U-Net-based Pancreas Tumor Segmentation from Abdominal CT Images

H S Saraswathi[1], Mohamed Rafi[2]
Department of Computer Science & Engineering, Jain Institute of Technology, Davangere-577003, India[1]
Department of Studies in Computer Science & Engineering, UBDT College of Engineering, Davangere-577004, India[2]

***Abstract*—There is no doubt that pancreatic cancer is one of the most deadly types of cancer. In order to diagnose and stage pancreatic tumors, computed tomography (CT) is widely used. However, manual segmentation of volumetric CT scans is a time-consuming and subjective process. It has been shown that the U-Net model is highly effective for semantic segmentation, although several deep learning models have been proposed. In this study, we propose a U-Net-based method for pancreatic tumor segmentation from abdominal CT images and demonstrate its simplicity and effectiveness. Using the U-Net architecture, the pancreas is segmented from CT slices in the first stage, while tumors are segmented from masked CT images in the second stage. For validation set of NIH dataset and according to the proposed method's dice scores, the pancreas segmentation and tumor segmentation performance was outstanding, demonstrating its potential to identify pancreatic cancer efficiently and accurately.**

***Keywords*—U-net; deep learning; segmentation; computed tomography images; hyper parameters; PDAC**

I. Introduction

The pancreas is an accessory organ, an exocrine gland of the digestive system, and an endocrine gland that generates hormones. In a healthy adult, the pancreas weighs around 100g, ranges in length from 14 to 25 cm, and has a volume of roughly 72.4 to 25.8 cm^3. It has an extended lobular form. The five anatomical divisions are the uncinate process, neck, body, and tail. Release of digestive enzymes that assist the digestion of fatty foods is the primary duty of the exocrine pancreas. The endocrine gland assists in regulating blood sugar levels and cell nutrient uptake. Exocrine pancreas cancer is referred to as "pancreatic cancer". It is one of the most common cancers, particularly in western countries and Japan. Pancreatic cancer is the melanoma that occurs in Americans the second most frequently. It accounts for 5% of all cancer mortality in that country and is more common in African Americans. It strikes men more commonly than women. The incidence increases after age 50. Pancreatic cancer develops when the cells of the pancreas undergo abnormal DNA modifications; this leads to the cells' uncontrolled division and growth into tumors. Occasionally, the liver, abdominal wall, lymph nodes, lungs, or bones may become infected by this tumor. The risk factors for getting pancreatic cancer include smoking, being overweight, having long-term diabetes, having a substantial family history of the illness, eating a lot of processed food and red meat, and having chronic pancreatitis. It is the fourth most common reason for cancer-related deaths in the west. It is predicted to overtake cancer as the second-deadliest illness in 10 years. It has an annual incidence rate of 12.50 per 100,000 persons, accounting for 3% of all cancer cases in America. One of the most fatal cancers, pancreatic cancer has a five-year mortality rate of less than 10%. Computed tomography (CT) scans of pancreatic tumors must be accurately segmented in order to make a diagnosis, determine a course of treatment, and track the disease's progression. Radiologists must manually segment images, which takes time and is prone to inter-observer variation. Deep learning models for automated segmentation have shown the potential in overcoming these constraints. The U-Net model is a convolutional neural network architecture designed for semantic segmentation tasks, particularly in biomedical image processing, including the segmentation of CT (Computed Tomography) images [15]. The U-Net architecture is named after its U-shaped design, which consists of two main parts: the contracting path and the expansive path. In this study the U-net model is proposed to segment both pancreas and hen to segment pancreas tumour. Each U-Net in our suggested system has a unique set of hyperparameters, and they are all cascaded together. When initializing the weights of the first U-Net, we employ Kaiming initialization to prevent the issues caused by disappearing or exploding gradients. A multi-class cross-entropy loss function, used by the second U-Net, is suitable for segmentation tasks. Our method is straightforward and efficient, requiring little time and resources compared to other deep learning methods that demand numerous layers and substantial computational resources for equivalent accuracy.

II. Literature Review and Related Work

The bulk of efficient deep learning algorithms for semantic segmentation of the pancreas and PDAC from abdominal CT scans are based on the U-Net, ResNet, AlexNet, and VGG-net models, as well as variants of these models.

