# Automated GI Tract Segmentation with U-Net: A Comparative Study of Loss Functions

Bindu Madhavi Tummala [1,*], Subba Reddy Chavva [2], Yallamandaiah S. [3], A. Radhika [4], Dasari Chinna Veeraiah [1], Rishitha Jaladi [1], and Aruna Kumari Peruri [1]

[1] Department of Computer Science and Engineering, Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada, India
[2] Department of Information Technology, Velagapudi Ramakrishna Siddhartha Engineering College, India
[3] Department of Electronics and Communication Engineering, Vignan's Nirula Institute of Technology and Science for Women, Guntur, India
[4] Department of Computer Science and Engineering, SRK Institute of Technology, Vijayawada, India
Email: bindumadhavi@vrsiddhartha.ac.in (B.M.T.); chavvasubbareddy@gmail.com (S.R.C.); yallamandaiah@gmail.com (Y.S.); radhikaankala@gmail.com (A.R.); veeranani222@gmail.com (D.C.V.); rishitha.jaladi03@gmail.com (R.J.); arunakumarip1003@gmail.com (A.K.P.)
*Corresponding author

***Abstract*—Medical image processing has significantly transformed the landscape of healthcare, particularly in the diagnosis and treatment of various diseases. Gastrointestinal (GI) cancer has emerged as a rapidly growing concern, with an estimated 5 million new cases reported annually. To achieve this precision, healthcare professionals now leverage the cutting-edge Magnetic Resonance Imaging (MRI) modality known as MR-Linacs, which provides a daily view of tumor positions. However, a bottleneck in this process arises during the manual segmentation of healthy organs at risk, such as the stomach and intestines, from the obtained medical images. This task, performed by radiologists, is time-consuming and can significantly prolong treatment times, consequently amplifying the suffering of patients. Thus, automation of the GI tract segmentation helps oncologists seamlessly. Our research proposes a model that automates the segmentation of the GI tract. This study presents a U-Net model that segments the stomach and intestines from MRI scans. The dataset, sourced from the UW-Madison Carbone Cancer Center, comprises RLE-encoded masks for training annotations alongside 16-bit grayscale PNG images. Each case consists of multiple scan slices, split either by time or the entire case. Our approach used various combinations of loss functions on U-Net to improve the accuracy and efficiency of the automated segmentation of the gastrointestinal tract. Our model achieves high accuracy with the Dice+BCE loss function when compared with other loss functions. On the training dataset, the U-Net model with Dice+BCE loss function received the highest dice score of 0.9082 and IOU score of 0.8594. On the validation dataset, the model obtained a dice score of 0.8974 and an Intersection over Union (IoU) score of 0.8181. This research contributes to addressing the challenges associated with manual GI tract segmentation, offering a viable solution through automated segmentation using deep learning techniques.***

***Keywords*—Magnetic Resonance Imaging (MRI), combined loss functions, U-Net, UW-Madison Carbone cancer center, gastro-intestinal tract segmentation**

Manuscript received January 21, 2024; revised March 20, 2024; accepted April 7, 2024; published December 6, 2024.

## I. Introduction

The Gastrointestinal (GI) tract, is also known as the digestive tract. It encompasses multiple organs, including the small intestine, large intestine, and stomach. Gastrointestinal (GI) diseases are a significant global health issue that affects millions of individuals yearly. According to the World Health Organization (WHO), gastrointestinal diseases account for approximately 9% of the global burden of disease, and they are responsible for an estimated 1.5 million deaths per year [1]. In 2018, Gastrointestinal (GI) cancers are predicted to have resulted in 3.4 million fatalities and 4.8 million new cases worldwide. Globally, 26% of new cases of cancer and 35% of cancer-related fatalities are caused by GI malignancies. There will likely be 7.5 million new cases of GI cancer and 5.6 million GI cancer-related deaths worldwide by 2040, an increase of 58% and 73%, respectively. segmentation of the big intestines, stomach, and small bowel that is accurate and trustworthy, is essential for the diagnosis and treatment of these diseases. However, manual segmentation is a time-consuming, labor-intensive process and is vulnerable to both inter- and intra-observer variability. Due to these limitations, this approach may result in diagnostic inconsistencies. Therefore, there is a need for automated segmentation methods to assist clinicians in accurately identifying and analyzing the abnormalities present in the gastrointestinal tract. Automated segmentation can reduce the time and effort required for segmentation, increase the accuracy, and consistency of results, and improve overall patient care. It helps oncologists with effective radiation therapy. Oncologists can use MR-Linacs to Observe the exact position of the tumor and keep an eye out for the precise dosage based on the presence of cancer cells, which can change from day to day. The objective is to accurately locate the position of the stomach and intestines and effectively distribute the radiation intensity to the tumor while avoiding the surrounding organs. The primary goal is to precisely identify the position of the stomach and intestines, optimizing the distribution of radiation intensity to target the tumor while minimizing exposure to surrounding organs. Without the use of deep learning technologies, segmentation is a hard and time-consuming procedure that might easily increase treatment times ranging from several minutes to an hour daily.

The following are some of this research's key contributions:

(1) This research presented a deep learning-based U-Net model for the segmentation of organs such as large bowel, small bowel, and stomach in the Gastrointestinal tract.
(2) The suggested U-Net model's performance in segmenting the stomach, small intestine, and large intestine was assessed using the UW-Madison GI tract image segmentation dataset.
(3) Model performance metrics including model loss, dice coefficient, and IOU coefficient are used to evaluate the model.

Our work builds on these previous studies by implementing a combination of loss functions to achieve both improved segmentation accuracy and computational efficiency.

### A. *Motivation*

Oncologists encounter challenges in administering high radiation doses precisely to tumors using X-ray beams, while simultaneously ensuring avoidance of the stomach and intestines. Patients may find it challenging to tolerate treatments that are 15 min to an hour longer due to the laborious manual process of drawing the contour of these organs. So, we addressed the task of creating an automated system for segmenting the gastrointestinal tract from MRI images.

### B. *Gap and Contribution to Work*

Existing research on Gastrointestinal (GI) tract segmentation has made notable progress, yet a critical gap persists in achieving optimal accuracy. Current methodologies often utilize singular loss functions during training, limiting their ability to comprehensively address the complex anatomical structures and variations inherent in the GI tract. Moreover, there is a lack of thorough investigation and incorporation of diverse loss functions in current models, presenting an untapped opportunity. Each loss function captures unique aspects of the GI tract, and their combination could potentially lead to a more robust and accurate segmentation. The absence of a systematic examination of various loss function combinations within the U-Net framework represents a significant gap in the current state of research.

