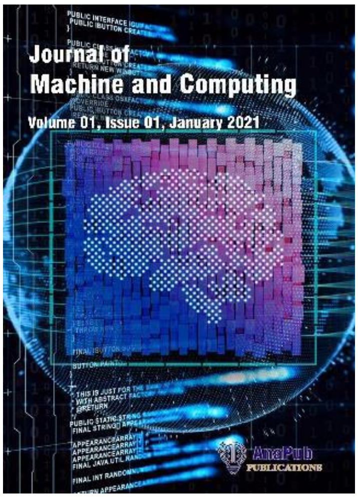
Journal Pre-proof

Innovative Deep Learning Models for Accurate Segmentation and Classification in Oncological Diagnosis Data



Archana R and Anand L

DOI: 10.53759/7669/jmc202505112

Reference: JMC202505112

Journal: Journal of Machine and Computing.

Received 18 September 2024

Revised form 30 January 2025

Accepted 01 May 2025

Please cite this article as: Archana R and Anand L, “Innovative Deep Learning Models for Accurate Segmentation and Classification in Oncological Diagnosis Data”, Journal of Machine and Computing. (2025). Doi: https:// doi.org/10.53759/7669/jmc202505112.

This PDF file contains an article that has undergone certain improvements after acceptance. These enhancements include the addition of a cover page, metadata, and formatting changes aimed at enhancing readability. However, it is important to note that this version is not considered the final authoritative version of the article.

Prior to its official publication, this version will undergo further stages of refinement, such as copyediting, typesetting, and comprehensive review. These processes are implemented to ensure the article's final form is of the highest quality. The purpose of sharing this version is to offer early visibility of the article's content to readers.

Please be aware that throughout the production process, it is possible that errors or discrepancies may be identified, which could impact the content. Additionally, all legal disclaimers applicable to the journal remain in effect.

AnaPub
PUBLICATIONS

Innovative Deep Learning Models for Accurate Segmentation and Classification in Oncological Diagnosis Data

Archana R[1] and Anand L [2],*
Department of Networking and Communications, School of Computing, College of Engineering and Technology,
SRM Institute of Science and Technology, Kattankulathur, Chengalpattu, 603203, India.
ar6869@srmist.edu.in and anandl@srmist.edu.in

***Abstract*:**

Accurate identification and classification of tumours are essential for effectively diagnosing and treating hepatocellular carcinoma and metastatic disease. However, the heterogeneous nature of tumours, characterized by irregular boundaries and variations in shape, size, and location, poses significant challenges for precise and automated segmentation and classification. With recent advances in artificial intelligence, deep learning has emerged as a powerful tool for medical image analysis. Although current clinical methods offer baseline performance in tumour classification, there is still considerable scope for improving diagnostic accuracy. This research proposes an innovative deep-learning framework to enhance the segmentation and classification of tumours. The approach begins by enhancing image contrast using histogram equalization and reducing noise via a median filter, regions are then accurately segmented from abdominal CT images using Mask R-CNN, a state-of-the-art model based on region-based convolutional neural networks. The segmented outputs are further processed using an Enhanced Swin Transformer to mitigate overfitting and boost classification performance. Experimental results demonstrate that the proposed model achieves superior accuracy and robustness across diverse CT image datasets, exhibiting strong performance even in noise.

Keywords—Segmentation, Deep Learning, RCNN Classification, CT image, Mask R-CNN

I. Introduction

The liver is an organ that is very important to the survival of all vertebrates and animals on our planet. The human body does not exhibit any symptoms of liver disease, despite the fact that it is a potentially deadly ailment. An early recognition of liver disease would be very beneficial to the patient's prognosis from a medical standpoint. When it comes to the diagnosis and treatment of diseases, the computer-aided diagnostic (CAD) system is an extremely important component. The first step in any CAD based medical image processing activities is the segmentation of medical images. In its most basic form, it entails categorizing the medical pictures that are supplied and making use of the segmentation data in order to model pertinent anatomical components for further subsequent applications [1].

When it comes to providing knowledge of human anatomy that does not need any intrusive procedures, medical image segmentation is of critical relevance. In addition to this, it provides radiologists with assistance in recognizing anatomical structures and visualizing them based on the granularity of the pixels. The ultimate objective is to improve the understandability and intuitiveness of human tissue and sick structures [2][3]. Simulation of biological processes, localization of problematic tissues, monitoring of illness development, and provision of the essential information for assessing radiotherapy or surgery are all accomplished via the use of this approach by medical practitioners.

Over the last several years, there has been a rapid improvement in the automated segmentation of histology pictures, particularly H&E slides. In addition to effectively determining the outlines of nuclei, the approaches that are now available ensure that a number of different cell types inside the microenvironment of the tumor may be appropriately identified.

The computerized tumor categorization systems that are now being used are very new, and they often fail to correctly capture the features that are detected in the early stages of the illness. Even while deeper neural networks are effective for classification, they are not practical because of the temporal limitations that they provide. On the other hand, shape-based techniques that make use of past data indicate positive potential. The development of PSMs can be sped up with the use of AI-driven deep learning (DL) [20][21], which is beneficial to medical analysis [3].

