



Dense Depthwise Separable Convolution with Multi-Encoder Network for Liver Tumor Segmentation

Pallavi Babu^{1*} Meesala Shobha Rani¹

¹*School of Computer Science and Engineering, REVA University, Bengaluru, India*

* Corresponding author's Email: pallavi8805@gmail.com

Abstract: Liver tumor segmentation from Computed Tomography (CT) images plays a significant role in diagnosis and treatment planning. However, automatic segmentation techniques face challenges because of irregularity, fuzzy boundaries and heterogeneity of tumor tissues. This research proposes Dense Depthwise Separable Convolution with Multi-Encoder Network (DDSC-MENet) for liver tumor segmentation. The MENet is developed in this research with down-sampling structure for liver tumor segmentation, which contains a whole down-sampling path with four encoders along with the same structures. The decoder connections in MENet allow the model to obtain feature representations from multiple encoder branches and perform integration of multi-scale and multi-modal information. The DDSC-MENet enhances the capacity of segmentation of tumors having different shape, sizes or intensity to the existing regular model. The Convolution Neural Network with Multiple Parametric Exponential Linear Units (CNN-MPELU) is applied in the classification stage for classifying liver tumors, which helps to resolve the vanishing gradient issues and quickens the convergence. The DDSC-MENet achieves dice of 94.58% and 92.62% for LiTS2017 and 3DIRCADB datasets respectively, which is better than Multi-Scale Feature Attention Network (MS-FANet).

Keywords: Computed tomography, Convolution neural network, Dense depthwise separable convolution, Liver tumor segmentation, Multi-encoder network, Multiple parametric Exponential linear unit.

1. Introduction

Liver is a large and highly structured granular organ of humans, and it plays a crucial role in different metabolic functions. However, modern lifestyle factors have exceptionally led to cases of liver disorders like cirrhosis, acute and chronic viral hepatitis and fatty liver [1]. It is important to detect liver cancer in its early stages because this greatly increases the choice of treatment. However, due to the limited availability of accurate physical examinations for liver cancer, imaging and radiology remains as essential tools [2]. Liver segmentation with the identification of associated tumors is a significant step in treatment planning such as ablation or Selective Internal Radiation Therapy (SIRT) [3]. However, manual segmentation is time-consuming and this has made the treatment planning to become more complex [4]. In regular clinical tasks, especially in recognition of hepatic tumor, CT has gained many

advantages in clinical application because of its high signal-to-noise ratio, high elliptical image resolution, fast scanning speed and low cost and so on [5]. CT image segmentation of liver tumors must be done accurately because it supplies important information for the differentiation of the more detailed tumor features like the shape or the size of the lesion and the exact location of the tumor which assist the clinicians in performing a better therapy plan [6, 7]. However, segmentation of liver tumor is challenging, and it is considered a crucial aspect. The irregular shapes and sizes of the liver tumors with low contrast with surrounding healthy tissues and the difficulty in accurately distinguishing tumor boundaries present significant segmentation challenges [8].

Many computational methods have been developed in the existing research to improve diagnostic performance, but these computer aided systems are not performing optimally when dealing with the challenging task of liver lesion segmentation

and detection [9]. Several factors make this process complex such as the absence or low differentiation of the liver from the surrounding structures, high contrast between the liver tissue and tumors, differing number of tumors and their size, irregular growth and mutual arrangement of tumors, and the presence of very small tumors [10-12]. These challenges have been overcome by Deep Learning (DL) models which have been demonstrated to be useful. Recent works from existing research show better outcomes when using DL methods, especially CNN-based models employed for the segmentation of liver tumor from CT scans. The gray scale and texture-based image processing techniques were applied to automatically segment liver tumors [13]. However, the CNN was shown better performance in a wide range of computer vision applications and its bias assumption limits model performance in learning dependencies to local perceptual fields which loss capturing of long-range feature association possibilities [14]. It is not enough to adjust image inputs with various sizes, textures and shapes which leads to data loss and model variability [15]. The contributions are summarized as follows:

- The decoder connections in MENet enable the model to obtain feature representations from multiple encoder branches and perform integration of multi-scale and multi-modal information.
- This proposed DDSC-MENet enhances the capacity of segmentation of tumors having different shapes, sizes or intensities to the existing regular model.
- The CNN-MPELU is applied for classifying liver tumor which helps to resolve the vanishing gradient issues and quickens the convergence.

This research paper is organized in the subsequent manner: Section 2 describes literature review and Section 3 details proposed methodology with the process. Section 4 generates results with discussion and Section 5 concludes the manuscript with future direction.

2. Literature review

The recent research related to liver tumor segmentation are analyzed in this research with advantages and limitations.

Zhang [16] developed a U-shaped network with Scale-Axis-Attention (SAA-Net) for liver tumor segmentation. The scale attention fused receptive fields from numerous scales that were helpful for multi-scale target segmentation. The axis attention models the spatial long-range dependencies

effectively. The scale and axis attention were integrated with adaptive global pooling named as SAA. Moreover, the model was better than self-attention with computational resource utilization but it integrated scale and spatial attention mechanisms for enhancing performance. The SAA-Net improves liver tumor segmentation by efficiently concentrating on relevant features at multiple scales. However, it struggled with tumor segmentation because of less contrast against liver tissue surroundings thereby reducing segmentation performance.

Chen [17] suggested a MS-FANet for liver tumor segmentation. The Residual Attention (RA) block and Multi-scale Atrous Down-sampling (MAD) are developed in the encoder part of MS-FANet to acquire sufficient variable features and the tumor features at various ranges. Moreover, Dual-path Feature (DF) filter and Dense Up-sampling (DU) were developed in feature reduction for reducing efficient features for better liver tumor segmentation. The MS-FANet improves segmentation accuracy by focusing on significant tumor regions among various scales thereby capturing both global and fine-grained information efficiently. However, it struggled with small and irregular tumor segmentation for accurate boundary detection.

