SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V13I7P111

Volume 13 Issue 7, 155-165, July 2026
© 2026 Seventh Sense Research Group®

Original Article

A Multi-Scale Residual Learning-Based Sequential Model for Hepatocellular Carcinoma Detection

M. Indumathi[1], M. Vanitha[2], V. Palanisamy[3]

[1,2,3]*Department of Computer Applications, Alagappa University, Tamil Nadu, India.*

[1]*Corresponding Author : mindumathi112@gmail.com*

Received: 30 April 2026 Revised: 06 June 2026 Accepted: 19 June 2026 Published: 29 July 2026

Abstract *- Accurate recognition of Hepatocellular Carcinoma (HCC) from medical images is very important to diagnose the disease early and come up with appropriate treatment plans. For instance, in this paper, an efficient HCC diagnosis model called MS-ResNet is designed by combining the proficiencies of deep learning. In particular, the proposed approach uses CT imaging in its process to detect HCC, but unlike other approaches, the new method can take advantage of multi-scale features to improve the overall efficiency of the diagnostic framework with the DL technique. It involves an improved preprocessing step where normalization and CLAHE techniques are used to enhance the quality of the input data. Further, in this work, multi-scale features have been extracted via an enhanced residual network. These extracted features have been further modeled through sequences using a Bi-LSTM approach and attention mechanisms. Subsequently, the predictive capability of the model in recognizing HCC is tested through the following evaluation metrics: accuracy, precision, recall, F1-score, and Dice score. From the experimental analysis of results presented in the work, the proposed method achieves an accuracy of 96.40%, precision of 95.70%, recall of 96.40%, F1-score of 94.90%, and dice score of 94.90%. These results exceed those obtained by other classifiers such as CNN, VGGNet, and ResNet, and are compared with the several recently developed liver tumor segmentation and classification methods.*

Keywords *- Hepatocellular Carcinoma, CT image analysis, Multi-Scale residual network, Bi-LSTM, Attention mechanism, Feature extraction.*

1. Introduction

Hepatocellular Carcinoma (HCC) is a frequently diagnosed primary liver malignancies, contributing significantly to global cancer mortality. Early and accurate diagnosis is essential for improving patient survival and guiding effective treatment planning [1]. Among various imaging modalities, Computed Tomography (CT) is widely used due to its ability to provide detailed anatomical and pathological information.

However, manual interpretation of CT images considerable amount of time to interpret the results manually, which creates a need for automated and reliable detection methods [2].

With the rapid growth in the availability of medical imaging data, the development of modern computer-aided diagnostic systems has gained significant momentum [3]. These systems support clinicians by enhancing detection accuracy while reducing the effort involved in manual analysis. As a result, automated image analysis has become an essential tool for enabling timely and consistent diagnosis in liver cancer detection [4].

Recent advancements in deep learning have shown considerable potential in the analysis of medical images, particularly in tumor detection and classification. Convolutional Neural Networks (CNNs) and residual networks have been widely adopted for extracting spatial features from CT images [5].

Despite their success, many existing approaches primarily focus on single-slice analysis, often overlooking the contextual relationships between adjacent slices. This limitation can affect the accurate identification of HCC, as tumor characteristics typically span across multiple slices and exhibit complex spatial continuity. [6]

To tackle the identified issues, this research develops a hybrid framework that integrates multi-scale residual learning with sequential modeling for improved HCC detection.

The proposed MS-ResNet model extracts rich multi-scale features and captures inter-slice dependencies using a Bi-LSTM architecture. An attention mechanism is further incorporated to emphasize the most relevant features, leading to more precise predictions.


This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

The principal advancements introduced in this work are as follows:

- A novel hybrid framework combining multi-scale residual feature extraction and sequential learning for HCC detection.
- Incorporation of an enhanced preprocessing stage using intensity normalization with CLAHE to improve contrast and feature representation for reliable learning.
- Integration of Bi-LSTM-based sequence modelling to effectively capture inter-slice contextual dependencies.
- Introduction of an attention model to increase the ability of feature differentiation and boost the performance of classification.
- Conducting thorough experiments by applying performance measures such as accuracy, precision, recall, F1-score, and Dice Score.
- Convincing proof that the proposed technique is superior to baseline classifiers, namely CNN, VGGNet, and ResNet, and is evaluated against existing literature for liver lesion identification and diagnosis.
- Concluding on the validation of the proposed technique via CT medical images for HCC detection.

The remaining part of the paper will be organized as follows. Section 2 will present the literature review on this topic. Section 3 will highlight the methodological framework used in the research. The findings and discussions will be presented in Section 4. The conclusion of this study is presented in Section 5, and Section 6 provides the study limitations and future research directions.

2. Related Works

This section examines existing studies on Hepatocellular Carcinoma (HCC) detection using deep learning techniques. Prior research mainly focuses on CNN-based models and medical imaging analysis for improved diagnosis. Methods such as multi-phase imaging, transfer learning, and attention mechanisms have shown significant progress. The review helps identify gaps that guide the proposed work.

In [7], a deep learning approach is introduced for non-invasive HCC mutation prediction from multi-phase CT images. Here, an attempt is made to leverage the complementary nature of features extracted from various contrast phases for improved tumor representation. Through cross-phase feature learning, the heterogeneity in tumors is well understood, along with enhanced predictive performance in terms of mutation prediction. Experimental results indicate that the proposed model outperforms both single-phase and conventional machine learning techniques, thus illustrating the potential of multi-phase CT image-based deep learning for accurate HCC characterization.

A deep learning-based framework for automatic segmentation of hepatocellular carcinomas from contrast-enhanced CT images is introduced in [8]. In this work, efforts have been made to improve tumor localization through accurate delineation of tumor borders utilizing advanced convolutional architectures. Due to the learned representations of complex spatial relationships and intensity variations in tumor regions, the segmentation performance has been improved by this method.