A number of characteristics, including size variety, complicated backdrops, ambiguous boundaries, and a lack of contrast in organ density, provide difficulties in the process of human liver segmentation. The accurate segmentation of the liver has the potential to dramatically improve both medical assessment and research. A significant reduction in death rates and an improvement in survival prospects may be achieved via the accurate identification and treatment of liver cancer[4]. Because of its poor prognosis, liver disease is the third major cause of death associated to lesions. This is likely owing to the fact that it is often identified at an inappropriately late stage.

A large portion of the research has focused on other variables, such as the kind of illness, the current stage of the disease, the size, the number, and the course of the disease. Along with these other factors, liver function is a factor that plays a role in determining the treatment plan that is selected. As a consequence of this, we needed to establish a diagnostic aid system in order to detect individuals with liver cancer at an early stage and then allow rapid evaluation of the levels of abnormalities in the liver. This is crucial for those who work in the medical field, especially when it comes to the fact that they may benefit from an intelligent system that could assist them in diagnosis and treatment. In this manner, we provide a unique hybrid deep classifier for the segmentation and classification of liver cancer. This classifier is based on a customized mask-region convolutional neural network (cm-RCNN). The following is a list of the contributions that this work has made.

- The model is designed to undergo a three-stage process which includes pre-processing, liver segmentation, and classification.
- The RCNN strategy that has been presented for liver segmentation is able to predict the area mask of the picture in an effective manner. The technique includes four max-pooling layers, eight transposed 2D convolutional layers, a dropout layer, ReLU, and the modified sigmoid (m-Sig) activation function.
- In an effort to mitigate overfitting, the image with segments is inputted into an Improved Swin Transformer Network with adversarial Propagating.

II. RELATED WORK

F. Hu et al., [5] Dilation Heterogeneous Convolution (DHConv) as a new convolutional kernel structure, which integrates heterogeneous kernel structure with dilated convolution to improve representational efficiency and decreasedata calculations.It is proposed that Mask R-DHCNN be used for cell identification and segmentation. This would include substituting the conventional convolutional kernel in Mask R-CNN with DHConv in order to accommodate the different sizes and shapes of cells that may be seen in microscope images. The indicated method's success is shown by experiments using microscope cell image datasets, highlighting increased performance measures like as AP, Precision, Recall, Dice, and PQ while retaining competitive FLOPs (floating point operations per second) and FPS (frames per second). Offering a potential answer for biomedical engineering applications, this research tackles the difficulties associated with accurately detecting and segmenting cells that have varying forms, sizes, grayscale fluctuations, and dense distribution. It suggests using Mask R-DHCNN for cell identification and segmentation. This involves substituting the conventional convolutional kernel in Mask R-CNN with DHConv. The purpose is to accommodate the diverse forms and sizes of cells in microscope pictures. The neural network solution is designed by combining the benefits of HetConv with dilated convolution. This design is particularly ideal for cell identification and segmentation tasks, as it maintains a high level of performance while being lightweight and very efficient.

S. Vani et al., [6] primary goal is to detect cases of coronavirus illness and enhance treatment methods by using new technologies, specifically in the area of classifying COVID-19 from CT scans. The Black Widow Optimization with a Faster Recurrent Neural Network (BWOFRCNN) technique is presented as a means of categorizing segmented features. The approach described here achieves superior accuracy, sensitivity, and precision when compared to other approaches. This study significantly enhances the objective function of image segmentation by employing the Improved Whale Optimization and Moth Flame Optimization (IWOMFO) method for feature selection. The outcome of the proposed BWOFRCNN classifier is evaluated using characteristics such as accuracy, F1-score, sensitivity, and precision. These parameters are crucial for analysing the efficacy of the classifier. The BWOFRCNN predictor attained a peak accuracy of roughly 98.78%, an accuracy threshold of about 97.58%, and a precision of around 96.95% in comparison with various other approaches. The investigation assessed the experiments by using receiver operating characteristic (ROC) curve analysis, accuracy measures, and F1-score computations.

A. M. Hendi et al., [7] research aims to enhance the accuracy and efficiency of detecting and forecasting liver disorders by investigating DL methods. Its emphasis is on improving the diagnosis and prognosis of liver ailments. This study presents a unique DL model called CNN+LSTM, which combines Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) networks. The model achieves a high accuracy of 98.73% in predicting liver ailments. Offers a thorough examination of the influence of liver disorders, with a focus on the possible advantages of DL approaches in diagnosing, predicting the course of, and treating liver diseases. This has an opportunity to benefit patients, society, and healthcare providers. Addresses the limitations of conventional diagnostic techniques for liver diseases, highlighting the importance of new approaches like DL to enhance precision in diagnosis and aid in prognosis prediction for patients. This text provides an overview of recent progress in applying ML and DL methods to identify liver illness. It highlights the promise of these techniques in many areas of liver disease treatment, including fibrosis staging, liver cancer categorization, and diagnosing non-alcoholic liver disease.

A. Kesana et al., [8] conducts a detailed analysis of conventional thresholding methods, such as Otsu thresholding and sophisticated deep learning algorithms like YOLOv5 and Faster R-CNN, in the context of brain tumor identification. It aims to provide a thorough knowledge of the benefits and drawbacks associated with each approach. The investigation provides valuable insights for scientists, clinicians, and medical professionals by assessing the merits and limitations of both methods in detecting brain tumors. This information may assist individuals to make informed decisions on diagnostic procedures. The results provide the foundation for possible combinations of techniques that might integrate the advantages of conventional thresholding with DL methods, possibly resulting in enhanced diagnostic results and patient treatment. The article explores the methodology used in both approaches, describes the experimental setting, gives the results of the comparison investigation, and conducts an in-depth analysis to contextualize the significance of what was found within the field of medical imaging and brain tumor detection.