Wang [18] introduced a Multi-Scale-Aware and Twin-Split Attention module (MSA-TSA) for liver tumor segmentation. The MSA link semantic gap and minimizes detailed data loss. Moreover, TSA recalibrates feature map channel response according to segmentation results from a 3D view for classifying tumors. The skip connection with MSA was developed to clean spliced features from the encoding phase for reducing single-scale feature detail and overcoming semantic gaps. The MSA-TSA improves segmentation by efficiently capturing context data at various scales which enables to recognize tumors accurately in different shapes and sizes. However, it was complex to interpret attention weights that complicated the decision-making process thereby affecting performance.

Liu [19] presented a Spatial and Spectral-learning Double-branched Aggregation Network (S2DA-Net) for liver tumor segmentation. The Fourier Spectral-learning Multi-scale Fusion (FSMF) employed a Fourier-based network to acquire phase and amplitude data which captures feature and detained data without numerous parameters. The Multi-axis Aggregation Hadamard Attention (MAHA) integrated spatial data enhancing discriminative features when reducing computational cost. In Group Multi-Head Cross-Attention Aggregation (GMCA) decoder path extracts local data and long-term dependencies which enhances the localization

abilities through combining features from various breaches. However, S2DA-Net struggled with tumor variability and anatomical structures among various patients that affect generalizability.

Di [20] implemented a 3D U-Net for automatic liver tumor segmentation from CT images. The liver region was separated into homogeneous super-pixels by applying LI-SLIC based hierarchical interactive approach in that super-pixels are decomposed based on intensity standard deviation to tumor boundaries accurately. Every pixel in liver regions is classified into tumor and non-tumor through Support Vector Machine (SVM). Lastly, voting based model was developed to identify tumour regions from super-pixels based on pixel-wise classification results. The 3D U-Net captures spatial information which enables to apply context relationship among 3D thereby improving performance. However, it struggled with overfitting due to noise which affects the segmentation performance.

Tiange Zhang et al. [21] suggested an Edge-guided multi-scale adaptive feature fusion network (MAEG-Net) for liver tumor segmentation. The multi-scale adaptive feature fusion efficiently integrates the multi-scale data to guide tumor segmentation in various sizes. To report blurred image boundaries, the edge-aware guidance was used to enhance its learning capability. The feature fusion model improves the channel partition selection to fuse features adaptively from every encoder layer. The MAEG-Net enhances the performance through conserving boundary details and utilizing multi-scale contextual data by adaptive feature fusion. However, it struggled with high heterogeneous tumor generalization because of overfitting to edge features.

In summary, the existing methods have limitations such as less contrast against liver tissue surroundings, struggle with small and irregular tumor segmentation for accurate boundary detection and overfitting due to noise. Moreover, complex interpreting attention weights complicated the decision-making process, and struggled with tumor variability and anatomical structures among various patients that affect generalizability. To overcome these limitations, this research proposes a DDSC-MENet for liver tumor segmentation.

3. Proposed method

The LiTS2017 and 3DIRCADB datasets are applied in this research and are preprocessed by wiener filter and data augmentation. The Wiener filter is used to remove noise and artefacts in liver images which enhances the reliability and quality of the image. The data augmentation is applied by replacing some training images with new images which enhances the training dataset to see a huge variety of liver tumors. Then, the preprocessed images are segmented by using DDSC-MENet which achieves better segmentation by factorizing spatial and channel-wise convolution. The segmented images are provided in the CNN-MPELU classifier which classifies the liver tumor. Fig. 1 depicts the process of the proposed methodology.

3.1 Dataset

Two publicly available liver tumor segmentation datasets such as LiTS2017 and 3DIRCADB datasets are applied in this research, The detailed explanation of these two datasets is given in the following:

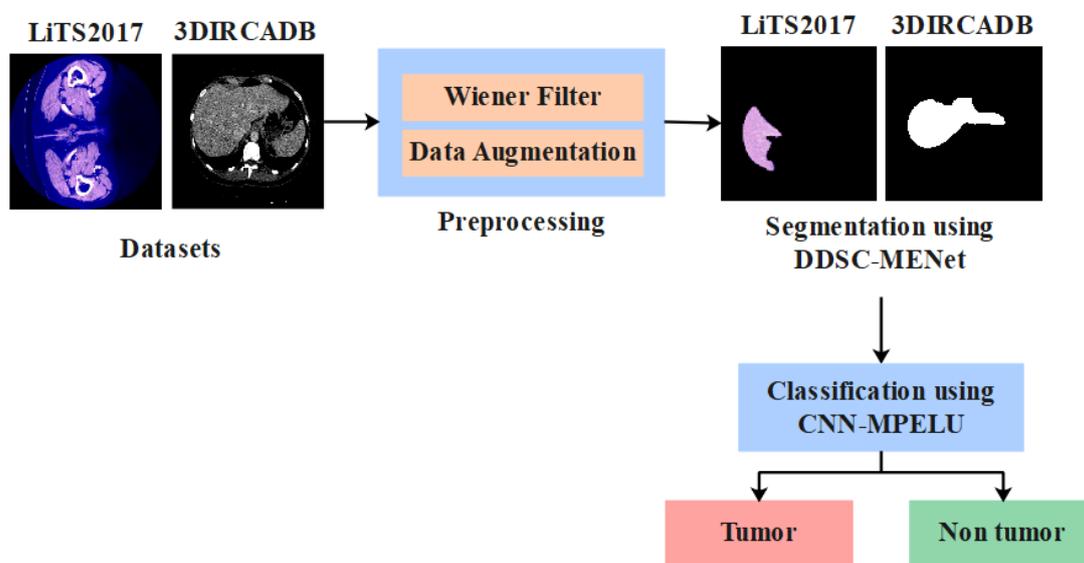


Figure. 1 Process of proposed methodology

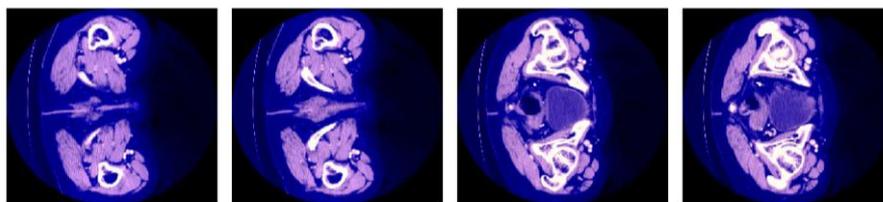


Figure. 2 Sample images for LiTS2017 dataset

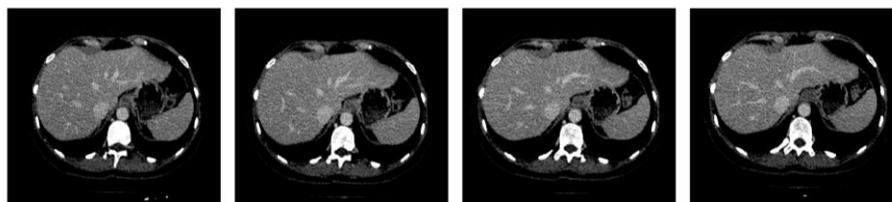


Figure. 3 Sample images for 3DIRCADB dataset

LiTS2017 [22] is one of the largest and most representative datasets for liver tumor segmentation including 201 CT scans (131 labelled, 70 unlabeled) from six medical centers. Randomly, the 131 labelled CT scans were divided into a training set of 111 and a validation set of 20. The 3DIRCADB dataset [23] is commonly used in current works as a training dataset. This dataset is composed of 3D CT scans of 10 women and 10 men with hepatic tumors in 75% of cases. Fig. 2 and Fig. 3 depict the sample images for LiTS2017 and 3DIRCADB datasets.

3.2 Preprocessing

The liver image is preprocessed by using a wiener filter and data augmentation to remove noise and increase the data size. The detailed explanation of these two preprocessing is given in the following:

3.2.1. Wiener filter

Wiener filter is applied in this research to remove noise and artefacts in liver images, which enhances the reliability and quality of the image [24]. It is a crucial step for preprocessing and aims to eliminate noise when preserving significant information in liver images. The wiener filter enhances the Signal-To-Noise Ratio (SNR) of medical images which makes it easier to differentiate healthy and abnormal liver tumor regions. The wiener filter is adaptive, and it adjusts itself based on the local mean and variance of the image which allows to selection of smooth noisy areas when maintaining edges and textures for identifying tumor boundaries. This adaptability is better than other filtering methods, such as Gaussian and median filters which blur the significant details and do not effectively handle noises present in the input images. Moreover, the wiener filter removes Gaussian and salt-pepper noise from the input image, thereby leading to better segmentation performance.

3.2.2. Data augmentation

The data augmentation is applied by replacing some training images with new images, which enhances the training dataset to see a huge variety of liver tumors [25]. The following techniques such as random contrast, random brightness, random scale, gaussian noise, coloring, cropping, horizontal flipping and rotation are applied to enhance the data size. Fig. 4 and Fig. 5 depicts the augmented images for LiTS2017 and 3DIRCADB datasets.

3.3 Segmentation

The proposed DDSC-MENet is used for liver tumor segmentation which achieves better segmentation by factorizing spatial and channel-wise convolution. This makes creating dense connections between layers essential within the DDSC as it enhances feature propagation and reuse throughout the network. This structure helps to maintain fine details which is important for distinguishing and accurately identifying liver tumor regions.

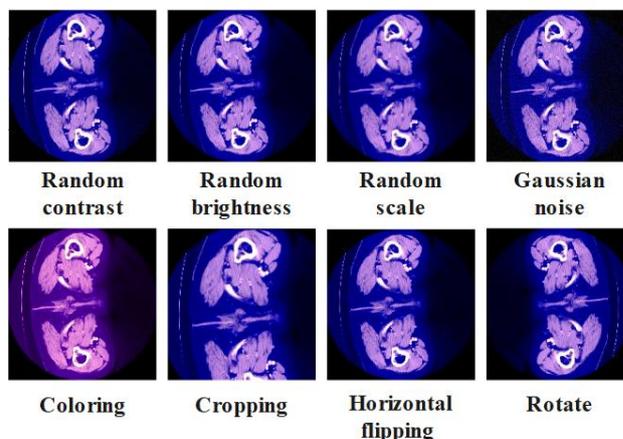


Figure. 4 Augmented images for LiTS2017 dataset

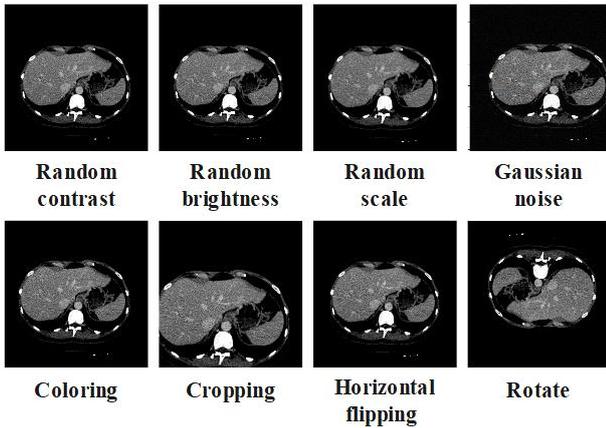


Figure. 5 Augmented images for 3DIRCADB dataset

3.3.1 Down-sampling

The MENet is developed in this research with down-sampling structure for liver tumor segmentation in which the whole down-sampling route is comprised of four encoders with the same structure. The input of these four encoders is based on four modalities and the skip-connection is used to integrate model features extracted through every encoder. Compared to conventional V-Net uses one encoder to extract features from four modalities, the complexity of extracting image feature through encoder minimized the encoder to extract features from a particular model. This one-to-one feature extraction enhances capability of down-sampling procedure. The encoder in V-Net architecture is designed to compress and structure the image data effectively. In encoder, every encoder block contains 1-3 DDSC layers and down-sampling layers. The convolution formula is given in Eqs. (1) and (2),

$$i_s = i + (s - 1)(i - 1) \quad (1)$$

$$o = \left\lceil \frac{i_s + 2p - k}{s} + 1 \right\rceil \quad (2)$$

Where, i is the highest integer that does not exceed the range. Here, DDSC with step-size of 2 is used to exchange pooling layer. This process doubles channels and has solution when maintaining computational complexity. The convolutional kernel size in encoder part is $3 \times 3 \times 3$ and DDSC layer utilizes batch normalization, Rectified Leaky Unit (ReLU) functions and the final layer uses the sigmoid function. Moreover, ResNet short-circuiting is incorporated within every encoder block to prevent gradient disappearing because of excessive deep network structure.