Our research addresses this gap by introducing a novel paradigm that significantly advances existing methodologies in GI tract segmentation. Unlike previous approaches, our work systematically explores and employs diverse combinations of loss functions within the U-Net framework. This systematic exploration serves as the primary motivation for our proposed research. Recognizing the limitations of existing segmentation methods, our study seeks to systematically explore and exploit different combinations of loss functions. By doing so, we aim to enhance the model's capacity to accurately delineate the complex structures of the GI tract. Through this approach, we anticipate achieving substantial improvements in segmentation accuracy and contributing significantly to the field of GI tract segmentation.

Following is how the remaining manuscript is organized. In Section II, a description of earlier research on the topic is offered. In Section III, the proposed technique is defined. In Section IV, the study is concluded, and Section V covers the important findings.

## II. Literature Review

This section shows the related works for our project, which has been done over the past years. It includes 8 research papers [2–12] and all the authors have used the "UW-Madison GI Tract image segmentation" dataset posted on the Kaggle platform. Sharma [2] used several types of encoders as the backbone of U-Net architecture to perform segmentation tasks on GI Tract scans. The researchers compared the performance of different encoders (ResNet, EfficientNet, VGG16, and MobileNet) and found that EfficientNet achieved the highest IoU score. An advantage of this research is that the optimal model achieved an IoU score of 85.3% utilizing the EfficientNet backbone. One limitation of this paper is the use of Intersection over Union (IoU) as a performance metric, but it may benefit from including additional metrics like the Dice coefficient to provide a more comprehensive evaluation of the models. Chou *et al.* [3] proposed a method for better the treatment of cancer. This research examines the accurate segmentation of the stomach and intestines in MRI scans. The primary benefit of this research is, to segment the organ areas, they used mask R-CNN and U-net strategies. The Mask R-CNN model outperforms the U-Net model. The Study's limitation is that considering a more appropriate loss function might increase the performance.

Sharma *et al.* [4] developed a U-Net model using a Transfer Learning model as the basis for the segmentation of the gastrointestinal system. Six pre-trained transfer learning models have been installed as a backbone to evaluate the performance of the suggested U-Net model: VGG19, DenseNet121, InceptionV3, SeResNet50, InceptionResNetV2, and EfficientNet B0. The primary benefit of this paper is, they had proposed a deep learning-based U-Net model for the small pictures to effectively enhance and extract local characteristics for segmentation. The models are evaluated using model performance indicators such as dice coefficient, IoU coefficient, and model loss. One Constraint of this research is the use of deep learning models, especially with transfer learning, which can introduce complexity in implementation, training, and interpretation. Nemani and Vollala [5] developed a model using U-Net++ as the decoder and LeViT as an encoder. In this research, to separate the various organs from an image, they developed a hybrid CNN-transformer architecture. Dice and Jaccard loss functions were utilized in this scenario.

The Main Advantage of this research is hybrid CNN-transformer architecture combines the strengths of both CNNs and Transformers, resulting in a more efficient and effective segmentation model. LeViT and UNet++ were utilized as encoder and decoder, respectively, in this design. Zhang *et al.* [6] proposed a method using the U-Net 2.5D Network. It is suggested to use the FCN U-Net network, which has a traditional encoder-decoder design. In comparison to FCN, U-Net can segment data more accurately while using less training data. The U-Net network implements convolution-based image segmentation. The advantage of this research is employing U-Net2.5D, specifically with EfficientNetB1 as the backbone architecture, the research capitalizes on cutting-edge neural network advancements. The Constraint for this research is relying solely on a single metric may not provide a comprehensive assessment of segmentation performance. Li *et al*. [7] examined several 2D, 2.5D, 3D, and their combined dimensions for segmentation. The weighted sum of Hausdorff distance and the dice similarity coefficient are used to implement the loss function, and the outcomes are adequate. The positive aspect of this research is it ensures that sparse inter-slice information and dense intra-slice information are adequately represented, leading to more accurate segmentation.

Ronneberger *et al.* [8] proposed a model known as the U-Net model for Biomedical Image Segmentation. The network design consists of two recurrent 3×3 convolutions, followed by rectified linear units, 2×2 max pooling operations, and stride 2 downsampling operations for each. We reduce the batch to a single image to reduce overhead and make the most of the GPU RAM by prioritizing large input tiles over a large batch size. The main advantage of this research is the U-Net architecture allows for good localization and context usage simultaneously. One constraint for this research is the model trained with limited annotated images and heavy data augmentation, there may be a risk of overfitting to the specific characteristics of the training data. Bindu Tummala and Barpanda [9] proposed a two-step training procedure using an encoder-decoder-based architecture to segregate liver tumors. The 3D-IRCADb1 dataset is taken into consideration for training and testing due to its tumor complexity. The main advantage of this paper is the adoption of MDICE, a combined loss function, enhances the learning capability of the network. This loss function likely helps in addressing the class imbalance and optimizing the segmentation performance.

Zhang *et al*. [10] proposed a model that combines many cutting-edge deep learning architectures, such as Edge U-Net for grayscale data segmentation, U-Net++ with VGG19 encoder for 2.5D data, and Inception-V4 for initial classification, to provide thorough and precise segmentation. The main advantage of this research is, that it shows promise as a productive and precise tool for physicians, representing a breakthrough in the area of GI tract image segmentation for radiation planning. Roshan *et al*. [11] proposed the Unet2.5D(Se-ResNet50) model, the suggested Unet2.5D model, which has demonstrated encouraging results with a Dice Coefficient of 0.848, showing precise segmentation. The positive aspect of this paper is, that it highlights the significance of cutting-edge architectures for better segmentation. One limitation of this study is Advanced architectures are complex and demand a lot of computer power and knowledge. Zhou *et al.* [12] proposed a method that uses a unique sampling technique to address class imbalance in medical image segmentation. One benefit of this research is by merging two UNet-based models into an ensemble model, segmentation performance is improved, providing a thorough and efficient method for medical picture analysis.

Observations on GI tract segmentation based on the related work are:

(1) Computational efficiency can be enhanced.
(2) Preprocessing quality needs improvement.
(3) Accuracy can be improved.
(4) One Loss function is applied. It is possible to combine different loss functions.