As per [9], there exists a multi-task deep learning framework that can be used in predicting early recurrence of hepatocellular carcinoma. Here, pre-training is employed to train multiple related tasks in one go. With this approach, several feature representations have been learned which are informative and generalized. This results in improved prediction of recurrence risk and better clinical decision-making processes.

According to [10], there exists a deep fusion framework that incorporates multi-phase CT scans combined with selected preoperative clinical parameters for early prediction of hepatocellular carcinoma recurrence. In this way, both imaging features and clinical features from different CT phases are integrated into one framework to improve the prediction capability.

A phase attention residual network has been designed in [11] for the segmentation of liver tumors from multi-phase CT images. The proposed network takes advantage of phase information in its design and uses attention mechanisms that help in considering the complementary information contained in various phases for better results. By combining residual learning with attention, the framework manages to achieve better results in terms of accurate liver tumor region segmentation.

An innovative deep learning approach for the early identification of hepatocellular carcinoma has been introduced in [12]. In the proposed approach, the researchers have made use of enhanced images and multi-phase CT scans. In this way, they were able to capture various features of the tumors and detect any cancerous changes present in the liver tissue.

In [13], a deep learning algorithm for detecting liver lesions with fused MR T1 in-phase, out-of-phase, and CT images has been developed. The proposed method applies the YOLOv8 neural network for fast and precise lesion detection using multi-modal medical imaging information, specifically MRI and CT. By using information about various modalities, the algorithm achieves high detection accuracy and can be used for liver lesion localization purposes.

In [14], a predictive CT-based model for MVI in patients with HCC has been developed. Using imaging-based features obtained during contrast-enhanced CT scanning, the authors propose an algorithm that predicts the presence of microscopic vascular invasion in HCC cases. As demonstrated by the

results, the proposed imaging-based algorithm can successfully predict the degree of tumor aggressiveness and clinical outcomes.

In [15], a deep learning approach to liver tumor segmentation with the use of a hybrid ResUNet architecture has been developed. The ResUNet model combines two approaches to segmentation, i.e., ResNet for deep learning and UNet for semantic segmentation. The combined algorithm improves image segmentation due to the ability to extract various features from CT images and combine them into a single feature map.

A densely connected hybrid U-Net model was developed to capture spatial information within and across CT slices, enabling effective identification of liver and tumor structures from volumetric scans, has been proposed in [16]. The proposed method is beneficial for liver tumor segmentation because it integrates dense connections with the U-Net structure. Integration of dense connectivity and the U-Net helps with efficient gradient flow and feature reuse. Therefore, it can help improve the efficiency of liver tumor segmentation.

A modified version of the U-Net architecture has been proposed for the segmentation of liver cancer from computed tomographic scans in [17]. A novel class balancing method has been applied to address the problem of data imbalance. With the introduction of the class balancing technique, learning of minority tumor classes was improved. Therefore, more precise identification and classification of liver and tumor parts could be achieved.

An automated deep learning-based approach has been proposed in [18] to accurately diagnose liver tumors through integration of magnetic resonance imaging data and clinical data. Integration of multiple sources of data makes this approach effective for liver tumor segmentation. Through the combination of imaging features and clinical data, liver tumors could be detected and classified in a more precise and clinically relevant manner.

In [19], an automated approach for detecting HCC from dynamic contrast-enhanced MRI through a deep learning technique has been designed. This framework makes use of the unique imaging properties, such as contrast enhancement and temporal characteristics, that can further assist in tumor identification. Using these techniques, a significant improvement in detection can be achieved. This paper shows the significance of using advanced imaging modality and deep learning techniques for accurate HCC detection.

In [20], a novel framework for detecting liver cancer using deep learning by utilizing Generative Adversarial Network (GAN), ResNet, and Vision Transformer has been developed. The GANs-based technique can help solve the problem of class imbalance and provide additional diversity to the training dataset. ResNets will be used to extract local features, while Vision Transformer will help capture global contextual information, which can be combined using an advanced fusion technique to achieve superior classification results.

In [21], a deep learning model that uses UNet++ to detect and segment liver tumors based on magnetic resonance images is suggested. It improves feature extraction by means of nested and dense skip connections, thus being able to extract not only local but also global features. By optimizing feature propagation within the network, it can provide precise segmentation of areas affected by the tumor. This paper demonstrates the potential of advanced versions of U-Net in the task of liver tumor segmentation.

According to previous research, the performance of deep learning models in HCC detection is rather high, yet there are several difficulties, including data variation and poor generalization. These problems make a solid foundation for this study.

3. Proposed Methodology

In this subsection, the Multi-Scale Residual Learning-Based Sequential Model (MS-ResNet) framework to detect hepatocellular carcinoma from CT images will be discussed. The MS-ResNet framework aims at utilizing both the fine-grained local features and the relationships between different slices. First, the input CT image data is preprocessed. Then, multi-scale features are extracted using the MS-ResNet algorithm. After that, the sequence is generated from the learned features while retaining spatial context. Features obtained in the previous step are used with Bi-LSTM, and the attention mechanism before classification is performed. The workflow of the proposed MS-ResNet model is illustrated in Figure 1.

3.1. Data Collection

For the current research, the authors employ a CT images database created for the LiTS (Liver Tumor Segmentation) Challenge, which is available on Kaggle. The image dataset comprises 130 contrast-enhanced CT images of the abdomen and is normalized to be in 256 × 256 pixels resolution in PNG format. Annotations for both liver and tumor lesions are provided in this dataset to allow the development of highly precise segmentation models. Since the database was collected from various medical facilities, its diversity cannot be underestimated.

3.2. Data Preprocessing

Preprocessing is an essential process during the analysis of medical images, thereby improving the integrity and usefulness of the data, minimizing variations, and making diagnostic features more visible. In the current research, the preprocessing stage will be conducted on the CT scan images before extracting the features. This includes the following steps.