3.3.2 Up-sampling

The architecture of a decoder is same as encoder and every up-sampling stage contains 2-3 DDSC layers with an up-sampling layer. It is mostly involved in the reconstruction of the image information. At up-sampling, deconvolution is performed with a stride of 2 to raise the resolution of the pre-trained image as shown in Eq. (3).

$$o = s(i - 1) + 2p - k + 2 \quad (3)$$

In the decoder part of the network, feature map size is doubled while number of channels is minimized to half. Additionally, each stage of decoder integrates features from the corresponding down-sampling stage. For the final down-sampling layer, the $1 \times 1 \times 1$ convolution kernel is adopted to set number of output channels equivalent to classes. Lastly, SoftMax layer normalizes the values of each channel to provide pixel-per-pixel segmentation. Furthermore, ResNet-style shortcut connections are used in each of the decoder blocks for better performance.

3.3.3 Fusion strategy

The feature map fusion process in MENet occurs primarily in two ways: First, they integrate feature maps from respective stages of four encoders at the encoder-reconstruction level. Second, they join feature maps from equivalent stages of encoder-decoder. When down-sampling, this research employs skip connections to fuse modal features obtained through every encoder. At each down-sampling stage, the feature map output is sent to a similar stage in parallel encoder rough skip connections while at the same time being input to the next down-sampling stage. Finally, the required four-modal fusions are attained after several fusion rounds in the down-sampling stages of the chosen architecture. In each of these down-sampling stages, these fused maps integrate the feature information of all four modes.

Subsequently, a $1 \times 1 \times 1$ convolution is utilized to feature maps generated after fusing multiple stages to change the number of channels. The network integrates numerous skip connections throughout the network to capture information that is not lost during four compression paths.

The MENet enhances final contour prediction but also reduces the time it takes for the model to converge. To recover the information not captured by the encoder, connections between the encoder and

decoder are used which add the corresponding feature maps from the two processes.

This enables the decoder to retrieve necessary information for up-sampling by map. By incorporating higher resolution data, the model increases precision by recovering detailed features from the original image. The reconstruction of a cross-layer feature using the encoder and decoder visualizes the potential for coupling feature maps from downsampling with new features from upsampling. As the network deepens detailed information from the feature map is retained which addresses the loss of details. This process assists in the retention of vital feature information hence enhance the segmentation achievements.

3.3.4 Dense Depthwise Separable Convolution

The DDSC network is inspired through U-Net is fully convolutional with feature connections between corresponding stages. It contains of an encoder part followed by a decoder part with skip connections between these two. The encoder receives the image input and extracts high semantic features of image, while decoder reconstructs high semantic features into actual image size. The skip connections are employed to connect the multi-scale features at the encoder and decoder part. Unlike U-Net, the DDSC replaces most layers of standard convolution with DDSC layers, leading to a commendably low computational complexity. This makes it possible to develop an advanced network with the ability to learn feature information at a compound level. Furthermore, for transferring the contextual feature information, this research establishes the remaining skip connections among encoder and decoder. The DDSC network is composed of three main components: DDSC blocks, subsampling layers and up-sampling layers. The DDSC block includes five densely connected layers that contain the batch

normalization layer, ReLU activation function and DDSC with a size of 3×3 . The subsampling layer is a max-pooling layer by kernel size and stride of 2 while the use of an up-sampling layer is a 3×3 transferred convolution layer. For traditional convolution, output feature map F is calculated by Eq. (4), when assuming stride and padding of one.

$$F_{k,l,n} = \sum_{i,j,m} K_{i,j,m,n} \cdot I_{k+i-1+j-1,m} \tag{4}$$

Where, I is an input feature map, K is a convolution kernel size. While DDSC is composed of depthwise and pointwise convolutions. The output feature map F for depthwise separable convolution is calculated as Eq. (5). Comparing parameters of DDSC with standard convolution is attained by Eq. (6).

$$F'_{k,l,n} = \sum_{i,j} K'_{i,j,m} \cdot I_{k+i-1+j-1,m} \tag{5}$$

$$\frac{k \times k \times M + M \times N}{k \times k \times M \times N} = \frac{1}{N} + \frac{1}{k^2} \tag{6}$$

The DDSC uses less parameters than standard convolution which allows the network to acquire contextual data. The segmented images are given as input to the classification process for classifying liver tumors. Fig. 6 shows the systematic diagram of DDSC-MENet.

3.4 Classification

The generalized CNN architecture primarily consists of three key layers such input convolution layer, hidden convolution layer and output convolution layer. Liver images are fed into the input layer as data and the output features are derived from the output layer [26]. The hidden layer located between the input and output layers is crucial in the convolution process with neurons acting as its fundamental components.

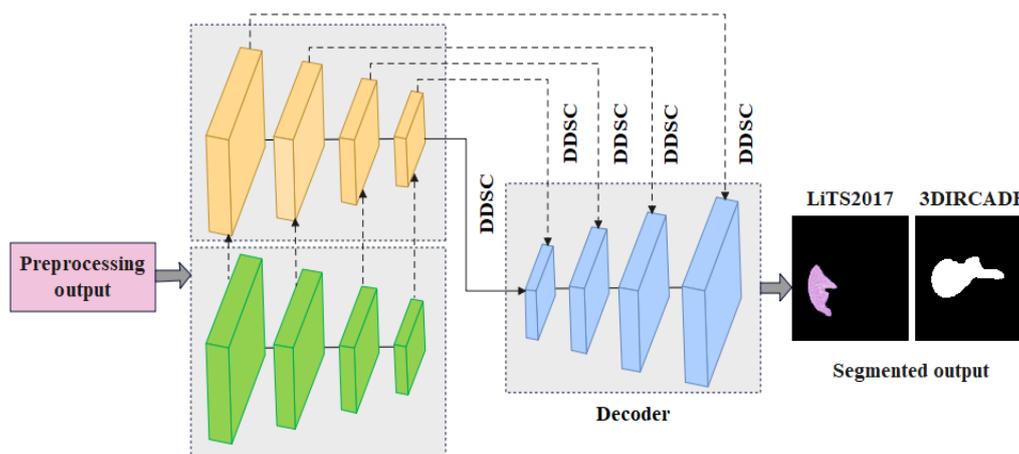


Figure. 6 Systematic diagram of DDSC-MENet

Each neuron in this network has unique learning weights and biases. The number of inputs received by each neuron corresponds to the number of inputs from other neurons and a weighted sum is computed. The result is then processed through a ReLU activation function. Convolution layers apply filters to the input images which generate feature maps.