## III. Proposed Methodology

The primary objective of this project is to propose a U-Net model that can precisely partition healthy organs to aid radio-oncologists. The model's performance has been improved using various techniques while adhering to the fundamental U-Net design, and its computational needs were reduced. Before training, the model employs sophisticated preprocessing methods such as standardizing pixel values, enhancing contrast through histogram equalization, and reducing noise via median filtering. These steps collectively optimize image quality, facilitating effective feature extraction and learning. By standardizing, enhancing, and filtering the image data, the model can better discern relevant features and patterns during training, ultimately enhancing its performance across various tasks and datasets. The model simplifies input images to a compact 256×256-pixel size. This strategic down sampling not only reduces computational overhead but also focuses computational resources on pertinent features, thereby enhancing efficiency without sacrificing accuracy. Combining different types of loss functions in the model allows for a more comprehensive assessment of the segmentation task by considering various aspects such as pixel-wise accuracy, perceptual similarity, and adversarial robustness. It enables the model to learn from diverse sources of information, leading to a more nuanced understanding of the data and improved generalization capabilities. By leveraging multiple loss functions, the model can effectively balance competing objectives, resulting in more accurate and robust segmentation results. It facilitates the exploration of the most effective combination of loss functions to optimize performance for specific datasets or applications. The model incorporates hyperparameter tuning strategies, including dynamic adjustments to epochs, learning rate, batch size, and adaptive learning rate strategies. Hyperparameters like epochs determine the number of training iterations, learning rate regulates the step size in adjusting model weights, batch size controls the number of samples processed before updating the model, and adaptive learning rate strategies dynamically adjust the learning rate during training. This meticulous optimization enhances the model's adaptability to dataset intricacies, leading to more efficient convergence and improved performance.

This section describes the proposed methods for the segmentation of the GI tract. Information regarding the input dataset used for segmentation will be presented in Section A. The various pre-processing techniques discussed in Section B were used to improve the dataset before further processing. The proposed U-Net model, which was used in our project is shown in Section C. The diagram in Fig. 1 illustrates the conceptual model designed for segmenting the Gastrointestinal (GI) tract.

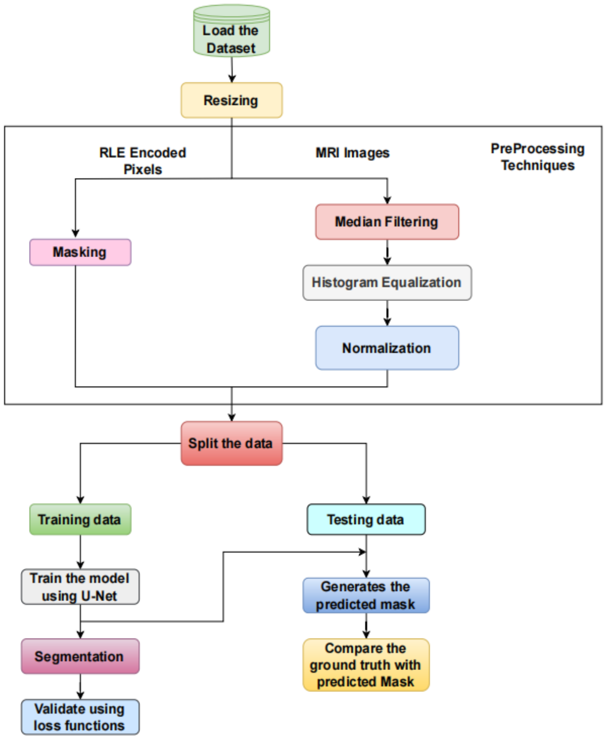



Fig. 1. Proposed model for segmentation of GI tract.

We have used the predefined network in our research. The decision to utilize a predefined network, specifically the U-Net model, is grounded in its well-established success in semantic segmentation tasks, particularly in medical imaging. This predefined structure provides a robust foundation, minimizing the need for intricate architectural design and saving valuable research time. Furthermore, pre trained U-Net models, often initialized with weights learned from diverse datasets, facilitate effective transfer learning. The predefined model acts as a feature extractor, enhancing the network's ability to discern complex anatomical structures even with limited task-specific data. Time and resource efficiency are paramount in medical research, and using a predefined U-Net model significantly reduces computational costs compared to training a model from scratch.

### *A. Dataset Description*

The dataset used in this research is "UW Madison GI Tract Image Segmentation" [13]. It contains MRI pictures of actual cancer patients who underwent one to five scans on various days while receiving radiation treatment. The images are presented as 16-bit grayscale PNG files, and the training annotations are offered as RLE-encoded masks. There are many sets of scan slices for each patient. It also contains a CSV file that consists of RLE-encoded pixels of the large bowel, small bowel, and stomach corresponding to each MRI scan. These RLE-encoded pixels can be converted into masks. The UW-Madison Carbone Cancer Centre is a leader in MRI-Linac [14] based radiotherapy and has been treating patients with MRI-assisted radiotherapy [15] based on their daily anatomy since 2015. In Madison, Wisconsin, there is a public land-grant institution called the Institution of Wisconsin-Madison. The dataset comprises 38,496 MRI scans, varying in size and sourced from multiple patients, encompassing a total of 85 cases. The MRI scan consists of 3 classes namely large bowel, small bowel, and stomach. Fig. 2 displays the MRI scan from the UW Madison GI tract segmentation dataset.

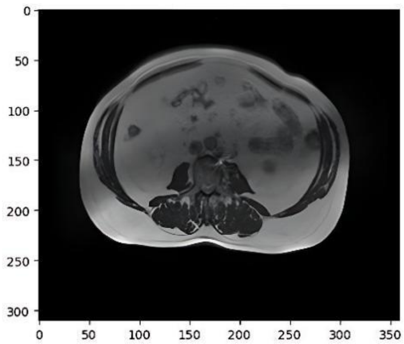



Fig. 2. A sample image from the dataset.

### *B. Data Pre-processing*

Data is pre-processed to improve its accuracy as well as reliability. It is the process of converting unprocessed data into a form that can be analyzed. In this study, the pre-processing steps included resizing, masking, bias correction, median filtering, Histogram equalization, smoothing, and Normalization. In the next sections, a detailed explanation of several pre-processing stages is provided.

#### *1) Resizing*

Resizing is the process of altering an image size. It involves modifying the number of pixels it contains. When an image is scaled, certain pixels from the original image are either added or subtracted. Our dataset contains MRI images of various sizes. The size 266×266 is the most common across all images in the sample (67.33% occurrence), followed by sizes 276×276, 310×360, and 234×234 in frequency-descending order. We standardized the dimensions of all images to 256×256 pixels to ensure uniformity in image sizes for generating masks of segmented areas [16].

*2) Masking*

In this research, our data set contains Run-Length Encoding (RLE) encoded pixels. we need to generate masks from these RLE-encoded pixels [17]. Converting RLE-encoded pixels into a mask is an essential step in image segmentation. The RLE format typically consists of pairs of values: the starting pixel index of a run and the length of that run. By converting RLE-encoded pixels into a mask, we can effectively represent the segmentation information in a binary format. Fig. 3 represents the mask generated from RLE-encoded pixels.