The first of research was proposed by the Pancreas Segmentation in Abdominal CT Images with U-Net Model by [1] in their study using a convolutional neural network (CNN) called the U-Net model. The segmentation procedure used the Pancreas CT dataset from The Cancer Imaging Archive (TCIA) database, which included computed tomography images from 82 patients. The report provides a thorough explanation of the outcomes of this segmentation method. The Dice and Jaccard similarity coefficients, used to measure the efficacy of the U-Net model for pancreatic segmentation, produced values of 0.78 and 0.66, respectively. In [2] the study being reviewed introduces a framework based on convolutional neural networks (CNNs) for the segmentation of

pancreatic ductal adenocarcinoma (PDAC) mass and surrounding vessels in CT images. The proposed framework first localizes the pancreas region from the whole CT volume using a 3D-CNN architecture and 3D Local Binary Pattern (LBP) map of the original image. Then, segmentation of the PDAC mass is performed using 2D attention U-Net and Texture Attention U-Net (TAU-Net), which is introduced by fusing dense Scale-Invariant Feature Transform (SIFT) and LBP descriptors into the attention U-Net. The Dice score for PDAC mass segmentation in portal-venous phase by 7.52% compared to state-of-the-art methods in terms of DSC. In [3] AX-Unet: A Deep Learning Framework for Image Segmentation to Assist Pancreatic Tumor Diagnosis the author presents AX-Unet, a deep learning framework incorporating a modified atrous spatial pyramid pooling module to reduce information loss during down sampling and achieve information decoupling between channels. The proposed explicit boundary-aware loss function also tackles the blurry boundary problem. Segmentation of pancreatic ductal adenocarcinoma (PDAC) and surrounding vessels in CT images using deep convolutional neural networks and texture descriptors. To prevent information loss during down sampling and accomplish information decoupling between channels, this author introduces AX-Unet, a deep learning system integrating a modified Atrous spatial pyramid pooling module. The blurry border issue is addressed by the explicit boundary-aware loss function that has been proposed. With a DSC of 85.9±5.1% and a Jaccard of 77.9±3.4%, experimental data show that AX-Unet surpasses state-of-the-art techniques in pancreas CT image segmentation. Additionally, the model's extracted feature output reveals notable variations in the pancreatic area between healthy individuals and patients with pancreatic tumours, which can help doctors detect pancreatic tumors earlier. In [4] study is done on Artificial Intelligence-based Segmentation of Residual Tumor in Histopathology of Pancreatic Cancer after Neoadjuvant Treatment. In this study, digitized H&E-stained slides from resected pancreatic cancer after NAT were used to train a modified U-net model to segment tumour, normal ducts, and residual epithelium classes. The DenseNet161 encoder provided the highest mean segmentation accuracy, and a promising F1 score of 0.86 was attained for tumour segmentation. This work demonstrates that AI-based residual tumour burden assessment is viable; might be created as a tool for the impartial assessment of treatment response, and could be used to direct adjuvant treatment decisions. Experimental results demonstrate that AX-Unet outperforms state-of-the-art methods in pancreas CT image segmentation with a DSC of 85.9 ± 5.1% and a Jaccard of 77.9 ± 3.4%. Additionally, the extracted feature output of the model shows significant differences in the pancreatic region of normal people and patients with pancreatic tumors, which can assist physicians in the screening of pancreatic tumors. In [5] the authors of this research provide a brand-new self-learning architecture for training the PDAC segmentation model with a significantly bigger patient population and a mixture of annotated and unannotated venous or multi-phase CT images. Unannotated images are combined by two teacher models with various PDAC segmentation specialties to produce pseudo-annotations, which can then be improved by a teaching assistant model that recognizes related vessels around the pancreas. On both manually annotated and fictitiously annotated multi-phase images, a student model is subsequently trained. The findings demonstrate that the suggested approach offers a 6.3% Dice score absolute improvement over the robust baseline of nn-UNet trained on annotated pictures, obtaining a performance (Dice = 0.71), which is comparable to the inter-observer variability between radiologists.

In [6], the authors of this research suggest a brand-new, completely 3D cascaded framework for segmenting the pancreas in 3D CT scans. A 3D detection network (PancreasNet) that locates the pancreas regions and two distinct scales of a 3D segmentation network (SEVoxNet) that segment the pancreas in a cascading fashion depending on the detection findings of PancreasNet make up the framework's two primary parts. On the publicly available NIH pancreatic segmentation dataset, the suggested method produces cutting-edge results with a mean Dice Similarity Coefficient (DSC) of 85.93% and a mean Jaccard Index (JI) of 75.38%. Deep neural networks' limitations in segmenting the pancreas were explored by Zhou et al. [7] due to its complex and changing backdrop regions. Using the anticipated segmentation map, they reduced the input region. We provide a fixed-point model that accepts the segmentation mask as input and output. On the NIH pancreas segmentation dataset, where they tested their model, they outperformed the state-of-the-art by more than 4% (in terms of DSC). A DSC of 82.37±5.68% on average was attained. In order to overcome the issue of spatial non-smoothness of inter-slice pancreatic segmentation, Cai et al. [8] presented a stacked CNN-RNN model. Convolutional long short-term memory (CLSTM) units make up the RNN sub-network. Deep supervision and multi-scale feature map aggregation were used to modify the 2D CNN sub-network. The investigations make use of the NIH-CT dataset and 79 abdomen T1-weighted MRI scans. In the CT dataset and the MRI dataset, they obtained DSC of 83.35.6% and 80.77.40%, respectively.