In deeper models, convolution layers are closer to the input for learning lower-level features such as edges and lines. Layers further down the model learn higher-level features such as shapes and specific objects. The layers get more information as it combines to produce complex patterns. The output neuron is given into the convolution layer as shown in Eq. (7).

$$z_{i,j,k} = b_k + \sum_{u=0}^{f_h-1} \sum_{v=0}^{f_w-1} \sum_{k'}^{f_{n'}-1} x_{i',j',k'} \cdot W_{u,v,k',k} \quad (7)$$

Where, $z_{i,j,k}$ is a neuron output at row i , column j and feature map k of convolution layer l , f_h and f_w are height and width of relevant fields, $f_{n'}$ is the number of features inside the layer. The $x_{i',j',k'}$ is a neuron output at row i , column j and feature map k of previous convolution layer $l - 1$, b_k is a bias for k , the $w_{u,v,k',k}$ is a connection weight among neuron k on layer l and its input is positioned at row u , column v and feature map k' .

In CNN, the activation function is highly required which is applied before the pooling layer. The MPELU activation function is applied which aims to generalize and unify the ReLU and Exponential Linear Unit (ELU). Moreover, the MPELU is capable to adaptively switching between ReLU and ELU which is given in Eq. (8). The cross-entropy loss function is given in Eq. (9),

$$MPELU(x) = \lambda \begin{cases} x & x > 0 \\ \alpha(e^{\beta x} - 1) & x \leq 0 \end{cases} \quad (8)$$

$$S = -\sum_x p(x) \log(p(x)) \quad (9)$$

Where, α is a hyperparameter learnable to improve its representation ability, $\beta > 0$, S is cross-entropy loss, x is a discrete random variable, $p(x)$ is a probability distribution function. The high entropy loss provides more uncertain distributions whereas less entropy loss provides definite distribution of variables. The overall pseudocode is given as follows:

Pseudocode:

Initialize input liver CT scan image

Preprocess the image using a wiener filter to remove noise and enhance the quality

Apply data augmentation to expand the training dataset

for each encoder block in the Multi-Encoder Network:

for each DDSC layer in the block:

Perform depthwise convolution on spatial dimensions

Perform pointwise convolution to combine channels

Apply Batch Normalization

Apply ReLU activation function

Use skip connection to transfer features to the corresponding decoder block

Down-sample using DDSC with stride = 2

Add ResNet shortcut to avoid gradient issues

for each down-sampling stage:

Integrate feature maps from corresponding stages of all encoders using skip connections

Fuse the feature maps to combine multi-scale and multi-modal information

for each decoder block:

for each DDSC layer in the block:

Perform up-sampling using deconvolution with stride = 2

Apply Batch Normalization

Apply ReLU activation function

Fuse feature maps from encoder to corresponding decoder stage via skip connection

Apply 1×1 convolution to predict output segmentation map (number of channels equals number of classes)

Apply SoftMax to normalize pixel-wise classification

Output segmented liver tumor regions

Feed the segmented images into the CNN-MPELU classifier

for each layer in CNN:

Perform $z_{i,j,k} = b_k +$

$$\sum_{u=0}^{f_h-1} \sum_{v=0}^{f_w-1} \sum_{k'}^{f_{n'}-1} x_{i',j',k'} \cdot W_{u,v,k',k}$$

Perform $MPELU(x) = \lambda \begin{cases} x & x > 0 \\ \alpha(e^{\beta x} - 1) & x \leq 0 \end{cases}$

Calculate $S = -\sum_x p(x) \log(p(x))$

Output final classification result (tumor class)

4. Result analysis

The proposed DDSC-MENet is simulated in Python with the following system configurations: RAM 8GB, windows 10 OS and Intel i5 processor. The dice, Volumetric overlap error (VOE) and Relative volume difference (RVD) metrics are applied for calculating the segmentation performance of DDSC-MENet for the LiTS2017 and 3DIRCADB

datasets. The precision, f1-score, accuracy and sensitivity metrics are applied for calculating classification performance for LiTS2017 and 3DIRCADB datasets. The formula for all the metrics is given in Eqs. (10)- (16).

$$Dice = \frac{2TP}{2TP+FP+FN} \times 100 \tag{10}$$

$$VOE = 1 - \frac{TP}{TP+FP+FN} \times 100 \tag{11}$$

$$RVD = \frac{FP}{TP+FN} \times 100 \tag{12}$$

$$Precision = \frac{TP}{TP+FP} \times 100 \tag{13}$$

$$F1 - score = \frac{2 \times precision \times sensitivity}{precision + sensitivity} \tag{14}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \tag{15}$$

$$Sensitivity = \frac{TP}{TP+FN} \times 100 \tag{16}$$

Where, TP, TN, FP and FN are true positive, true negative, false positive and false negatives respectively.