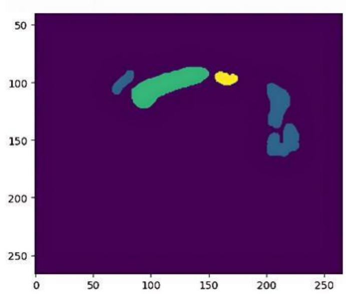
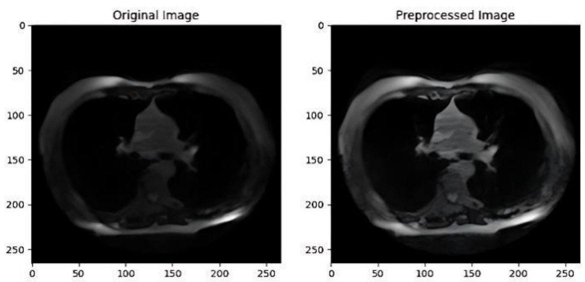


Fig. 3. MRI image.

*3) Histogram equalization*

This is a pre-processing technique that is used to enhance the contrast and visibility of medical images. It is especially helpful for improving low-contrast images because it produces a more uniform histogram by spreading the pixel intensities in the image. Fig. 4 displays the MRI scan before and after histogram equalization.




(a) (b)

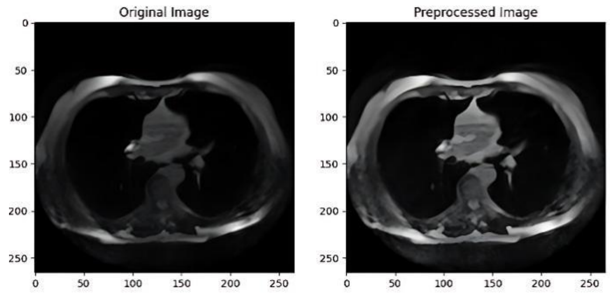
Fig. 4. MRI image: (a) Before histogram equalization, (b) After histogram equalization.

*4) Median filtering*

It is an image pre-processing technique that is used to remove noise from an image [18]. It is particularly useful for removing salt-and-pepper noise. Median filtering is a powerful tool for enhancing the quality of images. Fig. 5 displays the MRI scan before and after median filtering.

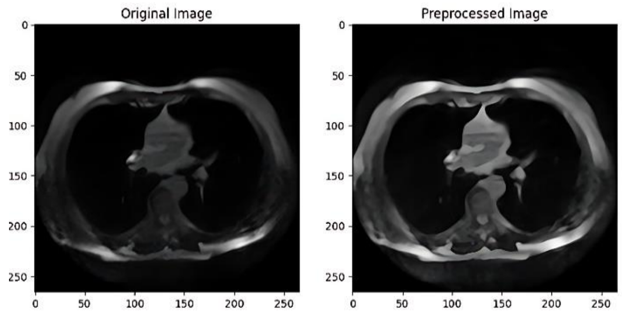
*5) Normalization*

Normalization is a pre-processing technique used to scale 25-pixel values in an image to a common range or distribution [19]. The goal of normalization is to raise an image's contrast and raise its overall quality. Fig. 6 displays the MRI scan before and after normalization (see Algorithm 1).




(a) (b)

Fig. 5. MRI image: (a) Before median filtering, (b) After median filtering.




(a) (b)

Fig. 6. MRI image: (a) Before normalization, (b) After normalization.

**Algorithm 1:** Preprocessing

**Input:** MRI Image of Gastro-Intestinal tract
**Output:** Preprocessed MRI Image
1:Read the input image from the specified input directory.
2: Convert the input image to a 32-bit float.
3: Perform bias correction using N4 Bias Field Correction.
corrected_image = N4BiasFieldCorrection(input_image)
4: Normalize the intensity of the corrected image.
normalized_image=corrected_image/max(corrected_image)
5: Apply histogram equalization using Adaptive Histogram Equalization (Adapthist).
img_eq=Adapthist(normalized_image,clip_limit=0.01)
6: Perform median filtering with a specified kernel size.
median_filtered_image=MedianFilter(img_eq,size=3)
7: Normalize the intensity of the smoothed image.
normalized_smoothed_image=Normalize(smoothed_image)
8: Save the preprocessed image to the specified output directory

## *C. Segmentation Using U-Net Model*

The U-Net model is a convolution neural network architecture that was initially created for biomedical image segmentation. The network is referred to as a "U-Net" due to its "U" shape, which has a contracting path (encoder) and an expanding path (decoder). The decoder network generates the segmentation mask after the encoder network down-scales the input image and extracts its features using a sequence of convolution and pooling layers, which up-samples the features. Several up-sampling and convolution layers comprise the decoder network, which reconstructs the image with a resolution that is the same as the original image. The encoder reduces the spatial dimensions in each layer while increasing the channels. In contrast, the decoder lowers the channels while raising the spatial dims.

Fig. 7 depicts the architecture for the U-Net model. While lowering the channels, the decoder, on the other hand, increases the spatial dims. The encoder is constructed by continually applying two 3×3 convolutions. Each convolution is followed by a ReLU and batch normalization. The spatial dimensions are subsequently compressed by using the 2×2 max pooling method. Once more, we quadruple the number of feature channels while halving the spatial dimensions for each down-sample step. When it comes to the decoder, In the expanding path, each step begins with an up sample of the conclude with a 2×2 transpose convolution, cutting the number of feature channels in half. feature map. A 3×3 convolutional is commonly used and it is concatenated with the appropriate feature map from the contracting path.

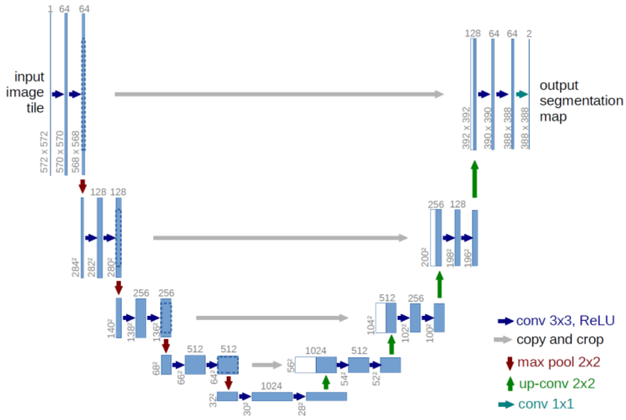



Fig. 7. U-Net architecture.

The model is trained using a loss function that compares the predicted mask with the actual mask during training. The optimization method then adjusts the network's parameters to minimize the loss function.

## *D. Loss Functions*

In our research we have used a variety of loss functions namely Dice loss, Focal loss, Tverskey loss, Binary Cross Entropy (BCE) loss, and a combination of these loss functions are also used.

### *1) BCE loss*

The BCE loss function calculates the loss by comparing each prediction to its corresponding true label, where the output is expected to be either 0 or 1 [20]. Specifically, for each prediction, the BCE loss penalizes deviations from the true label, assigning a higher loss for misclassifications with a larger discrepancy between the predicted probability and the ground truth label.

$$L_{BCE}(y, \hat{y}) = -(y\,log(\hat{y}) + (1 - y)log(1 - \hat{y})) \qquad (1)$$

Here, $\hat{y}$ is the predicted value by the prediction model.