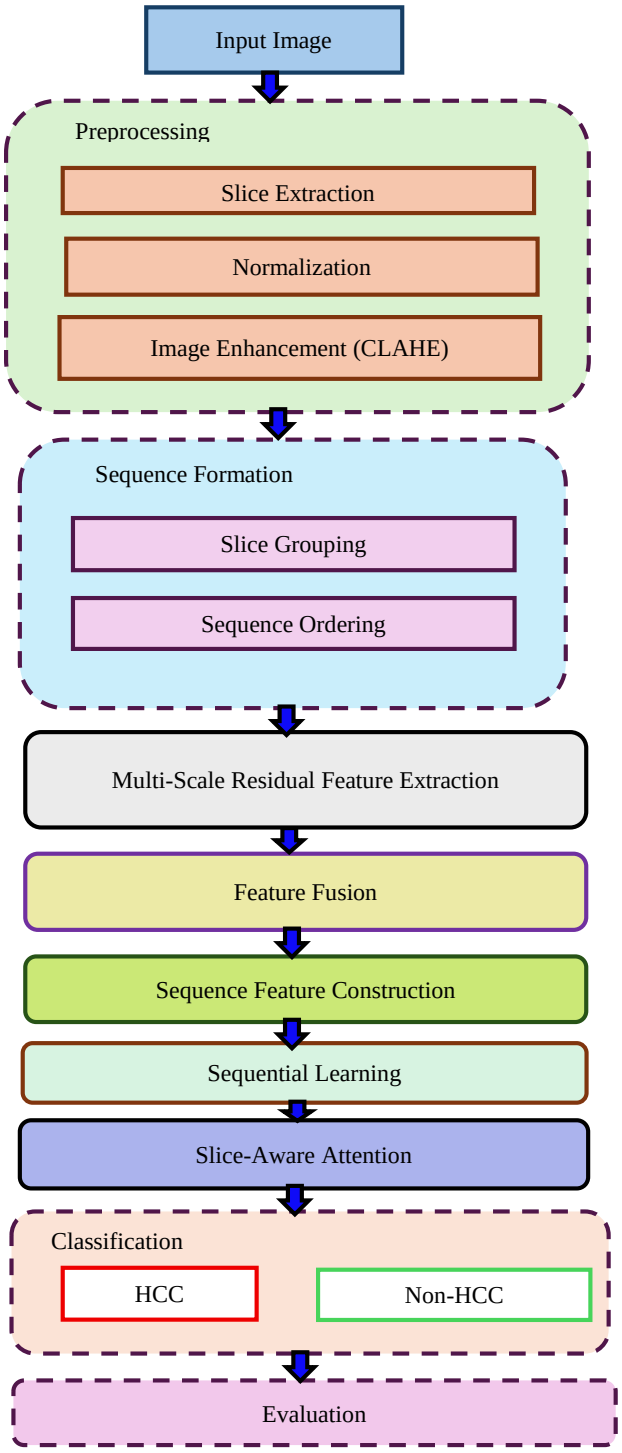

Fig. 1 The workflow of the proposed MS-ResNet model

3.2.1. Slice Extraction

CT scans are volumetric and can be represented as a 3D image volume $V \in \mathbb{R}^{H\times W\times D}$, correspond to the spatial dimensions of the volume, namely height, width, and slice count, respectively. The volume is decomposed into an ordered set of 2D slices presented in Equation (1):

$$S = \{I_1, I_2, I_3, \dots, I_D\} \quad (1)$$

where $I_i \in \mathbb{R}^{H\times W}$. It is important that the order of the slices is maintained in their original form.

3.2.2. Intensity Normalization

In order to minimize the variance of pixel intensities during each scan, normalization must be performed. This involves scaling the image I to a standardized range of values through min-max normalization:

$$I_{norm} = \frac{I - I_{\min}}{I_{\max} - I_{\min}} \quad (2)$$

where $I_{\min}$ and $I_{\max}$ represent the minimum and maximum intensity values in the image. This transformation ensures consistent intensity distribution and improves training stability.

3.2.3. Image Enhancement using CLAHE

In order to better highlight the fine details of certain anatomical structures, the image enhancement technique known as Contrast Limited Adaptive Histogram Equalization (CLAHE) is applied. CLAHE works differently from global histogram equalization by working on small image tiles in order to improve the local contrast of the image without introducing too much noise. This algorithm can work best in medical CT images where some tumors are visible at a low contrast when compared to other parts of the body. The contrast enhancement can be formulated as:

$$I_{enh}(x,y) = \text{CLAHE}\big(I_{norm}(x,y)\big) \quad (3)$$

where the amplification of contrast is controlled with a predefined clip limit in order to avoid increasing noise too much, through CLAHE, finer details can be better recognized, and tumor boundaries will become easier to discern. Therefore, it becomes much easier to learn the features of the image. The figure below shows the feature-improved image through the CLAHE filter.

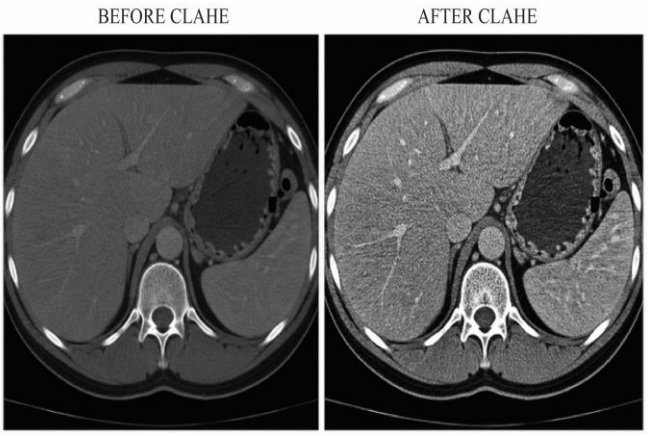

Fig. 2 Feature-enhanced image using CLAHE

3.3. Sequence Formation

After pre-processing, the generated slices are organized into sequences in order to facilitate the analysis of inter-slice dependencies in a sequential manner.