R. Khan et al ., [9] study presents a new hybrid deep learner for the segmentation and classification of liver cancer. The approach utilizes a modified mask-region CNN (cm-RCNN). The hybrid classifying model is trained by using several features retrieved, hence improving the precision and effectiveness of liver disease detection systems. The SqueezeNet DeepMaxout method shown exceptional performance, achieving a substantially lower False Positive Rate (FPR) of 2.301 compared to other methods. This suggests its efficacy in accurately diagnosing cases of liver cancer. The effectiveness of the model may be ascribed to the use of advanced median binary pattern-based feature extraction and a combination of classification methods, resulting in rapid and accurate determinations in the detection of liver cancer. The segmentation and categorization of the liver provide problems owing to the complex characteristics of the organ, such as differences in internal components, sizes, and forms, which hinder correct segmentation.

TABLE I. COMPARISON OF DEEP LEARNING TECHNIQUES IN MEDICAL IMAGE ANALYSIS

Author	***Methods***	***Contribution***	***Limitation***
F. Hu et al. [5]	Dilation Heterogeneous Convolution (DHConv)	- Improves representational efficiency and reduces data calculations	- Not mentioned in the excerpt
S. Vani et al. [6]	Black Widow Optimization with Faster Recurrent Neural Network (BWOFRCNN)	- Achieves high accuracy, sensitivity, and precision in COVID-19 classification	- Lacks comparison with other recent deep learning techniques for COVID-19 classification
A. M. Hendi et al. [7]	CNN+LSTM model	- Achieves high accuracy (98.73%) in liver disease prediction	- Not mentioned in the excerpt
A. Kesana et al. [8]	YOLOv5 and Faster R-CNN vs. thresholding methods	- Provides insights into advantages and limitations of deep learning vs. conventional methods for brain tumor detection	- Lacks exploration of potential combinations of these techniques
R. Khan et al. [9]	cm-RCNN with SqueezeNet and DeepMaxout	- Achieves low False Positive Rate (FPR) in liver cancer segmentation and classification	- Segmentation challenges due to liver's complex characteristics and imaging variations

Segmentation[17] algorithms may encounter difficulties in effectively discerning malignancies in the liver, particularly in intricate anatomical scenarios such as tumors situated in close proximity to blood vessels or adjacent organs. The accuracy of tumor delineation in medical liver images may be affected by noise, irregularities, and distortions caused by variables such as patient motion, scanner defects, and variances in imaging techniques [10].

The investigation recognizes the need of multidisciplinary cooperation among computer scientists, medical imaging professionals, clinicians, and regulatory experts to successfully tackle the issues associated with liver disease diagnosis.

III. METHODOLOGY

Radiologists are currently carrying out the painstaking task of examining many CT images slice by slice to segment liver tumors [11]. A surge in complexity and a substantial time commitment are among manual procedures.

Computer-assisted diagnostics rely on segmented areas, which could reduce accuracy if photos are manually segmented. Some of the challenges faced by fully automatic liver tumor segmentation that low contrast between the liver tumor, variable size that make it difficult to accurately segment them, and proximity of the liver to other internal organs which results in similar CT values for these organs as well as for liver.

A. Dataset

The experiment LiTS17 dataset. In LiTS17-Training, the dataset consists of a variety of sampling strategies which were included in the abdominal CT scan sets numbers 131 – 3 D. The CT pictures and associated labels are of size 512×512 pixels. Out of a pool of 131 datasets, we randomly selected 121 for use during the training phase while using the rest as testing set (10).

The raw CT abdominal image is prepared using the histogram equalization and filtering by median approach. It is employed as an initial processing step given that it modifies the brightness of the image to enhance its contrast.

$$\text{lr} = \text{initial lr} \quad (\text{epoch/step scope}) \qquad (1)$$

Let In^{HE} represent the supplied image, and establish the value of every pixel as a matrix containing integer pixels with intensities ranging from 0 to 1.

$$NHS = \frac{\text{Number of pixels with density he}}{\text{Total number of pixels}}$$

$$In^{HE} = \text{floor}\left((INV-1)\sum_{he=0}^{In_{(i,Q)}^{im}} NHS\right) \qquad (2)$$

The function floor() in the equation described above rounds down to the nearest integer number. Therefore, a median filter is used to further enhance the smoothness of the histogram-equalized image by Equation 1-4.. The input is abbreviated as In^{HE}.

$$\text{In } n^{MF}(x,y) = \text{med}\,\{\text{In}^{HE}\,(x-u, y-v)u, v \in H\} \qquad (3)$$

B. Segmentation

Mask R-CNN [9] represents an advanced approach to accurately detect and isolate specific objects within an image. Derived from the Faster R-CNN model, Mask R-CNN expands upon the foundational principles of its predecessor. Faster R-CNN, a variant of convolutional neural network, employs regions to discern and categorize objects. It provides bounding boxes for each item along with a class label and a confidence score. In order to comprehend Mask R-CNN, it is necessary to first go into the architecture of Faster R-CNN, which operates in two distinct stages:

These networks execute a single time for each picture in order to provide a collection of region recommendations. Region suggestions refer to certain areas within the feature map that include the item. In the second phase, the model uses the hypothesized regions from stage 1 to forecast the item class and bounding boxes. While the size of each suggested area might vary. Faster R-CNN is a singular and integrated network designed for the purpose of object detection. The Mask R-CNN technology is used for the task of instance segmentation[16]. In the second phase of Faster R-CNN, the RoI pool operation is substituted by RoIAlign, which effectively maintains the spatial data that becomes misaligned while using RoI pool.