### *2) Dice loss*

Dice loss quantifies the similarity between predicted and ground truth segmentation masks. It measures the overlap between the two masks by computing the ratio of twice the intersection of the masks to the sum of pixels in both masks [21]. Its formulation ensures robustness against class imbalance, producing sharper gradients during optimization.

$$DL\ (y, \hat{p}) = 1 - (2y\hat{p} + 1)/(y + \hat{p} + 1) \qquad (2)$$

To prevent the function from being undefined in certain situations, such as when both the ground truth and predicted values are zero, 1 is added to both the numerator and denominator. This ensures the calculation remains valid across various scenarios.

### *3) Focal loss*

Focal Loss is a modification of the Cross-Entropy loss, designed to address class imbalance in classification tasks. It introduces a modulating factor that down-weights the loss assigned to well-classified examples.

$$FL(p_t) = -\alpha_t (1 - p_t)^{\gamma} \log(p_t) \qquad (3)$$

Here, $(1 - p_t)^{\gamma}$ is a modulation factor that reduces the loss contribution from easy examples and focuses more on hard examples.

### *4) Tversky loss*

Tversky loss is designed to address class imbalance in segmentation tasks, especially in medical imaging. It introduces two hyperparameters, α and β, to adjust the emphasis on false positives and false negatives. Tversky loss allows for fine-tuning the model's sensitivity to different types of errors.

$$\text{Tversky Loss} = \frac{TP}{(TP + \alpha \times FP + \beta \times FN)} \qquad (4)$$

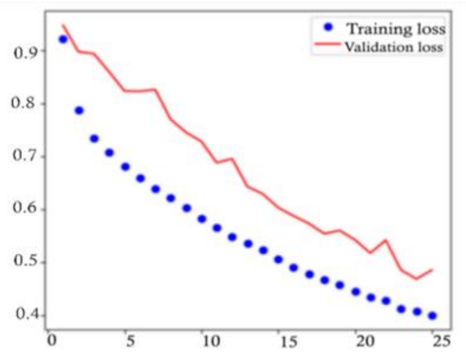
# IV. Result and Analysis

## *A. Hyper Parameter Tuning*

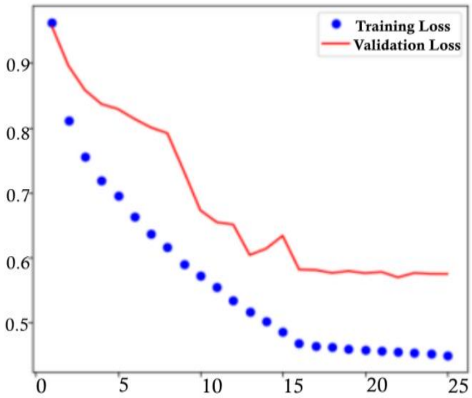
This is the process of selecting the best configuration of hyperparameters for a machine-learning model. They control aspects of the learning algorithm and influence the model's performance and behavior. Finding the ideal combination of hyperparameter values that produces the highest performance on a particular job or data collection is the aim of hyperparameter tweaking. Examples of hyperparameters include learning rate, batch size, number of epochs, and hidden layers in a neural network. These are the hyperparameters used in our proposed model. These are the hyperparameters used in our proposed model. The choice of a batch size of 64, 25 epochs, and a learning rate of 1.0000e–04 was made with a focus on optimizing computational efficiency and ensuring the convergence of the model [22].

## *B. Training and Validation Loss Analysis*

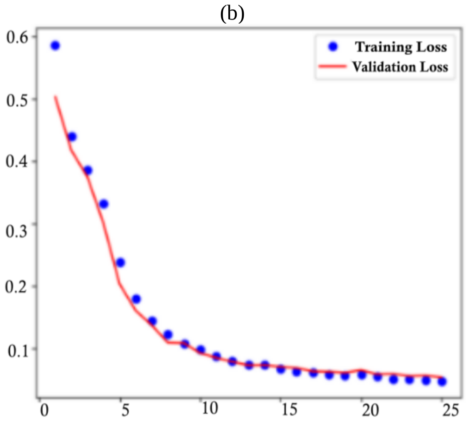
The proposed model was trained and evaluated using various loss functions [23]. The figures provided illustrate the training loss and validation loss for different combinations of these loss functions. By examining these figures, we can gain insights into the performance of the model across different training iterations. While the validation loss shows how successfully the model generalizes to new data, the training loss is the assessment of error or discrepancy between the expected and actual values during the training phase. The lower the validation loss, the better the model's performance. Comparing the trends and magnitudes of these losses can help us assess the effectiveness of different loss functions in optimizing the model's performance. Fig. 8 illustrates a comparison of model loss across different loss functions.



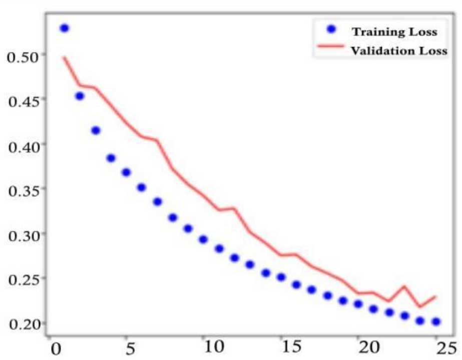

(a)



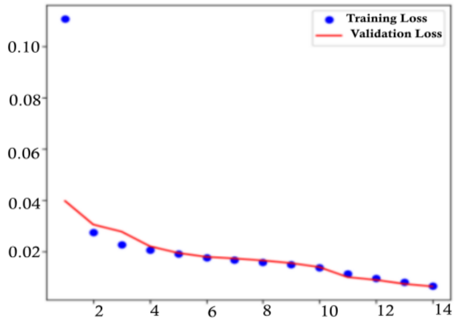

(b)




(c)




(d)



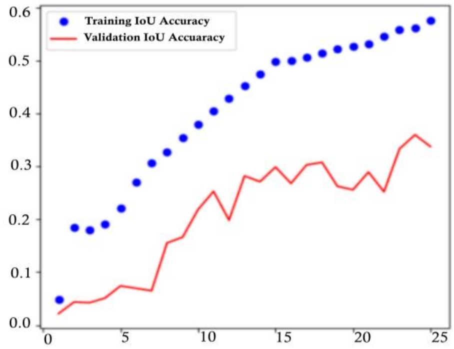

(e)

Fig. 8. Training loss and validation loss for focal tversky: (a) Loss, (b) BCE+Dice loss, (c) Focal+Dice loss, (d) dice loss, (e) BCE loss.