Table 1. Segmentation results for both datasets

Dataset	Methods	Dice (%)	VOE (%)	RVD (%)
LiTS2017	SENet	87.31	46.77	21.82
	U-Net	89.47	43.20	17.29
	RU-Net	91.66	41.86	14.38
	CU-Net	92.15	37.64	11.51
	DDSC-MENet	94.58	34.51	8.64
3DIRCADB	SENet	84.29	39.87	23.58
	U-Net	87.46	36.29	20.17
	RU-Net	89.53	33.52	16.36
	CU-Net	90.19	31.47	13.64
	DDSC-MENet	92.62	28.34	11.57

Table 1 and Fig. 7 denote the segmentation result for both LiTS2017 and 3DIRCADB datasets in terms of dice, VOE and RVD metrics. The existing state-of-art methods such as Squeeze-and-Excitation Network (SENet), U-Net, Recurrent U-Net (RU-Net) and Cascaded U-Net (CU-Net) are evaluated for both LiTS2017 and 3DIRCADB datasets. The DDSC-MENet achieves dice 94.58%, VOE 34.51% and RVD 8.64% for the LiTS2017 dataset which is better than SENet, U-Net, RU-Net and CU-Net. Similarly, the DDSC-MENet achieves dice 92.62%, VOE 28.34% and RVD 11.57% for the 3DIRCADB dataset which is better than SENet, U-Net, RU-Net and CU-Net.

Table 2 and Fig. 8 denote the activation function result for both LiTS2017 and 3DIRCADB datasets in terms of precision, f1-score, accuracy and sensitivity metrics. The existing state-of-art methods such as ELU, ReLU, Leaky ReLU (LReLU) and PELU are evaluated for both LiTS2017 and 3DIRCADB datasets. The CNN-MPELU achieves 96.32% precision, 96.03% f1-score, 97.48% accuracy and 95.76% sensitivity for LiTS2017 dataset which is better than ELU, ReLU, LReLU and PELU. Similarly, the CNN-MPELU achieves 94.66% precision, 94.08% f1-score, 96.25% accuracy and 93.51% sensitivity for 3DIRCADB dataset which is better than ELU, ReLU, LReLU and PELU.

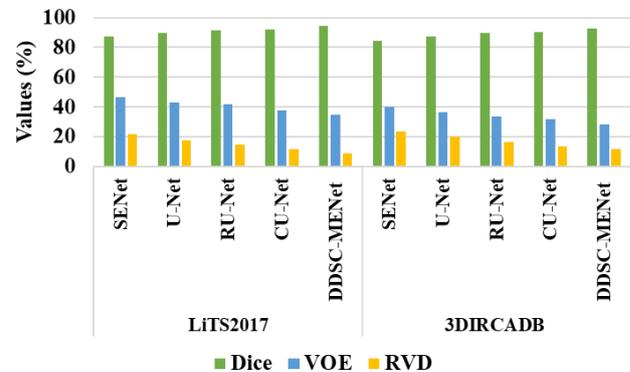


Figure. 7 Segmentation results for both datasets

Table 2. Activation function results for both datasets

Dataset	Methods	Precision (%)	F1-score (%)	Accuracy (%)	Sensitivity (%)
LiTS2017	ELU	90.46	90.04	90.35	89.63
	ReLU	91.73	90.97	92.56	90.24
	LReLU	93.51	92.95	94.23	92.41
	PELU	95.65	95.01	96.67	94.38
	MPELU	96.32	96.03	97.48	95.76
3DIRCADB	ELU	87.16	86.75	89.54	86.36
	ReLU	89.21	88.86	91.71	88.53
	LReLU	91.58	91.02	93.83	90.47
	PELU	92.34	92.50	94.38	92.68
	MPELU	94.66	94.08	96.25	93.51

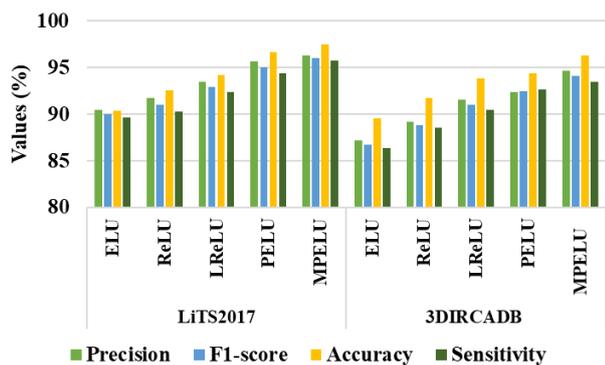


Figure. 8 Activation function results for both datasets

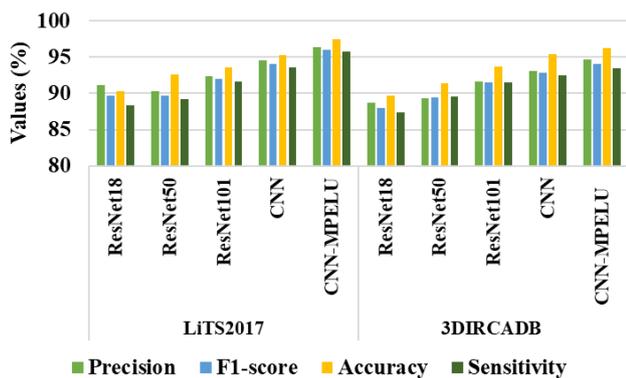


Figure. 9 Classification results for both datasets

Table 3 and Fig. 9 denote the classification result for both LiTS2017 and 3DIRCADB datasets in terms

of precision, f1-score, accuracy and sensitivity metrics. The existing state-of-art methods such as ResNet18, ResNet50, ResNet101 and CNN are evaluated for both LiTS2017 and 3DIRCADB datasets. The CNN-MPELU achieves 96.32% precision, 96.03% f1-score, 97.48% accuracy and 95.76% sensitivity for LiTS2017 dataset which is better than ResNet18, ResNet50, ResNet101 and CNN. Similarly, the CNN-MPELU achieves 94.66% precision, 94.08% f1-score, 96.25% accuracy and 93.51% sensitivity for the 3DIRCADB dataset which is better than ResNet18, ResNet50, ResNet101 and CNN.

4.1 Comparative analysis

The comparison of DDSC-MENet is given in this section for both LiTS2017 and 3DIRCADB datasets in terms of dice, VOE and RVD metrics. The existing methods such as SAA-Net [16], MS-FANet [17], MSA-TSA [18], S2DA-Net [19], 3D U-Net [20] and MAEG-Net [21] are compared with the proposed DDSC-MENet. The DDSC-MENet achieves dice 94.58%, VOE 34.51% and RVD 8.64% for LiTS2017 dataset. The DDSC-MENet achieves dice 92.62%, VOE 28.34% and RVD 11.57% for 3DIRCADB dataset. Table 4 denotes the comparative analysis for both datasets.