3.3.1. Slice Grouping

The ordered slice set $S = \{I_1, I_2, \dots, I_D\}$ is split into subsequences of fixed length via sliding window technique. For any particular window length L, the sequence is defined as follows:

$$\mathcal{X}^{(k)} = \{I_k, I_{k+1}, \dots, I_{k+L-1}\}, k = 1,2, \dots, D - L + 1 \quad (4)$$

This grouping enables the model to have access to local volumetric information, which in turn allows the identification of differences between anatomical structures in neighboring slices.

3.3.2. Sequence Ordering

Within each sequence $\mathcal{X}^{(k)}$, the slices are maintained in their original anatomical order:

$$I_k \to I_{k+1} \to \cdots \to I_{k+L-1} \quad (5)$$

Preserving this order is essential for modeling spatial continuity along the scan depth. It enables the subsequent sequential model to learn forward and backward dependencies, capturing the progression and consistency of tumor regions across neighboring slices.

3.4. Multi-Scale Residual Feature Extraction

For obtaining discriminative features from CT slices, a deep residual learning architecture is used. The objective of this phase is to extract spatial features at various scales for efficiently representing fine-texture and high-level semantic information related to hepatocellular carcinoma.

3.4.1. Residual Feature Learning

Each input slice I_i fed into the Residual Network (ResNet) to learn features at various levels. The concept of residual learning helps train deep neural networks, as it involves learning of the residual mapping given as:

$$y = F(x) + x \quad (6)$$

where x denotes the input feature map, $F(x)$ represents the learned residual function, and y is the output. This formulation helps preserve low-level information while enabling deeper feature abstraction.

3.4.2. Multi-Scale Feature Extraction

The tumor properties at various scales can be captured through feature extraction by applying the convolution operation on multiple layers of the residual network. The feature maps generated from various stages can be denoted by:

$$\{F_1, F_2, F_3, \dots, F_m\} \quad (7)$$

where each F_j represents the feature map corresponding to a unique scale of representation that can vary from low-level features to high-level semantic features.

3.4.3. Feature Representation

For each slice I_i, the multi-scale features extracted are fused together to create a full feature vector:

$$f_i = \mathcal{G}(F_1, F_2, \dots, F_m) \quad (8)$$

where $\mathcal{G}(\cdot)$ is a function for fusion, such as concatenation. This feature f_i contains all the necessary information about the input slice.

3.5. Feature Fusion

The features at multiple scales that have been obtained from different layers of the residual network are fused to represent a single entity for each slice. Suppose the set of feature maps is $\{F_1, F_2, \dots, F_m\}$, with each F_j representing one particular scale. Then, the following fusion process takes place using weighted summation:

$$F_{\text{fused}} = \sum_{j=1}^{m} w_j F_j \quad (9)$$

where $\|$ stands for concatenation and w_j the corresponding weight factor for each scale of the feature map. The fused feature map is subsequently transformed into a compact feature vector. $f_i = \phi(F_{\text{fused}})$using a pooling or flattening operation. The resulting feature vector serves as a comprehensive representation of features present in the input slice.

3.6. Sequential Feature Construction

After feature fusion, the feature vectors extracted at the slice level are grouped together systematically to facilitate sequential learning. For any sequence length L, the fused features will be grouped as follows:

$$\mathcal{X} = \{f_1, f_2, f_3, \dots, f_L\} \quad (10)$$

where each $f_i$represents the feature vector corresponding to the i-th slice. It must be noted that this arrangement maintains strict correspondence to the anatomical sequence of the slices, thus making the independent slices' features into a sequence.

3.7. Sequential Learning

In order to establish any dependencies between the adjacent slices, the derived feature sequence undergoes processing by a Bidirectional Long Short-Term Memory (Bi-LSTM) model. By processing information in both forward and

backward directions, the Bi-LSTM effectively captures contextual relationships and improves the interpretation of changes between adjacent slices. This is done as follows:

$$\overrightarrow{h_t} = \text{LSTM}\left(f_t, \overrightarrow{h_{t-1}}\right), \overleftarrow{h_t} = \text{LSTM}\left(f_t, \overleftarrow{h_{t+1}}\right) \quad (11)$$

and combined as:

$$h_t = \left[\overrightarrow{h_t} \parallel \overleftarrow{h_t}\right] \quad (12)$$

This bidirectional representation produces a context-aware sequence embedding, which is then utilized for further refinement and classification.

3.8. Slice-Aware Attention

To emphasize informative slices from the sequence, a slicing-specific attention model is used on the output of the Bi-LSTM. Given the sequence of hidden representations $\{h_1, h_2, \ldots, h_L\}$, Attention weights are computed to quantify the importance of each slice:

$$\alpha_t = \frac{\exp(W h_t)}{\sum_{k=1}^{L} \exp(W\, h_k)} \quad (13)$$

and used to obtain a context vector:

$$c = \sum_{t=1}^{L} \alpha_t\, h_t \quad (14)$$

This ensures that more significance is given to relevant slices, providing a better representation of the data.

3.9. Classification

After the attention mechanism generates the context vector c, it is forwarded to a fully connected layer for prediction. The final output is obtained by applying a sigmoid function to convert the network response into a probability value. It can be represented as:

$$\hat{y} = \sigma(W_c\, c + b) \quad (15)$$

where W_c and b are learnable parameters and $\hat{y}$ the predicted value. Depending on the threshold, the result can be HCC or non-HCC. Figure 3 illustrates the results of Hepatocellular Carcinoma (HCC) detection.