A pancreatic segmentation model based on 2020 First International Conference on Power, Control and Computing Technologies (ICPC2T) 402 bidirectional convolutional long short-term memory (BiCLSTM) networks and spatial context information (SCU-Net) was proposed by Hao Li et al. [9]. The CT abdominal scans were divided into multiple isometric blocks using a divide-and-conquer technique, and a multi-channel CNN was created to make use of spatial context information. To encourage the interaction between information flow from bidirectional sequences, the BiCLSTM network is introduced. For inter-slice constraint and regularisation, a new loss function was developed. DSC, Jaccard index, pixels-wise precision, and recall were the evaluation measures employed. They examined the NIH-CT dataset, which included 82 abdomen enhanced 3D CT scans, and found that the final prediction accuracy was 82.863%, the mean Jaccard Index was 68%, and the mean prediction error was 12% and mean recall of 80.2% and mean Precision of 82.2%. In [10] states the importance of huge number of training samples in the deep learning networks. Because of the considerable inter-patient structural changeability in both outline and size dimensions, the pancreas is a difficult

abdominal organ to section [16,17]. Convolutional neural networks (CNNs) give better performance [11, 12].

III. METHODOLOGY

A. U-Net Architecture

Two levels of the U-Net model are employed, consisting of downward-moving paths followed by upward-moving paths in U-shaped network architecture. Each block on the downward path is composed of two convolution layers with the Leaky ReLu activation function and a maximum pooling layer with stride 2. In the first block, we use 32 filters of size 3x3 for each of the two convolution layers. The journey upwards is paved with 2x2 transposed convolution and the joining of feature maps from the parallel downward path. Along this path, we encounter four blocks, each composed of two convolutional layers with 3x3 filters and a Leaky ReLU activation function. We chose the Leaky ReLU for its efficiency, swift convergence and sparse activation - when faced with negative input, it returns 0; for positive input, it returns the same value as shown in equation (1), where x stands for the input.

$$f(x)=\max(0,x) \quad (1)$$

Despite using the identical U-Net model at both levels, the hyperparameter values vary. The weights of the first U-Net (U-Net-1) are initialised using Kaiming Initialization, and the filter size is 3x3. This approach avoids the disappearing or bursting gradients problem brought on by exponentially growing or decreasing input signals by accounting for the nonlinearity of activation functions like ReLU. The equation below can be used to express each convolution layer's response.

$$y_l = W_l x_l + b_l \quad (2)$$

In this case, x is a k2 c x 1 vector that represents k x k pixels that are co-located in c input channels. The layer's spatial filter size is k.

The number of connections in a response is $n=k^2c$. The dimensions of the matrix W is a d x n, and each row of the matrix reflects the weights of a certain filter. In the output map, Y represents the response at a specific pixel, and b is a vector of biases. If l is the layer index and f is the activation, $xl = (y_{l-1})$.

Given in below Eq. is the necessary condition to prevent the gradients issue.

$$0.5n_l\mathrm{Var}[W_l]=1 \text{ for all } l \quad (3)$$

Where W_l denotes the random component of each of the element in W_l and the Var depicts variance. This results in a start of the structure indicated in below equation.

$$W_l \sim N(0,2/n_l) \quad (4)$$

Specifically, a zero-centered Gaussian with a $\sqrt{2/\mathrm{nl}}$ standard deviation.

R2 regularization is used in the U-Net-1 configuration to prevent overfitting. The penalty is basically increased by regularization as model complexity rises. All parameters, with the exception of the intercept, are penalized by the regularization term lambda (), making sure that the model properly generalizes the data and does not overfit. R2 regularization is useful when there are correlatively dependent characteristics. W denotes the weight matrix with values i and j.

There are two parts to it: an encoding route and a decoding path, where the encoder extracts high-level features from the input image and the decoder creates a segmentation mask using those features [18].