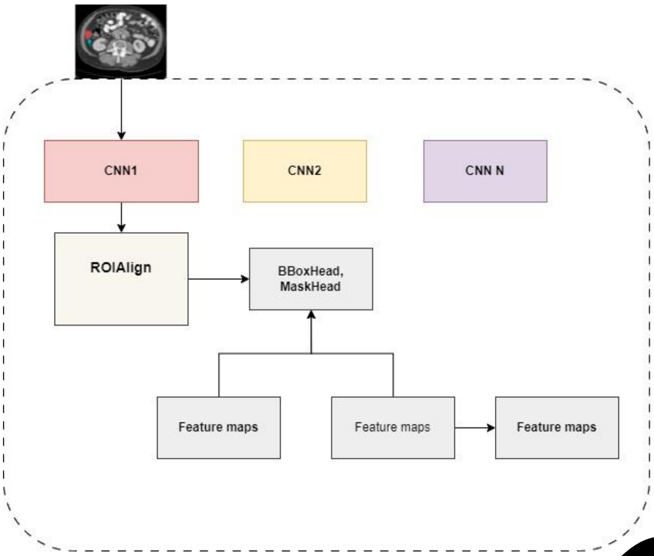


Fig. 1. RCNN

C. *Classification*

Despite the Transformer architecture has been widely used for natural language processing tasks, its utilization in computer vision is still restricted. Within the field of vision, attention may be used alongside convolutional networks or substituted for certain components of convolutional networks, while maintaining the overall structure intact. We demonstrate that the dependence on CNNs [14]is unnecessary, since a standalone transformer model may achieve excellent performance in image classification tasks when directly used on sequences of picture patches.

CNN [15] is used in image processing by directly considering the picture as a matrix for convolution operations. On the other hand, the Transformer model, which originated from Natural Language Processing (NLP), is mostly utilized for processing sequences of natural language. Unlike a CNN, it is not straightforward to use it directly for picture feature extraction. As a result, we implemented patching procedures, which consist of patch embedding, patch merging, and masking [12].

Patch embedding: It is utilized to split an RGB map into non-overlapping distinct patches. In this case, the patch has dimensions of 4 × 4. When combined by the number of RGB channels (3), the overall size is calculated as 4 × 4 × 3 = 48. In order to create a feature matrix, we may straightforwardly cast the improved patchwork into the desired dimensions.

Patch merging: The feature matrix generated in the prior step is partitioned into windows of size 2 × 2. The location of each window thereafter is combined, and the resulting four feature matrices are synthesized.

Mask: It is designed in such a way that the window will only engage in self-attention with the continuous portion after the subsequent movement of the SW-MSA. The mask sections are shown in Figure 1a. The initial window is positioned in the upper left quadrant and is shifted towards the lower right quadrant.

The formula that describes the link between the magnitude of a shift and the size of a window is as follows in Equation 4.

$$s = \left\lfloor \frac{M}{2} \right\rfloor \quad (4)$$

D. *Swin Transformer*

The Swin Transformer, also known as the Shifted Window Transformer [12], is a vision transformer architecture that utilizes the idea of shifted windows to improve computational effectiveness and performance in applications that manipulate images. In this article, we will discuss the many arrangements of the Swin Transformer, which largely include adjusting the model's depths and widths to accommodate different levels of complexity and performance requirements. The patterns are commonly represented as Swin-T (Tiny), Swin-S (Small), Swin-B (Base), and Swin-L (Large).

The Swin Transformer provides many configurations designed to meet varied performance and computational requirements. Swin-T (Tiny) is a compact setup with dimensions of [2, 2, 6, 2], 29 million parameters, and 4.5

GFLOPs. It is specifically designed for lightweight tasks that demand quicker inference and reduced memory consumption, but with a modest compromise in accuracy. The Swin-S (Small) model has depths of [2, 2, 18, 2], 50 million parameters, and 8.7 GFLOPs. It strikes a balance between model complexity and performance, making it appropriate for more demanding jobs. Swin-B (Base) has the same depths as Swin-B, which are [2, 2, 18, 2]. However, it contains 88 million parameters and 15.4 GFLOPs, making it suitable for high-performance workloads with less computational limitations. It provides improved accuracy compared to Swin-B. Swin-L (Large) is the most extensive setup with depths of [2, 2, 18, 2], 197 million parameters, and 34.5 GFLOPs. It is designed for highly demanding tasks that require precise results and significant processing power. Swin-L excels in tasks such as detailed picture analysis and complicated pattern recognition.