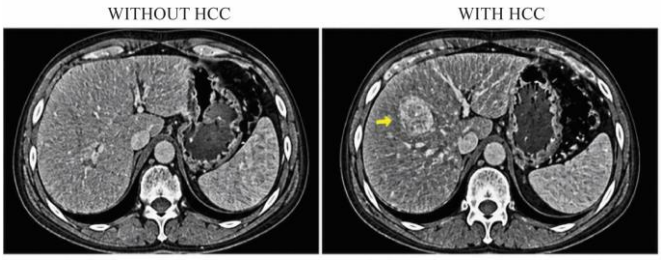

Fig. 3 Sample results of Hepatocellular Carcinoma (HCC) detection

3.10. Evaluation

Model performance will be assessed via common classification measures such as accuracy, precision, recall, F1-measure, and dice score. These performance measures guarantee that the proposed algorithm's performance will be properly evaluated based on its capability of classifying HCC, assessing the reliability of predictions, and determining the correspondence between the predicted areas and the manually annotated regions.

Algorithm Steps of the Proposed Model

Input: CT volume $V \in \mathbb{R}^{H\times W\times D}$
Output: Predicted label $\hat{y}$

Begin
1. Extract slices
   $S = \{I_1, I_2, \ldots, I_D\}$
2. Preprocessing
   for each $I_i \in S$do
   $I_i \leftarrow \text{Normalize}(I_i)$
   $I_i \leftarrow \text{CLAHE}(I_i)$
   end for
3. Form sequences of length L:
   $\mathcal{X} = \{I_1, I_2, \ldots, I_L\}$
4. for each slice $I_i \in \mathcal{X}$do
   Extract multi-scale features:
   $\{F_1^i, F_2^i, F_3^i\} \leftarrow \text{ResNet}(I_i)$
   Fuse features:
   $F_{\text{fused}}^i \leftarrow [F_1^i \parallel F_2^i \parallel F_3^i]$
   Compute feature vector:
   $f_i \leftarrow \text{GAP}(F_{\text{fused}}^i)$
   end for
5. Construct feature sequence
   $\mathcal{F} = \{f_1, f_2, \ldots, f_L\}$
6. Sequential learning
   $H = \text{BiLSTM}(\mathcal{F})$
7. Compute attention
   $c = \sum \alpha_t h_t$
8. Classification
   $\hat{y} = \sigma(Wc + b)$

End
Return $\hat{y}$

4. Results and Discussion

4.1. Experimental Setup

For this study, a CT medical imaging dataset has been used, which was created based on LiTS17 (Liver Tumor Segmentation) Challenge and was downloaded from Kaggle with 130 images that were contrast-enhanced abdominal CT scans [22]. All the images were preprocessed and resized to 256×256 and saved in PNG format. The liver and tumors have been labeled in these images. This is a diverse dataset since it was created from multiple clinical sites. The dataset consists of 80% of training images and 20% of test images. It has been trained in Python (version 3.8). For training, binary cross-

entropy was used along with the Adam optimization algorithm. This research has been executed on a computer with an Intel i7 CPU, 16GB RAM, and a NVIDIA GPU.

4.2. Comparative Analysis with Baseline Classifiers

A performance comparison was carried out to assess the capability of the proposed MS-ResNet model against baseline classifiers, namely CNN, VGGNet, and ResNet. These classifiers were independently implemented and trained using the same dataset, preprocessing procedures, and experimental settings as the proposed model to ensure a fair comparison. The results exhibit that the proposed MS-ResNet model outperforms the baseline models across all evaluation metrics, including accuracy, precision, recall, F1-score, and dice score.

4.2.1. Accuracy

Accuracy represents the proportion of instances that are correctly identified out of all evaluated samples. The accuracy offers a measure of the overall performance of the classifier, and it is often employed in comparing multiple classifiers under similar circumstances.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{16}$$

In this context, TP and TN refer to the true positive and true negative classification results, while FP and FN represent incorrectly predicted data. A higher value of accuracy implies good prediction performance. Figure 4 presents the difference in performance between different models on the grounds of accuracy. The accuracy values achieved by the CNN model and VGGNet model were 88.20% and 90.10%, respectively. On the other hand, ResNet produced an accuracy of 92.30%. It may be clearly seen that with increasing depth, the accuracy improves significantly. The MS-ResNet produced the highest accuracy of 96.40%, thereby exhibiting a superior model performance.

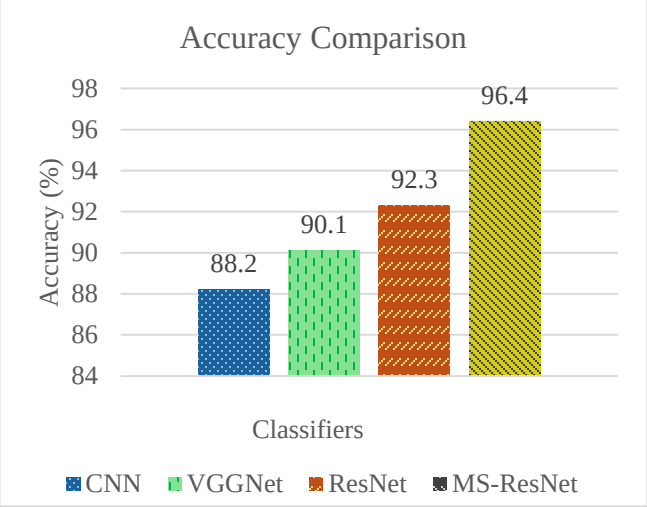

Fig. 4 Accuracy comparison of the proposed model with existing approaches

4.2.2. Precision

Precision evaluates the accuracy with which a model predicts positive examples through a measure of how many true positives there are among all the predicted positives. Precision gives an indication of how well a model can avoid false positives.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{17}$$

A high precision measure implies that the model performs very well when it comes to identifying relevant positive classes without any errors. The performance comparison results of the suggested MS-ResNet classifier with other classifiers can be seen in Figure 5. The precision comparison reveals that precision values have improved steadily with the increase in models, ranging from 87.10% for CNN, 89.40% for VGGNet, to 91.80% for ResNet. The suggested MS-ResNet model has a precision value of 95.70%, making it perform best when it comes to identifying positive cases.