Table 3. Classification results for both datasets

Dataset	Methods	Precision (%)	F1-score (%)	Accuracy (%)	Sensitivity (%)
LiTS2017	ResNet18	91.09	89.71	90.36	88.38
	ResNet50	90.25	89.75	92.63	89.27
	ResNet101	92.38	92.01	93.57	91.65
	CNN	94.56	94.04	95.29	93.53
	CNN-MPELU	96.32	96.03	97.48	95.76
3DIRCADB	ResNet18	88.71	88.02	89.67	87.36
	ResNet50	89.37	89.49	91.42	89.62
	ResNet101	91.62	91.57	93.68	91.54
	CNN	93.15	92.81	95.39	92.48
	CNN-MPELU	94.66	94.08	96.25	93.51

Table 4. Comparative analysis for both datasets

Dataset	Methods	Dice (%)	VOE (%)	RVD (%)
LiTS2017	SAA-Net [16]	84.49	NA	NA
	MS-FANet [17]	74.2	36.7	10.7
	MSA-TSA [18]	85.96	NA	NA
	S2DA-Net [19]	NA	44.87	44.11
	3D U-Net [20]	NA	37	14
	MAEG-Net [21]	71.84	38.64	12.38
	DDSC-MENet	94.58	34.51	8.64
3DIRCADB	MS-FANet [17]	78	31.3	15.5
	MSA-TSA [18]	83.67	NA	NA
	S2DA-Net [19]	NA	38.29	34.66
	DDSC-MENet	92.62	28.34	11.57

4.2 Discussion

This section explains the existing method limitations and advantages of the proposed DDSC-MENet for liver tumor segmentations. The existing method has limitations such as SAA-Net [16] struggled with tumor segmentation because of less contrast against liver tissue surroundings thereby reducing segmentation performance. MS-FANet [17] struggled with small and irregular tumor segmentation for accurate boundary detection. MSA-TSA [18] was complex in interpreting attention weights that complicated the decision-making process thereby affecting performance. S2DA-Net [19] struggled with tumor variability and anatomical structures among various patients that affect generalizability. 3D U-Net [20] struggled with overfitting due to noise which affects the segmentation performance. MAEG-Net [21] struggled with high heterogeneous tumor generalization because of overfitting to edge features. To overcome these limitations, this research proposed a DDSC-MENet to enhance the capacity of segmentation by a huge number of skip connections. The MENet with a special down-sampling structure is developed for liver tumor segmentation in that whole down-sampling route is comprised of four encoders through same structure. The decoder connections in MENet allow the model to obtain feature representations from multiple encoder branches and perform integration of multi-scale and multi-modal information. The DDSC-MENet enhances the capacity of segmentation of tumours having different shapes, size or intensities to the existing regular model. Moreover, the DDSC-MENet contribute to preserving high-frequency information and preventing significant distortion of the image after down-sampling and enhancing the performance of the tumor boundary extraction.

5. Conclusion

The DDSC-MENet is proposed in this research for liver tumor segmentation which enhances the capacity of segmentation of tumors having different shapes, sizes or intensities to the existing regular model. Initially, the LiTS2017 and 3DIRCADB datasets are processed by wiener filter and data augmentation. The wiener filter is utilized to remove noise and artefacts in liver images which improves the reliability and quality of the image. The data augmentation is applied by replacing some training images with new images which enhances the training dataset to see the huge variety of liver tumors. The MENet is developed in this research with the down-

sampling structure for liver tumor segmentation in that the whole down-sampling route is comprised of four encoders along with the same structure. The decoder connections in MENet allow the model to obtain feature representations from multiple encoder branches and perform integration of multi-scale and multi-modal information. The CNN-MPELU is applied in the classification stage for classifying liver tumor, which helps to resolve the vanishing gradient issues and quickens the convergence. The DDSC-MENet achieves dice of 94.58% and 92.62% for LiTS2017 and 3DIRCADB datasets, respectively. In future, clustering-based methods will be applied to further improve the segmentation performance.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

The paper conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, have been done by 1st author. The supervision and project administration, have been done by 2nd author.

Notation

Notation	Description
i	Highest integer
F	Output feature map
I	Input feature map
K	Convolution kernel size
$z_{i,j,k}$	Neuron output at row i , column j and feature map k of convolution layer
f_h and f_w	Height and width of relevant fields
f_n'	Number of features inside the layer
$x_{i',j',k'}$	Neuron output at row i , column j and feature map k of previous convolution layer $l - 1$
b_k	Bias for k
$w_{u,v,k',k}$	Connection weight among neuron k on layer l
α	Hyperparameter learnable
S	Cross-entropy loss
x	Discrete random variable
$p(x)$	Probability distribution function
TP	True positive
TN	True negative
FP	Fase positive
FN	False negative