Elbow U-Net is the name of this specific version, which was created in Python using the TensorFlow library. It is made to segment binary images with the intention of locating areas of interest that are labelled with pixel values of 1 while the backdrop is labelled with 0. The input image dimensions, which are given as an input layer to the model, are IMG_HEIGHT x IMG_WIDTH x IMG_CHANNELS. After that, the input layer is normalized by using a lambda layer to multiply each pixel's value by 255. A succession of convolutional layers with 16, 32, 64, 128, and 256 filters are used in the encoding path. In order to avoid overfitting, a dropout layer is placed after each convolutional layer. In order to minimise the spatial dimensions of the feature maps, the output of each convolutional layer is then passed through a max pooling layer. Transpose convolutional layers, also referred to as deconvolutional layers, and are used in the decoding path to upsample the feature maps back to their initial size. In order to maintain spatial information, the decoder also has skip connections that concatenate the feature maps from the associated encoding path layer. The first convolutional layer in the encoder is a transpose layer with 128 filters, and the next two layers are each convolutional layers with 128 filters. Until the final layer, which comprises of a transpose convolutional layer with 16 filters, followed by two convolutional layers with 16 filters each, this is repeated with progressively smaller filter sizes. A segmentation mask with image values between 0 and 1 is produced by a convolutional layer with one filter and a sigmoid activation function in the output.

Table I provides a detailed illustration of the U-Net model's construction. An expansive path follows a contracting path in the five-level U-Net network. The parameters Batchnorm, activation functions, dropout are shown in detail.

TABLE I. HYPERPARAMETER SETTINGS FOR PROPOSED U-NET MODEL

Encoder path (Contracting)					Decoder path (Expansive)			
1st Level	2nd Level	3rd Level	4th Level	5th Level	4th Level	3rd Level	2nd Level	1st Level
CNN Para Sizes: (16, 3, 3, 3) (16)	MaxPool2D WithIndices	MaxPool2D WithIndices	MaxPool2D WithIndices	MaxPool2D WithIndices	Concatenation	Concatenation	Concatenation	Concatenation
Batch_Norm Para Sizes:(16) (16)	BatchNorm Para Sizes:(32) (32)	BatchNorm Para Sizes:(64) (64)	BatchNorm Para Sizes:(128) (128)	BatchNor m Para Sizes:(256) (256)	BatchNo rm Para Sizes:(51 2) (512)	BatchNo rm Para Sizes:(25 6) (256)	BatchNo rm Para Sizes:(12 8) (128)	BatchNo rm Para Sizes:(64) (64)
Leaky_Relu	CNN Param Sizes:(64, 32, 3, 3) (64)	CNN Param Sizes:(128, 64, 3, 3) (128)	CNN Param Sizes:(256, 128, 3, 3) (256)	CNN Param Sizes:(512 , 256, 3, 3) (512)	CNN Param Sizes:(25 6, 512, 3, 3) (256)	CNN Param Sizes:(12 8, 256, 3, 3) (128)	CNN Param Sizes:(64 , 128, 3, 3) (64)	CNN Param Sizes:(32 , 64, 3, 3) (32)
BatchNorm Param Sizes: (16) (16)	Relu	Relu	Relu	Relu	Relu	Relu	Relu	Relu
CNN Param Sizes:(32, 16, 3, 3)(32)	Feature Dropout	Feature Dropout	Feature Dropout	Feature Dropout	Feature Dropout	Feature Dropout	Feature Dropout	Feature Dropout
Leaky_Relu	BatchNorm Param Sizes:(64) (64)	BatchNorm Param Sizes:(128) (128)	BatchNorm Param Sizes:(256) (256)	BatchNor m Param Sizes:(512) (512)	BatchNorm Param Sizes:(25 6) (256)	BatchNorm Param Sizes:(12 8) (128)	BatchNorm Param Sizes:(64) (64)	BatchNorm Param Sizes:(32) (32)

The proposed architecture of a U-Net model for medical picture segmentation is depicted in the Table II. In medical image analysis, U-Net is a common deep learning architecture, especially in segmentation tasks where the objective is to label each pixel in an image with its appropriate class. A contracting path is found on the left side and an expansive path is found on the right side of the U-Net architecture. While the expansive path gradually upsamples the feature maps to generate a segmentation mask, the contracting path uses a sequence of convolutional and pooling layers to capture the context of the input image. The earliest stages of the contracting process, which minimises the spatial dimensions of the feature maps, are convolutional and max pooling layers. When this procedure is performed numerous times, the number of filters in each convolutional layer rises at each level. By using batch normalisation and activation algorithms, the model performs better. After obtaining the feature maps from the contracting path, the expansive path uses transposed convolutional layers to progressively increase their spatial dimensions. The feature maps from the contracting path are concatenated with the corresponding feature maps from the expanding path in order to keep the high-resolution data. The output layer is then activated with a softmax function to generate the probability distribution.