4.2.3. Recall

Recall is an indicator of the accuracy of the model in determining all real positive cases from the data set. Recall represents the model's capability in recognizing relevant data without missing out on any positive data.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{18}$$

The higher the value of recall, the more accurate the model in detecting the real positive cases with few missed detections. The recall values show a similar pattern and begin at 86.50% for CNN, followed by 88.70% for VGGNet and 92.30% for ResNet, as seen in Figure 5. The proposed model attains the highest recall value of 96.40%. This shows the efficiency of the model in recognizing real cases of HCC.

4.2.4. F1 score

F1-Score is a balanced performance score metric in which the precision and recall metrics are balanced in a single evaluation process. In situations where classes do not have equal numbers, it helps balance the ratio of false positives and false negatives.

$$\text{F1-score} = \frac{2 \times (Precision \times Recall)}{Precision + Recall} \tag{19}$$

The higher the F1-Score in a performance score, the better the performance balance achieved by the model in relation to the number of correct predictions and incorrect predictions. An F1-Score, which balances the precision and recall, supports the robustness of the proposed model as depicted. Whereas CNN, VGGNet, and ResNet have F1-Scores of 85.80%, 89.00% and 93.20% respectively, MS-ResNet achieves a slightly higher percentage of 94.80%.

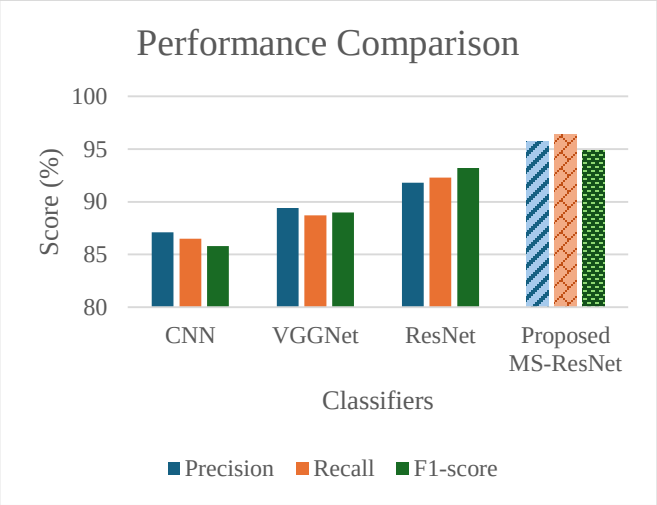

Fig. 5 Performance Comparison of the proposed MS-ResNet Model with baseline models

4.2.5. Dice Score

The dice similarity coefficient is a statistical measure for evaluating the amount of overlap between the prediction map and the target regions. This measure is frequently employed in biomedical image analysis to measure the accuracy of the target region segmentation.

$$\text{Dice Score} = \frac{2TP}{2TP + FP + FN} \qquad (20)$$

The more similar the predicted segmentation area is compared to the actual ground truth, the higher is the Dice score value. From Figure 6 below, one can conclude that the proposed method shows the best performance with the Dice score equal to 94.90%. In descending order are the scores obtained for CNN, VGGNet, and ResNet methods equal to 85.90%, 88.20%, and 90.70%, respectively.

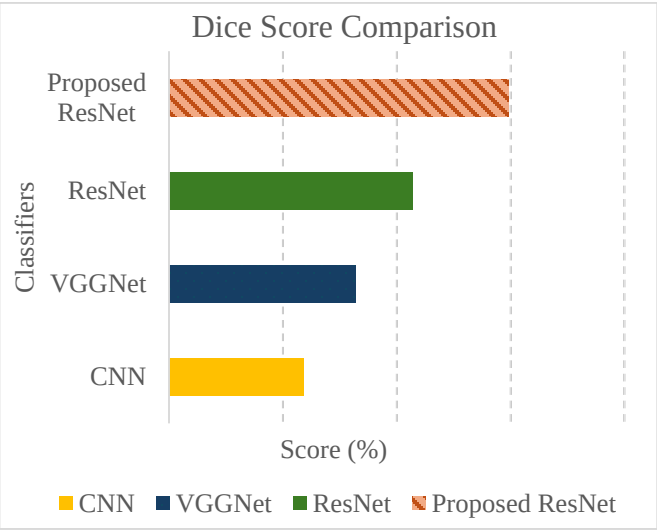

Fig. 6 Dice score comparison of the proposed MS-ResNet model with baseline models

Table 1. Comparative analysis of the proposed MS-ResNet model with baseline models

Metrics (%)	Classifiers			
	CNN	VGGNet	ResNet	Proposed MS-ResNet
Accuracy	88.20	90.10	92.30	96.40
Precision	87.10	89.40	91.80	95.70
Recall	86.50	88.70	92.30	96.40
F1-score	85.80	88.00	93.20	94.90
Dice Score	85.90	88.20	90.70	94.90

Table 1 presents the comparative performance of the proposed MS-ResNet model against existing approaches. The results clearly indicate that the proposed model achieves superior performance across all evaluated metrics that demonstrate its effectiveness over the baseline methods.

4.3. Comparison with State-of-the-Art Methods

To provide a more comprehensive evaluation of the proposed MS-ResNet model, its performance was compared with several recent state-of-the-art approaches reported for liver tumor analysis using the LiTS17 dataset. This comparison provides a broader evaluation of the proposed model and highlights its capability relative to advanced deep learning models developed for liver tumor segmentation and classification. The results presented in Tables 2 and 3 summarize the comparative evaluation based on dice score, accuracy, precision, recall, and F1-score. The comparison demonstrates the effectiveness of the proposed approach in accurately identifying tumor characteristics from CT images while maintaining robustness across multiple evaluation measures.

