *Journal of Cancer Research and Clinical Oncology*, *150*(10), 443. https://doi.org/10.1007/s00432-024-05977-y

Jiang, L., Hu, J., & Huang, T. (2025). Improved SwinUNet with fusion transformer and large kernel convolutional attention for liver and tumor segmentation in CT images. *Scientific Reports*, *15*(1), 14286. https://doi.org/10.1038/s41598-025-98938-5

Lin, W., Hasenstab, K., Cunha, G. M., & Schwartzman, A. (2020). Comparison of handcrafted features and convolutional neural networks for liver MR image adequacy assessment. *Scientific Reports*, *10*(1), 20336. https://doi.org/10.1038/s41598-020-77264-y

Meng, L., Tian, Y., & Bu, S. (2020). Liver tumor segmentation based on 3D convolutional neural network with dual scale. *Journal of Applied Clinical Medical Physics*, *21*(1), 144–157. https://doi.org/10.1002/acm2.12784

Nkinahamira, F., Feng, A., Zhang, L., Rong, H., Ndagijimana, P., Guo, D., Cui, B., & Zhang, H. (2024). Machine learning approaches for monitoring environmental metal pollutants: Recent advances in source apportionment, detection, quantification, and risk assessment. *TrAC Trends in Analytical Chemistry*, *180*, 117980. https://doi.org/10.1016/j.trac.2024.117980

Peden, C. J., Aggarwal, G., Aitken, R. J., Anderson, I. D., Foss, N. B., Cooper, Z., Dhesi, J. K., French, W. B., Grant, M. C., Hammarqvist, F., Hare, S. P., Havens, J. M., Holena, D. N., Hübner, M., Kim, J. S., Lees, N. P., Ljungqvist, O., Lobo, D. N., Mohseni, S., … Scott, M. (2021). Guidelines for Perioperative Care for Emergency Laparotomy Enhanced Recovery After Surgery (ERAS) Society Recommendations: Part 1—Preoperative: Diagnosis, Rapid Assessment and Optimization. *World Journal of Surgery*, *45*(5), 1272–1290. https://doi.org/10.1007/s00268-021-05994-9

Pomohaci, M. D., Grasu, M. C., Băicoianu-Nițescu, A.-Ș., Enache, R. M., & Lupescu, I. G. (2025). Systematic Review: AI Applications in Liver Imaging with a Focus on Segmentation and Detection. *Life*, *15*(2), 258. https://doi.org/10.3390/life15020258

Prakash, N. N., Rajesh, V., Namakhwa, D. L., Pande, S. D., & Ahammad, S. H. (2023). A DenseNet CNN-based liver lesion prediction and classification for future medical diagnosis. *Scientific African*, *20*, e01629. https://doi.org/10.1016/j.sciaf.2023.e01629

Rahman, H., Bukht, T. F. N., Imran, A., Tariq, J., Tu, S., & Alzahrani, A. (2022). A Deep Learning Approach for Liver and Tumor Segmentation in CT Images Using ResUNet. *Bioengineering*, *9*(8), 368. https://doi.org/10.3390/bioengineering9080368

Rather, I. H., Kumar, S., & Gandomi, A. H. (2024). Breaking the data barrier: a review of deep learning techniques for democratizing AI with small datasets. *Artificial Intelligence Review*, *57*(9), 226. https://doi.org/10.1007/s10462-024-10859-3

Razik, A., Malla, S., Goyal, A., Gamanagatti, S., Kandasamy, D., Das, C. J., Sharma, R., & Gupta, A. K. (2022). Unusual Primary Neoplasms of the Adult Liver: Review of Imaging Appearances and Differential Diagnosis. *Current Problems in Diagnostic Radiology*, *51*(1), 73–85. https://doi.org/10.1067/j.cpradiol.2020.10.001

Sakthisaravanan, B., & Meenakshi, R. (2020). OPBS-SSHC: outline preservation based segmentation and search based hybrid classification techniques for liver tumor detection. *Multimedia Tools and Applications*, *79*(31–32), 22497–22523. https://doi.org/10.1007/s11042-019-08582-1

Sethia, K., Strakos, P., Jaros, M., Kubicek, J., Roman, J., Penhaker, M., & Riha, L. (2025). Advances in liver, liver lesion, hepatic vasculature, and biliary segmentation: A comprehensive review of traditional and deep learning approaches. *Artificial Intelligence Review*, *58*(10), 299. https://doi.org/10.1007/s10462-025-11310-x

Sohaib, M., Shabani, S., Mohammed, S. A., & Parvin, B. (2025). 3D-Organoid-SwinNet: High-Content Profiling of 3D Organoids. *IEEE Journal of Biomedical and Health Informatics*, *29*(2), 792–798. https://doi.org/10.1109/jbhi.2024.3511422

Sun, K., Meng, Y., & Dong, S. (2022). FoSSA Optimization-Based SVM Classifier for the Recognition of Partial Discharge Patterns in HV Cables. *Computational Intelligence and Neuroscience*, *2022*, 1–12. https://doi.org/10.1155/2022/7566731

Wang, F., Liu, R., Hu, Q., & Chen, X. (2021). Cascade Convolutional Neural Network With Progressive Optimization for Motor Fault Diagnosis Under Nonstationary Conditions. *IEEE Transactions on Industrial Informatics*, *17*(4), 2511–2521. https://doi.org/10.1109/tii.2020.3003353

Wong, G. L., Yuen, P., Ma, A. J., Chan, A. W., Leung, H. H., & Wong, V. W. (2021). Artificial intelligence in prediction of non-alcoholic fatty liver disease and fibrosis. *Journal of Gastroenterology and Hepatology*, *36*(3), 543–550. https://doi.org/10.1111/jgh.15385

Xi, P., Guan, H., Shu, C., Borgeat, L., & Goubran, R. (2020). An integrated approach for medical abnormality detection using deep patch convolutional neural networks. *The Visual Computer*, *36*(9), 1869–1882. https://doi.org/10.1007/s00371-019-01775-7

Zhang, P., Gao, C., Huang, Y., Chen, X., Pan, Z., Wang, L., Dong, D., Li, S., & Qi, X. (2024). Artificial intelligence in liver imaging: methods and applications. *Hepatology International*, *18*(2), 422–434. https://doi.org/10.1007/s12072-023-10630-w

Zheng, R., Wang, Q., Lv, S., Li, C., Wang, C., Chen, W., & Wang, H. (2022). Automatic Liver Tumor Segmentation on Dynamic Contrast Enhanced MRI Using 4D Information: Deep Learning Model Based on 3D Convolution and Convolutional LSTM. *IEEE Transactions on Medical Imaging*, *41*(10), 2965–2976. https://doi.org/10.1109/tmi.2022.3175461