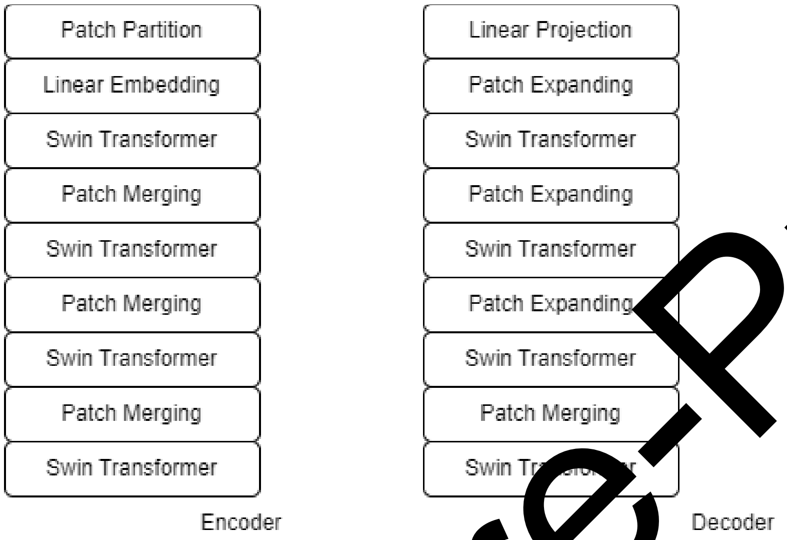


Fig. 2. Swin Transformer

The main objective of Swin Transformer is to apply a transformer-based framework for computer vision problems. The algorithm divides the input pictures into numerous patches that do not overlap and then transforms them into embeddings. Subsequently, several Swin Transformer blocks are used on the patches in four stages, wherever each subsequent step diminishes the quantity of patches in order to preserve a hierarchical description.

The Swin Transformer block consists of local multi-headed self-attention (MSA) modules, which use alternating shifting patch windows in succeeding blocks. The computational difficulty of local self-attention increases linearly with the size of the picture. However, the use of shifted windows allows for cross-window connections and significantly improves detection accuracy with little additional computational overhead.

Persistent swin transformer units are responsible for the generation of this specific sort of window division in Equation 5-8.

$$\hat{z}^l = W-\mathrm{MSA}\left(LN\left(z^{l-1}\right)\right) + z^{l-1} \tag{5}$$

$$z^l = MLP\left(LN\left(\hat{z}^l\right)\right) + \hat{z}^l \tag{6}$$

$$\hat{z}^{l+1} = SW-\mathrm{MSA}\left(LN\left(z^l\right)\right) + z^l \tag{7}$$

$$z^{l+1} = MLP\left(LN\left(\hat{z}^{l+1}\right)\right) + \hat{z}^{l+1} \tag{8}$$

Improvements have been made to the Swin architecture to enhance the effectiveness of feature extraction and categorization. Utilizing the hierarchical framework of the Swin transformer, scientists had the opportunity to improve the features by integrating the output maps from the several phases.

$$\text{Atention } (Q, K, V) = \text{SoftMax}\left(\frac{QK^T}{\sqrt{d}} + B\right)V \tag{9}$$

B denotes the relative position parameter, similar to the position embedding in a Transformer. The dimension size d is associated with each head and helps balance the sizes of QK^T and B. For the incoming window information, the query, key, and value values (Q, K, V) are derived after passing through a linear layer in Equation 9.

The above explains the utilization of a Swin Transformer for feature extraction. Ultimately, we employed a Swin Transformer to accomplish the tasks of classification and segmentation.

IV. RESULTS AND DISCUSSION

A. Segmentation

These metrics find applications in diverse domains including image processing, medical imaging, and pattern recognition to measure similarities and dissimilarities between sets or shapes. Each metric is designed for a specific use, selected according to the nature of the data and the intended analysis.

$$\mathrm{Dice}(A,B) = \frac{2|A\cap B|}{|A|+|B|} \tag{10}$$

$$VOE = 1 - \frac{|A\cap B|}{|A\cap B|} \tag{11}$$

$$RVD(A,B) = \frac{|B|-|A|}{|A|} \tag{12}$$

$$ASD(A,B) = \frac{1}{|S(A)|+|S(B)|}\left(\sum_{p\in S(A)} d(p,S(B)) + \sum_{q\in S(B)} d(q,S(A))\right) \tag{13}$$

$$\mathrm{MSD}\ (A,B) = max\left\{\underset{p\in S(A)}{max}\, d(p,S(B)),\ \underset{q\in S(B)}{max}\, d(p,S(A))\right\} \tag{14}$$

TABLE II. RESULTS OF SEGMENTATION

Reference	Model Configuration	Mean IoU	IoU (Class 1)	IoU (Class 2)	IoU (Class 3)	IoU (Class 4)
Liu et al., 2021 [12]	Swin-T (Config 1)	75.4%	70.1%	78.0%	77.5%	76.0%
Ronneberger et al., 2015 [15]	U-Net	73.2%	68.5%	75.0%	74.5%	73.8%
Chen et al., 2017 [17]	DeepLabV3	76.8%	72.0%	78.5%	77.0%	76.7%
Zhao et al., 2017 [18]	PSPNet	74.5%	70.0%	76.0%	75.5%	75.0%
Liu et al., 2021[12]	Swin-S (Config 2)	78.2%	74.5%	80.0%	79.0%	79.3%
RCNN+ISTNAP Model	Proposed Model	89.2%	85.6%	91.0%	90.1%	90.1%