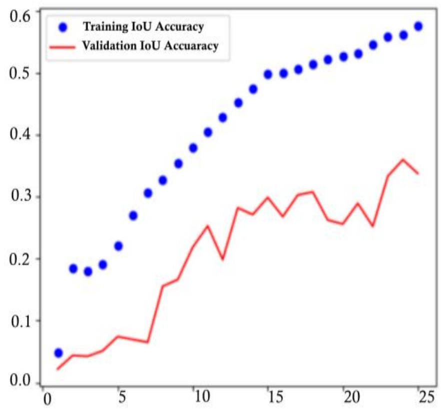
### C. IOU Coefficient Analysis

The IOU coefficient, also known as the Jacquard index, is a metric used to measure the similarity between two sets or Regions of Interest (ROIs). It provides a measure of the accuracy of the predicted sets compared to the ground truth. IOU Coefficient with higher values indicates better performance. The combined Dice and BCE loss functions during training and validation yielded the highest IOU coefficient of 0.8594, suggesting that the corresponding segmentation prediction most closely matched the actual data. The IOU coefficient curves during training and validation for various combinations of loss functions are shown in the figures below. Fig. 9 illustrates a comparison of IOU coefficient metrics across different loss functions.

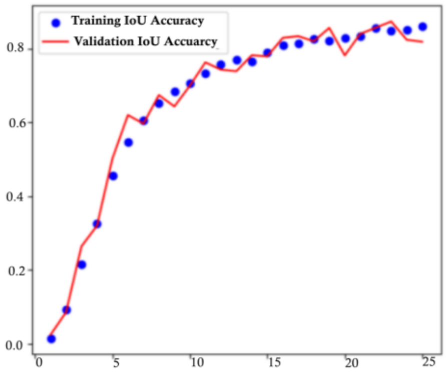
$$\text{IOU} = \frac{\textit{Area of Overlap}}{\textit{Area of Union}} \tag{5}$$



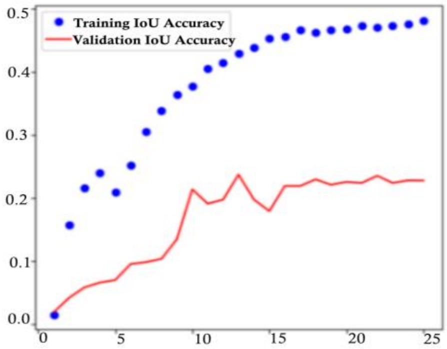

(a)



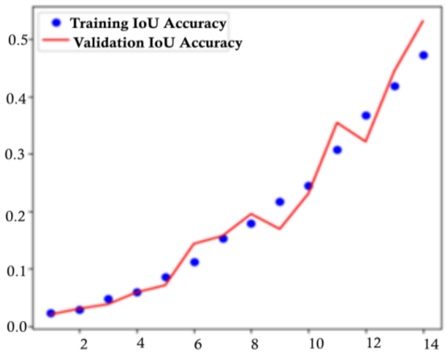

(b)




(c)




(d)




(e)

Fig. 9. IOU coefficient: (a) Focal tversky loss, (b) BCE+Dice loss, (c) Focal+Dice loss, (d) Dice loss, (e) BCE loss.

### D. *Dice Coefficient Analysis*

The Dice coefficient indicates how closely the expected and actual segmentation masks resemble one another. Calculating the amount of overlap between the predicted and ground truth segmentation serves as an evaluation metric for segmentation algorithms. The dice coefficient curves are shown in the figures below for various combinations of loss functions during training and validation. The highest dice coefficient of 0.9082 was achieved using the combined dice and BCE loss functions during training and 0.8974 during validation, indicating the corresponding segmentation prediction most closely matched the ground truth. Fig. 10 illustrates a comparison of Dice coefficient metrics across different loss functions.

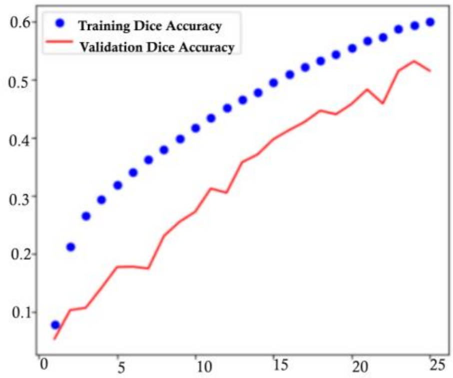
$$\text{Dice} = \frac{2|Intersection|}{(|Predicted| + |Ground\ Truth|} \tag{6}$$

where:

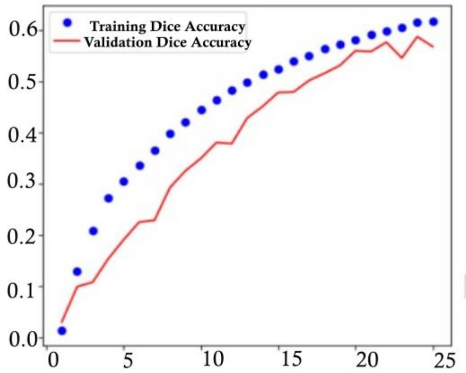
|Intersection| is the number of pixels that are both in the predicted mask and the ground truth mask.

|Predicted| is the total number of pixels in the predicted mask.

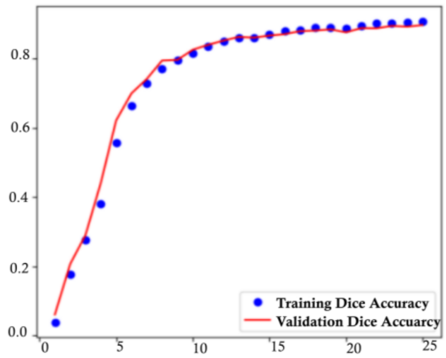
|Ground Truth| is the total number of pixels in the ground truth mask.



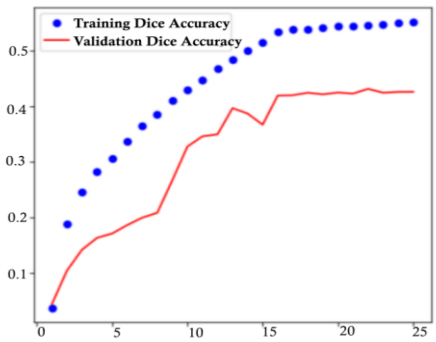

(a)



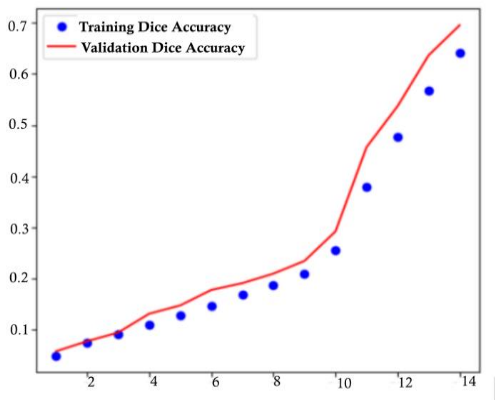

(b)




(c)




(d)

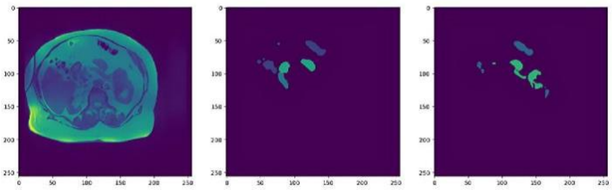



(e)

Fig. 10. Dice coefficient: (a) Focal tversky loss, (b) BCE+Dice loss, (c) Focal+Dice loss, (d) Dice loss, (e) BCE loss.