Table 2. Comparison of the proposed MS-ResNet model with State-of-the-Art methods based on the dice score

S.No	Model	Dataset	Dice score (%)
1	CAFCT-NET [23]	LiTS17	84.29
2	FasNet [24]	LiTS17	87.66
3	RHEU-Net [25]	LiTS17	70.19
4	PGC-Net [26]	LiTS17	73.63
		3DIRCA Db	74.16
5	PAKS-Net [27]	LiTS17	76.9
6	MS Liver Tumor Segmentation Algorithm [28]	LiTS17	74.8
7	DynTransNet [29]	ATLAS	86.12
		LiTS17	93.12
8	**Proposed MS-ResNet**	**LiTS17**	**94.90**

The findings reported in Table 2 highlight the strong performance of the proposed MS-ResNet model in liver tumor analysis. The model achieved a dice score of 94.90%,

outperforming several recent methods, including CAFCT-Net, FasNet, RHEU-Net, PGC-Net, PAKS-Net, the MS Liver Tumor Segmentation Algorithm, and DynTransNet. The enhanced performance of the proposed model is primarily due to the combined use of multi-scale feature learning and residual connections. This design allows the network to effectively capture both fine-grained tumor characteristics and contextual information from CT images. Consequently, the model is able to identify tumor regions more precisely, which results in better agreement between the predicted outputs and the corresponding reference annotations.

Table 3. Performance comparison of the proposed MS-ResNet model with State-of-the-Art liver tumor classification methods

Model	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
En-DeNet [30]	LiTS17	93.13	93.65	88.94	91.42
FLAS-UNet++ [31]	LiTS17	82.4	-	-	-
Hybrid Deep Ensemble Model [32]	LiTS17	95.4	94.3	93.9	94.1
Proposed MS-ResNet	**LiTS17**	**96.40**	**95.70**	**96.40**	**94.90**

Table 3 compares the proposed MS-ResNet model with recent state-of-the-art classification methods on the LiTS17 dataset. The proposed model attained an accuracy of 96.40%, precision of 95.70%, recall of 96.40%, and F1-score of 94.90% which demonstrates strong performance across all evaluation metrics. Compared with En-DeNet, FLAS-UNet++ and the Hybrid Deep Ensemble Model, MS-ResNet achieved better overall classification results. This improvement is mainly due to its multi-scale feature learning and residual connections, which help capture important tumor characteristics from CT images more effectively. The results highlight the capability of the proposed model to provide accurate and reliable liver tumor classification.

5. Conclusion

This study proposes a hybrid deep learning architecture, termed MS-ResNet, for accurate detection of HCC in CT images. Unlike conventional approaches that focus solely on spatial feature extraction, the proposed model adopts a multi-stage learning strategy, where multi-scale features are first extracted using an enhanced residual network and then modeled sequentially to capture inter-slice dependencies. Additionally, the use of slice-level attention in the current approach helps in highlighting informative features used in classification. Results obtained under identical experimental conditions show that the proposed model performs better than baseline classifiers like CNN, VGGNet, and ResNet across all evaluation metrics, such as accuracy, precision, recall, F1-score, and dice score. It is evident that the proposed model, which combines multiple scales of feature learning together with sequential modeling, has significantly contributed to improved performance through better tumor pattern identification and spatial correlation among slices. Specifically, the higher Dice scores obtained for all datasets indicate the superiority of the model regarding the localization of tumor areas. The future of research should be directed towards increasing the capacity of the architecture through its application on a large CT database, along with considering lighter architectures for improved computational capabilities.

Study Limitations and Future Research Directions

Despite the promising results obtained from the application of MS-ResNet on HCC detection, some limitations need to be discussed. The limitation of the model is its testing on an incomplete dataset of CT scans, which may result in reduced accuracy. Moreover, the sequence-based architecture employed by the authors may be affected by different image acquisition protocols, including slice thickness and other conditions. Also, the deep learning architecture applied in the study requires significant computational resources.

The future work could involve expanding the current framework to include multiple datasets to increase the model's robustness. Also, an extension to a 3D architecture may lead to increased accuracy. Employing lightweight models may decrease computation time. Segmentations and integration of clinical information could further enhance the current framework.