The last convolutional layer of the U-Net is activated using a softmax activation function to provide a probability map for each pixel, which is then thresholded to produce the final binary segmentation map.

The proposed framework is based on the application of two tiers of U-Net model as shown in the Fig. 1. The model has two sets of encoding and decoding blocks. This can help to improve the segmentation performance by capturing more complex features at different scales.

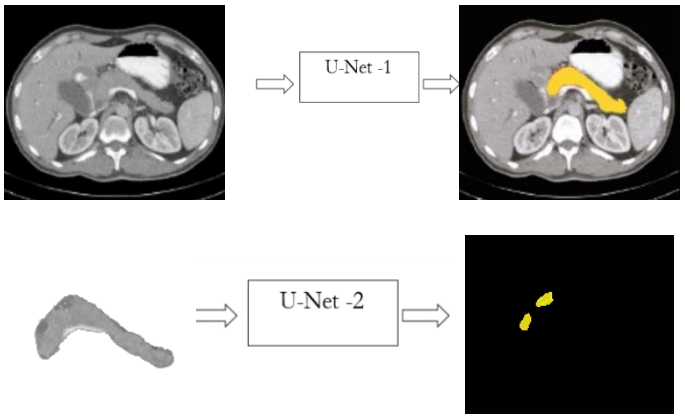

Fig. 1. U-Net based architecture for pancreas tumor segmentation.

IV. EXPERIMENTS AND RESULTS

A. Pancreas Segmentation Training

The outcomes of a pancreatic segmentation training method are displayed in Table III. The aim of the training is to correctly identify the pancreas in medical photographs using a machine learning system. When compared to the ground truth data, the algorithm's ability to predict the right pancreas segmentation is measured by the training loss. Better performance is indicated by a reduced training loss. How closely the anticipated segmentation matches the actual segmentation is determined by the dice coefficient. Better performance is indicated by a higher dice coefficient [14]. The training process starts with a relatively high training loss of 0.76 and a training dice coefficient of 0.82. As the training progresses, the algorithm gets better at identifying the pancreas, resulting in a decrease in training loss and an increase in the dice coefficient as shown in the Table II.

TABLE II. ARCHITECTURE OF PROPOSED MODEL

Epoch	Time(hrs)	Training Loss	Training Dice	Validation Dice
0	1.19	0.75599	0.82197	0.90297
1	1.19	0.72110	0.88970	0.91160
2	1.2	0.71554	0.90505	0.929217
3	1.2	0.71121	0.91528	0.93275
4	1.2	0.70814	0.92264	0.91778
5	1.2	0.70615	0.92812	0.94656
6	1.19	0.70434	0.93308	0.93837
7	1.19	0.70248	0.93663	0.95023
8	1.2	0.70103	0.94062	0.95054
9	1.19	0.70042	0.94320	0.93310
10	1.19	0.70020	0.94497	0.94846
11	1.2	0.69875	0.94772	0.94290
12	1.2	0.69725	0.94830	0.95534
13	1.2	0.69738	0.95020	0.95216
14	1.19	0.69677	0.95076	0.93893
15	1.19	0.69677	0.95171	0.93644
16	1.2	0.69662	0.95299	0.94623
17	1.2	0.69586	0.95253	0.94996
18	1.19	0.69561	0.95498	0.95489
19	1.2	0.69544	0.95585	0.95485

Based on the findings, the model was able to validate with a maximum accuracy of 0.9939 and a minimum loss of 0.0159 at the twelfth epoch. The accuracy values were about 0.95 in the early epochs but quickly increased to reach the high levels in the later epochs. With an 11-batch training run and an unspecified input image size, the model was trained over the course of 25 epochs. Also not stated was the loss function that was employed. It's important to note that the model's strong accuracy scores on the validation set indicate that the model generalizes well to new data. To verify the model's robustness, it's crucial to assess the model's performance on a different test set. Overall, the results of your U-Net-based pancreatic segmentation model are encouraging, and further analysis on a different test set may reveal more details about how well it performs.

B. *Lesion Segmentation Training*

Table III displays the outcomes of a training process for a machine learning algorithm that tries to recognize legions in medical photos. Each of the 20 training epochs, which each took between 0.4 and 1.6 hours to complete, was performed once. The table lists the training loss, the training dice coefficient, and the validation dice coefficient for each epoch.

Training was done for 20 epochs and took around 11 hours to complete as shown Fig. 2.