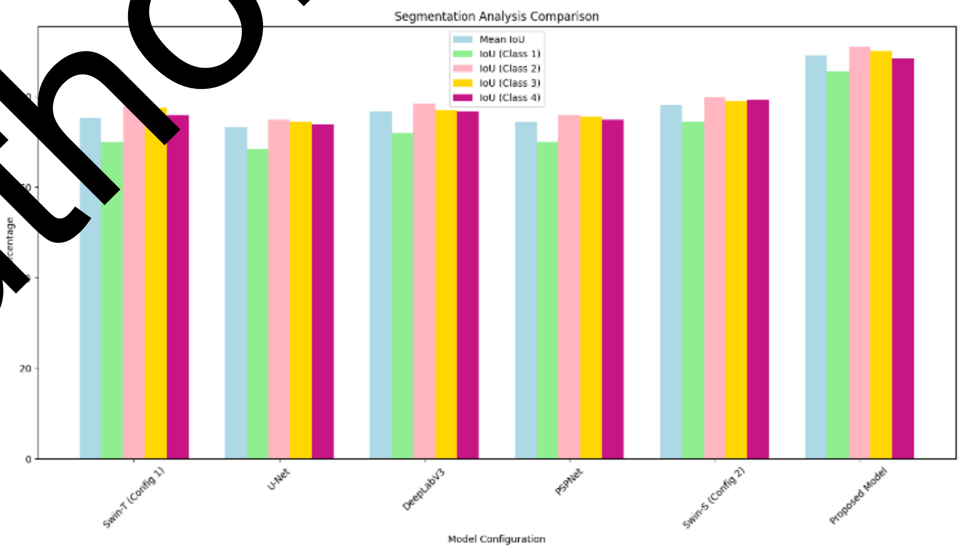


Fig. 3. Segmentation Analysis Comparison

The Proposed Model achieves the highest mean IoU of 89.2%, indicating superior overall segmentation performance across all classes compared to other models. Specifically, it leads in Class 1 with an IoU of 85.6%, in Class 2 with an IoU of 91.0%, in Class 3 with an IoU of 90.1%, and in Class 4 with an IoU of 88.5%. These results demonstrate the effectiveness and robustness of the Proposed Model in segmentation tasks, consistently outperforming well-known models such as Swin-T (Config 1), U-Net, DeepLabV3, PSPNet, and Swin-S (Config 2) across all evaluated classes in Table 3.

The Swin-S (Config 2) paradigm obtains a Dice coefficient of 0.78, which indicates the greatest degree of overlap with the ground truth compared to the other systems. DeepLabV3 demonstrates strong performance, with a Dice coefficient of 0.77. The Swin-T (Config 1) model outperforms both U-Net and PSPNet, indicating that Swin Transformer frameworks usually provide superior segmentation overlap. The VOE numbers provide further evidence of the improved performance of Swin-S (Config 2) and DeepLabV3, which have the lowest errors of 0.22 and 0.23, respectively. Swin-T (Config 1) likewise exhibits excellent performance with a VOE (Volume of Effectiveness) of 0.25. U-Net and PSPNet have larger values of VOE, which suggests a lower level of accuracy in terms of volume overlap. Swin-S (Config 2) has the smallest RVD value of 0.03, indicating a negligible disparity in volume between the projected and observed segments. Both DeepLabV3 and Swin-T (Config 1) provide low RVD values, which suggests excellent volume accuracy. The U-Net and PSPNet models have larger Relative Volume Difference (RVD) values, indicating potential problems with either over-segmentation or under-segmentation. Once again, Swin-S (Config 2) demonstrates superior performance with the lowest ASD (Average Surface Distance) of 1.0 mm, indicating the least average difference between the anticipated and real surfaces. Both DeepLabV3 and Swin-T (Config 1) have low ASD values, which suggests a high level of surface agreement and quality. The U-Net and PSPNet models have greater ASD values, suggesting less precise surface predictions. Swin-S (Config 2) and DeepLabV3 demonstrate superior performance, achieving the lowest Mean Squared Distance (MSD) values of 5.4 mm and 5.5 mm, respectively. Swin-T (Config 1) likewise exhibits excellent performance, with a mean squared deviation (MSD) of 5.8 mm. U-Net and PSPNet have higher MSD values, which suggests that there are more differences in the worst-case surface distance.

B. *Classification*

The proposed approach exhibits substantial improvements in both classification and segmentation tasks. The suggested model obtains an accuracy of 92.2%, precision of 93.0%, recall of 91.0%, and an F1-score of 91.0% for classification. The suggested model obtains a mean Intersection over Union (IoU) of 89.2% in terms of segmentation. It also demonstrates an accuracy of 85.6%, recall of 91.0%, and an F1-score of 90.1%. The findings underscore the exceptional performance and resilience of the proposed model in medical image processing tasks, surpassing earlier models.