References

[1] Sanghyeok Lim, YiRang Shin, and Young Han Lee, "Arterial Enhancing Local Tumor Progression Detection on CT Images using Convolutional Neural Network after Hepatocellular Carcinoma Ablation: A Preliminary Study," *Scientific Reports*, vol. 12, pp. 1-8, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[2] Sarthak Grover, and Surbhi Gupta, "Automated Diagnosis and Classification of Liver Cancers using Deep Learning Techniques: A Systematic Review," *Discover Applied Sciences*, vol. 6, pp. 1-32, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[3] Shanmugapriya Survarachakan et al., "Deep Learning for Image-Based Liver Analysis — A Comprehensive Review Focusing on Malignant Lesions," *Artificial Intelligence in Medicine*, vol. 130, pp. 1-23, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[4] Mohammadreza Elhaie et al., "Deep Learning for Hepatocellular Carcinoma Segmentation in MRI: A Systematic Review of Models, Performance, and Challenges," *Medicine*, vol. 104, no. 51, pp. 1-12, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[5] Samy A Azer, "Deep Learning with Convolutional Neural Networks for Identification of Liver Masses and Hepatocellular Carcinoma: A Systematic Review," *World Journal of Gastrointestinal Oncology*, vol. 11, no. 12, pp. 1218-1230, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[6] Chaopeng Wu et al., "A Review of Deep Learning Approaches for Multimodal Image Segmentation of Liver Cancer," *Journal of Applied Clinical Medical Physics*, vol. 25, no. 12, pp. 1-22, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[7] Jiapan Gu et al., "Multi-Phase Cross-modal Learning for Noninvasive Gene Mutation Prediction in Hepatocellular Carcinoma," *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Montreal, QC, Canada, pp. 5814-5817, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[8] Ilaria Manghi et al., "Deep Learning for Automatic Segmentation of Hepatocellular Carcinoma in Contrast Enhanced CT Scans," *Scientific Reports*, vol. 15, pp. 1-8, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[9] Jian Song et al., "Early Recurrence Prediction of Hepatocellular Carcinoma Using Deep Learning Frameworks with Multi-Task Pre-Training," *Information*, vol. 15, no. 8, pp. 1-14, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[10] Weibin Wang et al., "Deep Fusion Models of Multi-Phase CT and Selected Clinical Data for Preoperative Prediction of Early Recurrence in Hepatocellular Carcinoma," *IEEE Access*, vol. 8, pp. 139212-139220, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[11] Yingying Xu et al., "PA-ResSeg: A Phase Attention Residual Network for Liver Tumor Segmentation from Multiphase CT Images," *Medical Physics*, vol. 48, pp. 3752-3766, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[12] Shih-Wei Guo et al., "Efficiently Using Deep Learning to Distinguish Early-Stage Hepatocellular Carcinoma (HCC) from Non-HCC Based on Multi-Phase CT and Image Enhancement," *Proceedings of the 2024 7th International Conference on Artificial Intelligence and Pattern Recognition*, Xiamen China, pp. 537-546, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[13] Rhugved Bhojane et al., "Liver Lesion Detection from MR T1 In-Phase and Out-Phase Fused Images and CT Images Using YOLOv8," *International Conference Proceedings of Data Science and Applications*, pp. 121-135, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[14] Qi Yue et al., "Contrast-Enhanced CT Findings-based Model to Predict MVI in Patients with Hepatocellular Carcinoma," *BMC Gastroenterology*, vol. 22, pp. 1-11, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[15] Hameedur Rahman et al., "A Deep Learning Approach for Liver and Tumor Segmentation in CT Images using ResUNet," *Bioengineering*, vol. 9, no. 8, pp. 1-19, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[16] Xiaomeng Li et al., "H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation from CT Volumes," *IEEE Transactions on Medical Imaging*, vol. 37, no. 12, pp. 2663-2674, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[17] Yodit Abebe Ayalew, Kinde Anlay Fante, and Mohammed Aliy Mohammed, "Modified U-Net for Liver Cancer Segmentation from Computed Tomography Images with a New Class Balancing Method," *BMC Biomedical Engineering*, vol. 3, pp. 1-13, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[18] Shi-hui Zhen et al., "Deep Learning for Accurate Diagnosis of Liver Tumor Based on Magnetic Resonance Imaging and Clinical Data," *Frontiers in Oncology*, vol. 10, pp. 1-14, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[19] Killian Monnin et al., "Deep Learning for Automatic Detection of Hepatocellular Carcinoma in Dynamic Contrast-Enhanced MRI," *Abdominal Radiology*, vol. 51, pp. 2843-2856, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[20] Shivani Joshi et al., "Enhancing Liver Cancer Detection: An Innovative Deep Learning Approach Combining GAN, ResNet, and Vision Transformer," *Expert Systems with Applications*, vol. 298, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[21] Jing Wang et al., "A Deep-Learning Approach for Segmentation of Liver Tumors in Magnetic Resonance Imaging using UNet++," *BMC Cancer*, vol. 23, pp. 1-8, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[22] Liver and Liver Tumor Segmentation, Kaggle. [Online]. Available: https://www.kaggle.com/datasets/andrewmvd/lits-png

[23] Ming Kang et al., "Cafct-Net: A Cnn-Transformer Hybrid Network with Contextual and Attentional Feature Fusion for Liver Tumor Segmentation," *2024 IEEE International Conference on Image Processing (ICIP)*, Abu Dhabi, United Arab Emirates, pp. 2970-2974, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[24] Rahul Singh et al., "FasNet: A Hybrid Deep Learning Model with Attention Mechanisms and Uncertainty Estimation for Liver Tumor Segmentation on LiTS17," *Scientific Reports*, vol. 15, pp. 1-19, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[25] Liyan Sun et al., "A Multi-Scale Liver Tumor Segmentation Method Based on Residual and Hybrid Attention Enhanced Network with Contextual Integration," *Sensors*, vol. 24, no. 17, pp. 1-19, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[26] Yilin You et al., "Contour-Induced Parallel Graph Reasoning for Liver Tumor Segmentation," *Biomedical Signal Processing and Control*, vol. 92, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[27] Jiezhou He et al., "Towards Accurate Abdominal Tumor Segmentation: A 2D Model with Position-Aware and Key Slice Feature Sharing," *Computers in Biology and Medicine*, vol. 179, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[28] Lifang Chen, and Shiyong Luo, "Multi-Scale Liver Tumor Segmentation Algorithm by Fusing Convolution and Transformer," *Journal of Computer Engineering & Applications*, vol. 60, no. 4, 2024. [Google Scholar]

[29] Siming Zheng et al., "DynTransNet: Dynamic Transformer Network with Multi-Scale Attention for Liver Cancer Segmentation," *Frontiers in Oncology* , vol. 15, pp. 1-14, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[30] S. Suganeshwari et al., "En–DeNet Based Segmentation and Gradational Modular Network Classification for Liver Cancer Diagnosis," *Biomedicines*, vol. 11, no. 5, pp. 1-23, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[31] Quchen Peng et al., "Liver Tumor Segmentation and Classification using FLAS-UNet++ and an Improved DenseNet," *Technology and Health Care*, vol. 30, no. 6, pp. 1475-1487, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[32] B. Sravani, and M. Sunil Kumar, "A Hybrid Deep Ensemble Model for Precise Liver and Tumor Segmentation Using U-Net and W-Net Architectures," *Journal of Electronics Electromedical Engineering Medical Informatics*, vol. 8, no. 2, pp. 553-571, 2026. [CrossRef] [Google Scholar] [Publisher Link]