When compared to the real-world data, the algorithm's ability to predict the right legion segmentation is measured by the training loss. Better performance is indicated by a reduced training loss. How closely the anticipated segmentation matches the actual segmentation is determined by the dice coefficient. Better performance is indicated by a higher dice coefficient. The training dice coefficient of 0.70 and the initial training loss of 1.45 both point to subpar performance in the first epoch. However, as training goes on, the algorithm improves at recognizing legions, which causes the training loss to go down and the dice coefficient to go up.

The algorithm's capacity to recognize legions has greatly increased by the time training is complete, as seen by the training loss of 1.35 and training dice coefficient of 0.83. With a score of 0.69, the validation dice coefficient likewise performs well.

TABLE III. TRAINING OUTCOMES

Epoch	Time(hrs)	Training Loss	Training Dice	Validation Dice
0	0.40532	1.44871	0.70470	0.6492
1	0.41683	1.40648	0.75877	0.6681
2	0.40805	1.39184	0.77948	0.64474
3	0.40694	1.38603	0.78834	0.66002
4	0.40503	1.38146	0.79132	0.62748
5	0.41182	1.37874	0.79927	0.71489
6	0.43404	1.37206	0.80300	0.70638
7	0.40494	1.36892	0.80577	0.64888
8	0.40575	1.36734	0.81093	0.67519
9	0.40510	1.36397	0.81306	0.6871
10	0.40516	1.36620	0.81297	0.69063
11	1.66921	1.36206	0.81292	0.68888
12	0.41266	1.35946	0.82007	0.66742
13	0.40665	1.35924	0.81898	0.66363
14	0.40700	1.35743	0.81908	0.69324
15	0.40637	1.35359	0.82549	0.65909
16	0.40637	1.35794	0.82055	0.69015
17	0.40565	1.35338	0.82610	0.71414
18	0.40493	1.35071	0.82977	0.64887
19	0.40503	1.35127	0.82449	0.69984

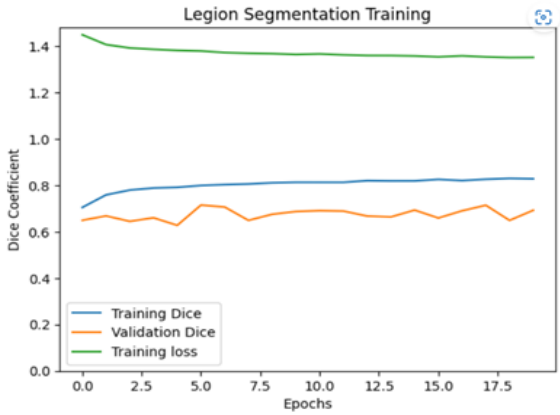

Fig. 2. Tumor segmentation training.

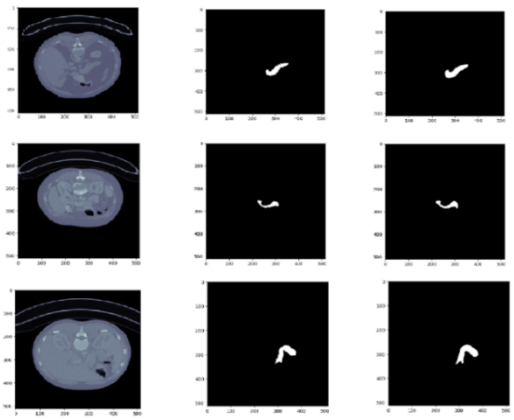

Fig. 3. Pancreas segmentation results.

In Fig. 3 by showing a visual comparison of the pancreas segmentation results generated using the suggested method and the ground truth segmentation, the usefulness of the proposed method is proved. The visual comparison is shown in a three-column arrangement, with the abdominal CT images in the first column, the ground truth segmentation in the second, and the segmentation results from the proposed method in the third.

The visual comparison is used to assess how well the suggested strategy performs in precisely segmenting the pancreas from abdominal CT images. The effectiveness of the suggested method in precisely segmenting the pancreas may be assessed by contrasting the segmentation results achieved by the proposed method with the ground truth segmentation.

It is possible to compare the segmentation results obtained by the suggested approach with the ground truth segmentation quickly and effectively by using a three-column format that offers a clear and concise depiction of the segmentation findings. Additionally, using abdominal CT scans in the first column offers a relevant and realistic context for assessing how well the suggested strategy works in a clinical scenario.

Fig. 3 also offers a visual comparison of the PDAC segmentation results produced by the suggested method and the ground truth segmentation in order to assess the performance of the proposed method. The visual comparison is shown in a three-column arrangement, with the abdominal CT images in the first column, the ground truth segmentation in the second, and the segmentation results from the proposed method in the third. The abdominal CT images, which were obtained from various patients and utilized to test the suggested procedure, are shown in the first column. The manual ground truth segmentation, carried out by a skilled radiologist, is displayed in the second column. The performance assessment of the suggested method uses the ground truth segmentation as a benchmark.