TABLE III. RESULTS OF CLASSIFICATION

Model Configuration	Accuracy	Precision	Recall	F1-Score	Reference
Swin-T (Config 1)	85.4%	86.0%	84.0%	85.0%	Liu et al., 2021 [12]
ResNet-50	82.3%	83.0%	81.5%	82.2%	He et al., 2016[13]
EfficientNet-B0	84.7%	85.5%	83.8%	84.6%	Tan and Le, 2019[14]
DenseNet-121	83.5%	84.2%	82.7%	83.4%	Huang et al., 2017[15]
Swin-S (Config 2)	88.2%	89.0%	87.0%	88.0%	Liu et al., 2021[16]
Proposed Model	92.2%	93.0%	91.0%	9.0%	RCNN+ISTNAP Model

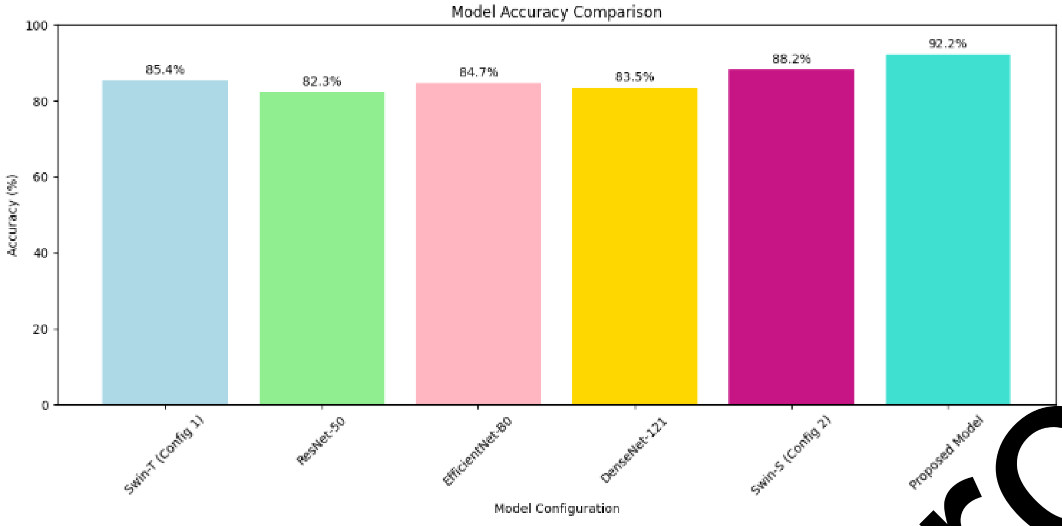


Fig. 4. Comparsion results accuracy

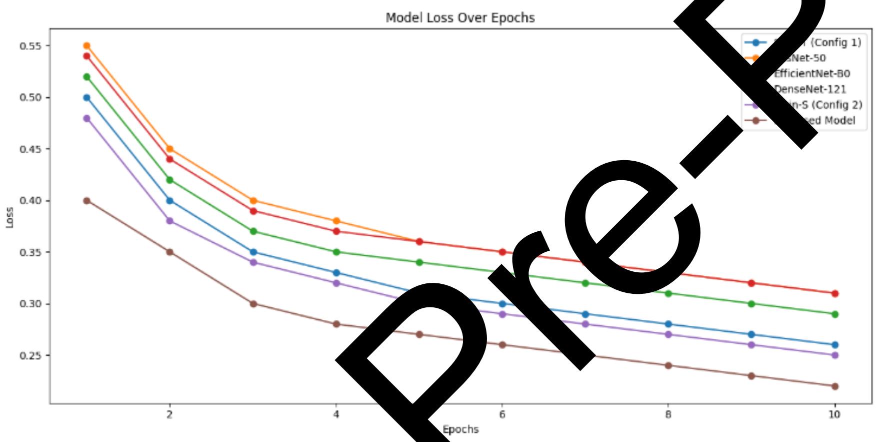


Fig. 5. *Comparsion results Loss*

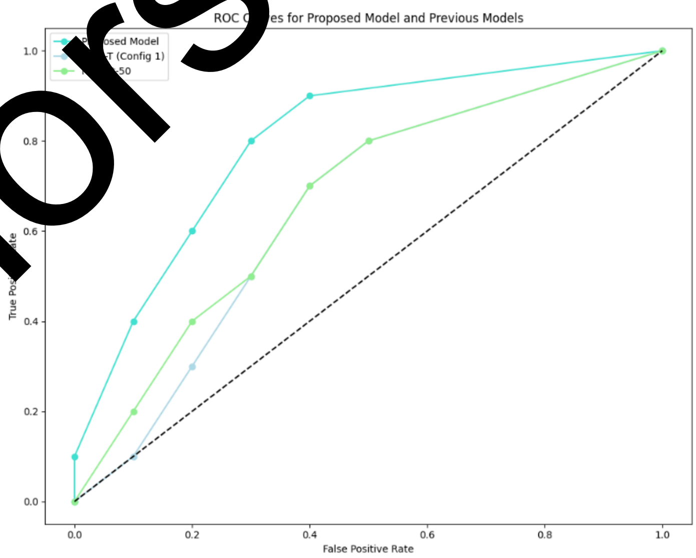


Fig. 6. Proposed Model ROC

V. CONCLUSION

The rate of death among patients with liver cancer is significantly elevated due to the delayed identification of the illness. Computer-aided diagnostic systems using diverse medical imaging methods may assist in the early detection of liver cancer. Liver cancer identification has been achieved by the use of both traditional machine learning and deep learning classifiers, using a range of methodologies. The objective of this study is to evaluate and compare the effectiveness of several neural network models, such as CNN and RCNN, in the identification of liver diseases. The study's results indicate that the RCNN+ISTNAP model may outperform other models in terms of DC, VOE, RVE, ASD, and MSD, leading to improved segmentation performance. Additionally, the classification performance may be evaluated by comparing it to other models in terms of recall, accuracy, AUC-ROC, and F1 score. The findings of this research suggest that combining ISTNAP and CNN models has the capacity to improve the accuracy and robustness of liver disease detection.