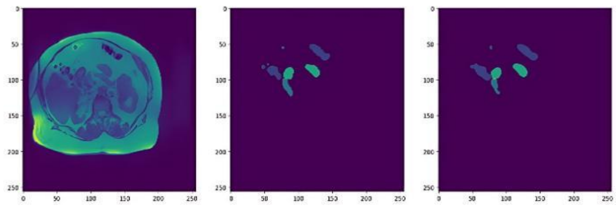
## *E. Testing for Various Loss Functions*

Images representing the results are shown in Fig. 11. The input images and the mask, which were obtained via RLE encoding, are shown in the figure. The predicted images are also shown in the figure using various combinations of loss functions. Here, the colors green, yellow, and blue represent the stomach, small bowel, and large bowel respectively. Fig. 11 illustrates a comparison of testing images across different loss functions.



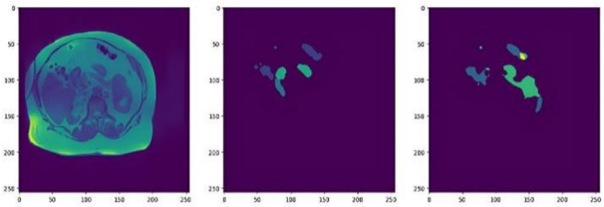
MRI Image Ground truth Predicted Mask

(a)



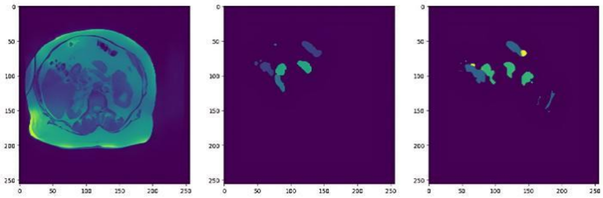
MRI Image Ground truth Predicted Mask

(b)



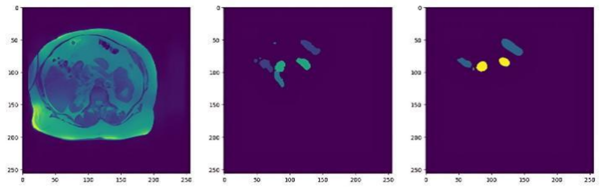
MRI Image Ground truth Predicted Mask

(c)



MRI Image Ground truth Predicted Mask

(d)



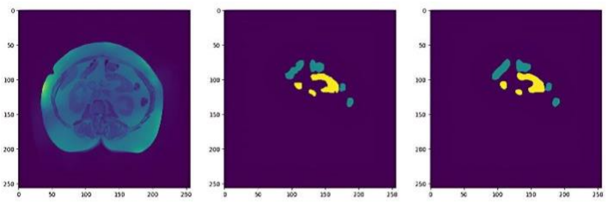
MRI Image Ground truth Predicted Mask

(e)

Fig. 11. Testing images, (a) Focal tversky loss, (b) BCE+Dice loss, (c) Focal+Dice loss, (d) Dice loss, (e) BCE loss.

## *F. Dice+BCE Loss*

Our Proposed U-Net model achieves good results by using the Dice+BCE loss function. Following are the test dataset images for the Dice+BCE loss function. Fig. 12 displays the testing images for the Dice+BCE loss function.



MRI Image Ground truth Predicted Mask

Fig. 12. Testing image for Dice+BCE loss.

TABLE I. COMPARISON OF MODEL LOSS, DICE, AND IOU COEFFICIENT FOR DIFFERENT COMBINATION OF LOSS FUNCTIONS

| Metrics | Training loss | YOU Coefficient | Dice Coefficient | Validation loss | Val_IOU Coefficient | Val_Dice Coefficient |
|---|---|---|---|---|---|---|
| **Focal Tversky** | 0.4002 | 0.5761 | 0.5998 | 0.4852 | 0.3374 | 0.5148 |
| **Dice+BCE loss** | 0.0488 | 0.8594 | 0.9082 | 0.0545 | 0.8181 | 0.8974 |
| **Dice loss** | 0.4489 | 0.4815 | 0.5511 | 0.5744 | 0.2275 | 0.4256 |
| **Focal+Dice loss** | 0.2020 | 0.5698 | 0.6173 | 0.2291 | 0.3860 | 0.5676 |
| **BCE loss** | 0.0066 | 0.4731 | 0.6409 | 0.0063 | 0.5312 | 0.6953 |

In this research, we conducted a comparative analysis of various combinations of loss functions applied to the U-Net model for automating the segmentation of the Gastrointestinal (GI) tract. Our results indicate that the Dice+BCE loss function produced the most promising outcomes, surpassing other combinations in terms of segmentation accuracy. The model trained with the Dice+BCE loss function achieved an Intersection over Union (IoU) score of 0.8594 and a Dice coefficient of 0.9082. The efficacy of the Dice+BCE loss function can be attributed to its capability to address different facets of segmentation accuracy. While the Dice loss emphasizes maximizing the spatial overlap between predicted and ground truth segmentation masks, the BCE loss penalizes the misclassification of individual pixels. By combining these two loss functions, the model becomes better equipped to capture both global and local features of the GI tract, resulting in enhanced segmentation performance. Our findings highlight the importance of selecting appropriate loss functions tailored to the specific characteristics of the segmentation task (see Table I).

Table II shows the Comparison of dice and IOU metrics for various models on the UW Madison GI tract Segmentation Dataset.