The segmentation results produced by the suggested method are shown in the third column. The outcomes demonstrate that the suggested approach is successful in precisely segmenting PDAC tumors in abdominal CT images. Analysis of similarities and differences between the segmentation results produced by the suggested method and the actual segmentation is done.

TABLE IV. COMPARISON WITH EXISTING WORK

Methods	DSC (%)	Jaccard (%)	Recall (%)	Precision (%)
CNN [1]	78	66	71	74
Unet and texture [2]	60	----	78.0	57.8
AX-Unet [3]	87.7	78.2	90.9	92.9
Unet with DenseNet [4]	83	----	----	----
nnUnet[5]	71	----	----	----
Fully Convolutional Network[13]	71	60	69	72
3D CNN[6]	88	71	84	82
Fixed Point[7]	82.37%	77	71	73
Attention Unet [19]	84	----	84.9	84.1
DenseASPP [20]	85	----	----	----
Cascaded FCN [21]	85.9	75.7	85.2	87.6
Proposed Model	88.2	79	82	86

In Table IV we compare several segmentation techniques of existing methods for medical picture analysis, especially using the NIH dataset, in this work. Bottom-up, Fixed-point, 3D Coarse-to-Fine, Holistically nested, RSTN, Recurrent Contextual Learning, Attention Unet, DenseASPP, (46), Cascaded FCN, AX-Unet, and the suggested technique are among the approaches investigated.

DSC, Jaccard index, recall, and accuracy were the four measures we used to compare the effectiveness of various approaches. In comparison to the other approaches, our proposed method had the highest values for all four criteria, demonstrating improved segmentation performance. Although some metrics had slightly lower values for the AX-Unet method, it still performed well overall.

V. PERFORMANCE EVALUATION METRICS

Typically, the segmentation's results is discussed here. Various challenges provide the ground truth against which processes are validated. Various standards are used to analyze the performance and accuracy of segmentation. The most popular statistical measures are based on a number of research articles, and the different problems are described below.

Assume that the segmented region is represented by X and the ground truth is by Y. To assess the precision of the segmentation, we may utilize a variety of indicators. The Volumetric Overlap Error (VOE) is one such measure. This calculates how much of the divided region overlaps with the actual scene. It is determined by dividing the sum of the pixels in x and y intersection by the sum of the pixels in X and Y's union. The result is then multiplied by 100 and removed from 1. Successful segmentation is indicated by a VOE value that is near to 0, whereas greater values signify disparities between the segmented pictures. The following is a formula for VOE.

$$VOE = ((|X \cap Y| / |X \cup Y|) - 1) * 100$$

Here dice similarity coefficient (DSC) is an additional measure. This gauges how well the split region's pixels correspond to the actual scene. A DSC score around 1 suggests effective segmentation, whereas a number near 0 implies inconsistencies across the segmented pictures. The DSC formula is displayed below:

$$DSC = 2 |X \cap Y| / (|X| + |Y|)$$

The segmented region's volume difference from the ground truth is measured by the Relative Volume Difference (RVD). It is computed by dividing the sum of the volumes of X and Y, deducting 1, and then multiplying the result by 100. Positive RVD values signify over-segmentation, whilst negative RVD values signify under-segmentation. The RVD formula is displayed below:

RVD = ((total volume of X / total volume of Y) - 1) x 100

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

The percentage of true positives (TP) that are accurately detected by the system, out of all real positive instances (TP + false negative (FN)), is known as recall, also known as sensitivity or true positive rate. In other words, recall gauges how well a system can identify all positive cases. A high recall score means that a significant number of positive cases may be recognized by the system.

The proportion of true positives that are successfully recognized by the system, out of all anticipated positive instances (TP + false positive (FP)), is measured by precision, also known as positive predictive value. To put it another way, precision assesses a system's capacity to accurately identify only pertinent positive cases. A high accuracy rating means that the system produces few false positives, which means that the instances that are projected to be positive are more likely to be accurate.

The validation set provided by NIH dataset has yielded the following results both for the segmentation of the pancreas and pancreas tumour.

The segmentation of Pancreas: dice for each case of 0.9283, dice global of 0.9300, VOE of 0.114, and RVD of -0.070.

The segmentation of PDAC: dice for each case of 0.5852, dice worldwide of 0.8900, VOE of 0.434, and RVD of -0.158.