References

- [1] B.M. Tummala, and S.S. Barpanda, "Liver tumor segmentation from computed tomography images using multiscale residual dilated encoder-decoder network", *International Journal of Imaging Systems and Technology*, Vol. 32, No. 2, pp. 600-613, 2022.
- [2] A. Hänsch, G. Chlebus, H. Meine, F. Thielke, F. Kock, T. Paulus, N. Abolmaali, and A. Schenk, "Improving automatic liver tumor segmentation in late-phase MRI using multi-model training and 3D convolutional neural networks", *Scientific Reports*, Vol. 12, p. 12262, 2022.
- [3] P.K. Balasubramanian, W.C. Lai, G.H. Seng, and J. Selvaraj, "Apestnet with mask r-cnn for liver tumor segmentation and classification", *Cancers*, Vol. 15, No. 2, p.330, 2023.
- [4] S. Di, Y.Q. Zhao, M. Liao, F. Zhang, and X. Li, "TD-Net: A hybrid end-to-end network for automatic liver tumor segmentation from CT images", *IEEE Journal of Biomedical and Health Informatics*, Vol. 27, No. 3, pp. 1163-1172, 2023.
- [5] R. He, S. Xu, Y. Liu, Q. Li, Y. Liu, N. Zhao, Y. Yuan, and H. Zhang", "Three-dimensional liver image segmentation using generative adversarial networks based on feature restoration", *Frontiers in Medicine*, Vol. 8, p. 794969, 2022.
- [6] Y. Wu, H. Shen, Y. Tan, and Y. Shi, "Automatic liver tumor segmentation used the cascade multi-scale attention architecture method based on 3D U-Net", *International Journal of Computer Assisted Radiology and Surgery*, Vol. 17, No. 10, pp. 1915-1922, 2022.
- [7] D.T. Kushnure, and S.N. Talbar, "HFRU-Net: High-level feature fusion and recalibration unet for automatic liver and tumor segmentation in CT images", *Computer Methods and Programs in Biomedicine*, Vol. 213, p. 106501, 2022.
- [8] R. Li, L. Xu, K. Xie, J. Song, X. Ma, L. Chang, and Q. Yan, "DHT-Net: Dynamic hierarchical transformer network for liver and tumor segmentation", *IEEE Journal of Biomedical and Health Informatics*, Vol. 27, No. 7, pp. 3443-3454, 2023.
- [9] H. Rahman, T.F.N. Bukht, A. Imran, J. Tariq, S. Tu, and A. Alzahrani, "A deep learning approach for liver and tumor segmentation in CT images using ResUNet", *Bioengineering*, Vol. 9, No. 8, p. 368, 2022.
- [10] B.M. Anderson, B. Rigaud, Y.M. Lin, A.K. Jones, H.C. Kang, B.C. Odisio, and K.K. Brock, "Automated segmentation of colorectal liver metastasis and liver ablation on contrast-enhanced CT images", *Frontiers in Oncology*, Vol. 12, p. 886517, 2022.
- [11] J. Wang, Y. Peng, S. Jing, L. Han, T. Li, and J. Luo, "A deep-learning approach for segmentation of liver tumors in magnetic resonance imaging using UNet++", *BMC Cancer*, Vol. 23, p. 1060, 2023.
- [12] F. Lyu, M. Ye, A. J. Ma, T. C. -F. Yip, G. L. -H. Wong, and P. C. Yuen, "Learning From Synthetic CT Images via Test-Time Training for Liver Tumor Segmentation", *IEEE Transactions on Medical Imaging*, Vol. 41, No. 9, pp. 2510-2520, 2022.
- [13] X. Wang, S. Wang, Z. Zhang, X. Yin, T. Wang, and N. Li, "CPAD-Net: Contextual parallel attention and dilated network for liver tumor segmentation", *Biomedical Signal Processing and Control*, Vol. 79, Part 2, p. 104258, 2023.
- [14] J. Ma, G. Yuan, C. Guo, X. Gang, and M. Zheng, "SW-UNet: a U-Net fusing sliding window transformer block with CNN for segmentation of lung nodules", *Frontiers in Medicine*, Vol. 10, p. 1273441, 2023.
- [15] B.C. Anil, and P. Dayananda, "Automatic liver tumor segmentation based on multi-level deep convolutional networks and fractal residual network", *IETE Journal of Research*, Vol. 69, No. 4, pp. 1925-1933, 2023.
- [16] C. Zhang, J. Lu, Q. Hua, C. Li, and P. Wang, "SAA-Net: U-shaped network with Scale-Axis-Attention for liver tumor segmentation", *Biomedical Signal Processing and Control*, Vol. 73, p. 103460, 2022.
- [17] Y. Chen, C. Zheng, W. Zhang, H. Lin, W. Chen, G. Zhang, G. Xu, and F. Wu, "MS-FANet: multi-scale feature attention network for liver tumor segmentation", *Computers in Biology and Medicine*, Vol. 163, p. 107208, 2023.
- [18] Z. Wang, J. Zhu, S. Fu, and Y. Ye, "Context fusion network with multi-scale-aware skip connection and twin-split attention for liver tumor segmentation", *Medical & Biological Engineering & Computing*, Vol. 61, No. 12, pp. 3167-3180, 2023.
- [19] H. Liu, J. Yang, C. Jiang, S. He, Y. Fu, S. Zhang, X. Hu, J. Fang, and W. Ji, "S2DA-Net: Spatial and spectral-learning double-branch aggregation network for liver tumor segmentation in CT images", *Computers in Biology and Medicine*, Vol. 174, p. 108400, 2024.
- [20] S. Di, Y. Zhao, M. Liao, Z. Yang, and Y. Zeng, "Automatic liver tumor segmentation from CT images using hierarchical iterative superpixels

- and local statistical features”, *Expert Systems with Applications*, Vol. 203, p. 117347, 2022.
- [21] T. Zhang, Y. Liu, Q. Zhao, G. Xue, and H. Shen, “Edge-guided multi-scale adaptive feature fusion network for liver tumor segmentation”, *Scientific Reports*, Vol. 14, No. 1, p.28370, 2024.
- [22] LiTS2017 dataset link: <https://www.kaggle.com/datasets/andrewmvd/liver-tumor-segmentation> (Accessed on Oct 2024).
- [23] 3DIRCADB dataset link: <https://www.kaggle.com/datasets/nguyenhoainam27/3dircadb> (Accessed on Oct 2024).
- [24] O.H. Kesav and G.K. Rajini, “Enhancing Brain Tumor Detection and Classification with Reduced Complexity Spatial Fusion Convolutional Neural Networks”, *International Journal of Intelligent Engineering & Systems*, Vol. 17, No. 1, pp. 263-277, 2024, doi: 10.22266/ijies2024.0229.25.
- [25] M. Ibrahim, M. Mahmoud, R.M. Albadawy, and H. Abdulkader, “Liver Multi-class Tumour Segmentation and Detection Based on Hyperion Pre-trained Models”, *International Journal of Intelligent Engineering & Systems*, Vol. 15, No. 6, pp. 392-405, 2022, doi: 10.22266/ijies2022.1231.36.
- [26] S.V. Vanmore, and S.R. Chougule, “Liver Lesions Classification System using CNN with Improved Accuracy”, *Journal of Advanced Applied Scientific Research*, Vol. 2454, p. 3225, 2022.