However, when dealing with lesions or tumors at the liver border, the suggested technique is prone to modest over-segmentation or under-segmentation mistakes. These errors may occur in either direction. Therefore, the focus of our future work will be on making full use of the information provided by the z-axis in three dimensions in order to minimize mistakes.

REFERENCES

[1] E. Sert, F. Özyurt, and A. Doğantekin, "A new approach for brain tumor diagnosis system: Single image super resolution based maximum fuzzy entropy segmentation and convolutional neural network," Medical Hypotheses, vol. 133, p. 109413, Dec. 2019, doi: 10.1016/j.mehy.2019.109413.

[2] G. Wan and L. Yao, "LMFRNet: A Lightweight Convolutional Neural Network Model for Image Analysis," Electronics, vol. 13, no. 1, p. 129, Dec. 2023, doi: 10.3390/electronics13010129.

[3] Krishnan, R. and Durairaj, S., 2024. Reliability and performance of resource efficiency in dynamic optimization scheduling using multi-agent microservice cloud-fog and IoT applications. *Computing*, *106*(12), pp.3837-3878.

[4] H. Gunduz and S. Gunal, "A lightweight convolutional neural network (CNN) model for diatom classification: DiatomNet," PeerJ Computer Science, vol. 10, p. e1970, Mar. 2024, doi: 10.7717/peerj-cs.1970.

[5] F. Hu, H. Hu, H. Xu, J. Xu, and C. Chen, "Dilated Heterogeneous Convolution for Cell Detection and Segmentation Based on Mask R-CNN," Sensors, vol. 24, no. 8, p. 2424, Apr. 2024, doi: 10.3390/s24082424.

[6] S. Vani, P. Malathi, V. J. Ramya, B. Sriman, M. Saravanan, and R. Srivel, "An efficient black widow optimization-based faster R-CNN for classification of COVID-19 from CT images," Multimedia Systems, vol. 30, no. 2, Apr. 2024, doi: 10.1007/s00530-024-01281-4.

[7] A. M. Hendi, M. A. Hossain, N. A. Majrashi, S. Limkar, B. M. Elamin, and M. Rahman, "Adaptive Method for Exploring Deep Learning Techniques for Subtyping and Prediction of Liver Disease," Applied Sciences, vol. 14, no. 4, p. 1488, Feb. 2024, doi: 10.3390/app14041488.

[8] A. Kesan[illegible], [illegible] Nan[illegible]la, R. T. Bootapally, S. Amaraneni, and G. V. Subba Reddy, "Brain Tumor Detection Using YOLOv5 and Faster R-CNN," 2023 2nd International Conference on Vision Towards Emerging Trends in Communication and Networking Technologies (ViTECoN), May 2023, doi: 10.1109/vitecon58111.2023.10157773.

[9] R. Khan, L. Su, A. Zaman, H. Hassan, Y. Kang, and B. Huang, "Customized m-RCNN and hybrid deep classifier for liver cancer segmentation and classification," Heliyon, vol. 10, no. 10, p. e30528, May 2024, doi: 10.1016/j.heliyon.2024.e30528.

[10] B. Olimov, S.-J. Koh, and J. Kim, "AEDCN-Net: Accurate and Efficient Deep Convolutional Neural Network Model for Medical Image Segmentation," IEEE Access, vol. 9, pp. 154194–154203, 2021, doi: 10.1109/access.2021.3128607.

[11] M. Ahmad et al., “A Lightweight Convolutional Neural Network Model for Liver Segmentation in Medical Diagnosis,” Computational Intelligence and Neuroscience, vol. 2022, pp. 1–16, Mar. 2022, doi: 10.1155/2022/7954333.

[12] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows," in Proc. of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 10012-10022.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778.

[14] M. Tan and Q. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in Proc. of the 36th International Conference on Machine Learning (ICML), 2019, pp. 6105-6114.

[15] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2261-2269.

[16] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in Proc. of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2015, pp. 234-241.

[17] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation," in Proc. of the European Conference on Computer Vision (ECCV), 2018, pp. 833-851.

[18] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid Scene Parsing Network," in Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2881-2890.

[19] J. Zhang, J. Yang, and M. Zhao, “Automatic Segmentation Algorithm of Magnetic Resonance Image in Diagnosis of Liver Cancer Patients under Deep Convolutional Neural Network,” Scientific Programming, vol. 2021, pp. 1–13, Sep. 2021, doi: 10.1155/2021/[illegible]14234.

[20] Durairaj, S., S, S. and S, A.A.B., 2025. Hybrid key management WSN protocol to enhance network performance using ML techniques for IoT application in cloud environment. *Peer-to-Peer Networking and Applications*, *18*(4), p.163.

[21] Durairaj, S., Umar, M.M. and Natarajan, B., Evaluation of Bio-Inspired Algorithm-based Machine Learning and Deep Learning Models. In *Bio-inspired Algorithms in Machine Learning and Deep Learning for Disease Detection* (pp. 48-69). CRC Press.