VI. Conclusion

In this study, we show the simplicity and efficiency of the U-Net model for semantic biomedical picture segmentation. It is possible to successfully segment the pancreas and tumours from abdominal CT scans using the proposed architecture, which is based on a variant of the fundamental U-Net model. By changing the fundamental U-Net model's hyper parameter values in accordance with the type of dataset, we may attain accuracy that is comparable to that of sophisticated state-of-the-art techniques. The model's performance was assessed using the National Institutes of Health (NIH) dataset, and an 88.2% (Dice Global) pancreas tumour segmentation accuracy was attained. The proposed method can be applied to MRI images and we also aim to extend our research work by combining the U net model with other deep learning models to improve the accuracy.

References

[1] Kurnaz, Ender, and Rahime Ceylan. "Pancreas segmentation in abdominal CT images with U-Net model." 2020 28th Signal Processing and Communications Applications Conference (SIU). IEEE, 2021.

[2] Mahmoudi, Tahereh. "Segmentation of pancreatic ductal adenocarcinoma (PDAC) and surrounding vessels in CT images using deep convolutional neural networks and texture descriptors." Scientific Reports 12.1 (2022): 3092.

[3] Yang, Minqiang. "AX-Unet: A deep learning framework for image segmentation to assist pancreatic tumor diagnosis." Frontiers in Oncology 12 (2022): 894970.

[4] Janssen, Boris V. "Artificial intelligence-based segmentation of residual tumor in histopathology of pancreatic cancer after neoadjuvant treatment." *Cancers* 13.20 (2021): 5089.

[5] Zhang, Ling. "Robust pancreatic ductal adenocarcinoma segmentation with multi-institutional multi-phase partially-annotated CT scans." Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part IV 23. Springer International Publishing, 2020.

[6] Wang, Wenzhe "A fully 3D cascaded framework for pancreas segmentation." 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). IEEE, 2020.

[7] Zhou, Yuyin. "A fixed-point model for pancreas segmentation in abdominal CT scans." International conference on medical image computing and computer-assisted intervention. Cham: Springer International Publishing, 2017.

[8] Cai, Jinzheng. "Pancreas segmentation in CT and MRI images via domain specific network designing and recurrent neural contextual learning." arXiv preprint arXiv:1803.11303 (2018).

[9] Li, Hao "Pancreas segmentation via spatial context based u-net and bidirectional lstm." arXiv preprint arXiv:1903.00832 (2019).

[10] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer International Publishing, 2015.

[11] Cai, Jinzheng. "Pancreas segmentation in CT and MRI images via domain specific network designing and recurrent neural contextual learning." arXiv preprint arXiv:1803.11303 (2018).

[12] Roth, Holger. "Towards dense volumetric pancreas segmentation in CT using 3D fully convolutional networks." Medical imaging 2018: image processing. Vol. 10574. SPIE, 2018.

[13] Heinrich, Mattias P., Max Blendowski, and Ozan Oktay. "TernaryNet: faster deep model inference without GPUs for medical 3D segmentation using sparse and binary convolutions." International journal of computer assisted radiology and surgery 13 (2018): 1311-1320.

[14] Hemmat, Mohammad Hossein Askari. "U-Net Fixed-Point Quantization for Medical Image Segmentation." ISBI, 2019.

[15] Man, Yunze, "Deep Q learning driven CT pancreas segmentation with geometry-aware U-Net." IEEE transactions on medical imaging 38.8 (2019): 1971-1980.

[16] Li, Jun. "Probability map guided bi-directional recurrent UNet for pancreas segmentation." arXiv preprint arXiv:1903.00923 (2019).

[17] Yang, Zhengzheng. "Pancreas segmentation in abdominal CT scans using inter-/intra-slice contextual information with a cascade neural network." 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2019.

[18] Huang, Huimin, "Unet 3+: A full-scale connected unet for medical image segmentation." ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, 2020.

[19] Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K. Attention U-Net: Learning Where to Look for the Pancreas. ArXiv Preprint ArXiv (2018) 1804:03999. doi: 10.48550/arXiv.1804.03999.

[20] P. Hu, "Automatic Pancreas Segmentation in CT Images With Distance-Based Saliency-Aware DenseASPP Network," in IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 5, pp. 1601-1611, May 2021, doi: 10.1109/JBHI.2020.3023462.

[21] Xue J, He K, Nie D, Adeli E, Shi Z, Lee SW. Cascaded Multitask 3-D Fully Convolutional Networks for Pancreas Segmentation. IEEE Trans Cybernet (2019) 51:2153–65. doi: 10.1109/TCYB. 2019.2955178.