TABLE II. COMPARISON OF DICE AND IOU METRICS FOR VARIOUS MODELS

| No | Model | IoU coefficient | Dice Coefficient |
|---|---|---|---|
| 1 | U-Net with EfficientNet-B3 as backbone [2] | 85.3% | – |
| 2 | U-Net with Resnet50 as backbone [2] | 84.9% | – |
| 3 | U-Net with VGC16 [2] | 82.7% | – |
| 4 | U-Net [3] | – | 51% |
| 5 | Mask R-CNN [3] | – | 73% |
| 6 | Inception V3 [4] | 76.87% | 60.49% |
| 7 | SeResNet 50 [4] | 75.88% | 58.2% |
| 8 | DenseNet121 [4] | 74.86% | 55.58% |
| 9 | VGG 19 [4] | 66.46% | 47.41% |
| 10 | U-Net Model [4] | 88.1% | 88.54% |
| **11** | **Our Proposed work** | **81.81%** | **89.74%** |

## V. CONCLUSIONS

In conclusion, the proposed automated deep learning-based approach for segmenting the stomach and intestine from MRI scans in patients with gastrointestinal cancer holds significant implications for improving treatment outcomes and reducing manual effort in tumor delineation. Different loss functions, such as Dice, BCE, Dice+BCE, Focal+Tversky, and Dice+Focal are compared to the proposed U-Net model. The results indicate that the U-Net model using the dice+BCE loss function outperforms the others, showcasing its potential for accurate and efficient automated segmentation in radiation therapy planning.

The limitations of the existing network architecture are:

(1) Lack of integration with other state-of-the-art architectures may restrict its ability to effectively handle complex segmentation tasks.

(2) There's an opportunity to boost accuracy, model memory, and training speed.

We postulate that the following future developments could enhance our findings even more and overcome the limitations of the existing network architecture.

(1) The integration of U-Net and Mask R-CNN architectures, with U-Net serving as the foundational structure within Mask R-CNN. This hybrid approach has the potential to leverage the strengths of both architectures, combining U-Net's proficiency in semantic segmentation with Mask R-CNN's capabilities in instance segmentation.

(2) By augmenting the training dataset through methods such as rotation, flipping, and scaling, researchers can increase the diversity and size of the data, potentially improving the model's ability to generalize to unseen scenarios.

(3) Utilizing attention mechanisms and curriculum learning techniques can elevate model accuracy by enabling the model to focus on relevant information and learn progressively complex concepts [24]. By incorporating attention mechanisms, the model can assign varying degrees of importance to different parts of the input data, while curriculum learning facilitates gradual learning from simpler to more complex examples, collectively enhancing the model's performance.

(4) Thoroughly evaluating multiple models allows for the identification of the most suitable architecture, leading to enhanced accuracy in machine learning tasks. By comparing performance metrics across different architectures, optimal model selection is achieved, ultimately improving predictive accuracy and model effectiveness.

(5) Optimize U-Net's design by refining skip connections can improve memory efficiency and training speed, leading to faster convergence during training and improving scalability to larger datasets and more complex tasks

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## AUTHOR CONTRIBUTIONS

Bindu Madhavi Tummala, Dasari Chinna Veeraiah, and Rishitha Jaladi conducted the research; Subba Reddy Chavva, Yallamandaiah S., and A. Radhika had analyzed the data and worked on the Methodology; Bindu Madhavi Tummala and Dasari Chinna Veeraiah worked on Testing; Rishitha Jaladi and Aruna Kumari Peruri wrote the paper; all authors approved the final version.

## REFERENCES

[1] M. Arnold *et al*., "Global burden of 5 major types of gastrointestinal cancer," *Gastroenterology*, vol. 159, no. 1, pp. 335–349, 2020. https://doi.org/10.1053/j.gastro.2020.02.068

[2] M. Sharma, "Automated GI tract segmentation using deep learning," arXiv preprint, arXiv:2206.11048, 2020.

[3] A. Chou, W. Li, and E. Roman, "GI tract image segmentation with U-Net and mask R-CNN," *Image Segmentation with U-Net and Mask R-CNN*, pp. 1–9, *2022.*

[4] N. Sharma *et al*., "U-Net model with transfer learning model as a backbone for segmentation of gastrointestinal tract," *Bioengineering*, vol. 10, no. 1, 119, 2023. https://doi.org/10.3390/bioengineering10010119

[5] P. Nemani and S. Vollala, "Medical image segmentation using LEVIT-UNet++: A case study on Gi tract data," in *Proc. 2022 26th International Computer Science and Engineering Conference (ICSEC)*, 2022, pp. 7–13.

[6] X. Zhang, X. Yue, T. Liu, and S. Wu, "A GI tract image segmentation algorithm via U-Net 2.5 d network," in *Proc. 2022 3rd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*, 2022, pp. 684–687.

[7] H. Li and J. Liu, "Multi-view u-net for automated GI tract segmentation," in *Proc. 2022 5th International Conference on Pattern Recognition and Artificial Intelligence (PRAI)*, Chengdu, China, 2022, pp. 1067–1072.

[8] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference*, 2015, pp. 234–241.

[9] 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.

[10] Y. Zhang, Y. Gong, D. Cui, X. Li, and X. Shen, "Deepgi: An automated approach for gastrointestinal tract segmentation in MRI scans," arXiv preprint, arXiv:2401.15354, 2024

[11] S. Roshan *et al.*, "A deep ensemble medical image segmentation with novel sampling method and loss function," *Computers in Biology and Medicine*, 108305, 2024.

[12] H. Zhou *et al.*, "Improvement of deep learning model for gastrointestinal tract segmentation surgery," *Frontiers in Computing and Intelligent Systems*, vol. 6, no. 1, pp. 103–106, 2023.

[13] Kaggle. (2022). UW-Madison GI tract image segmentation competition. [Online]. Available: https://www.kaggle.com/competitions/uw-madison-gi-tract-image-segmentation

[14] J. J. Lagendijk, B. W. Raaymakers, and M. V. Vulpen, "The magnetic resonance imaging–linac system," *Seminars in Radiation Oncology*, vol. 24, pp. 207–209, 2014. https://doi.org/10.1016/j.semradonc.2014.02.009

[15] D. A. Jaffray and M. K. Gospodarowicz, "Radiation therapy for cancer," *Cancer: Disease Control Priorities*, vol. 3, pp. 239–248, 2015.

[16] O. Rukundo, "Effects of image size on deep learning," *Electronics*, vol. 12, no. 4, 2023.

[17] S. Jagadeesh, T. Venkateswarlu, and M. Ashok, "Run length encoding and bit mask-based data compression and decompression using verilog," *Int. J. Eng. Res. Technol*, vol. 1, no. 7, pp. 1–7, 2021.

[18] S. Suhas and C. R. Venugopal, "MRI image preprocessing and noise removal technique using linear and nonlinear filters," in *Proc. 2017 International Conference on Electrical,* Electronics, *Communication, Computer, and Optimization Techniques (ICEECCOT)*, 2017, pp. 1–4.

[19] H. S. Obaid, S. A. Dheyab, and S. S. Sabry, "The impact of data pre-processing techniques and dimensionality reduction on the accuracy of machine learning," in *Proc. 2019 9th Annual Information Technology, Electromechanical Engineering and Microelectronics Conference*, 2019, pp. 279–283.

[20] S. Jadon, "A survey of loss functions for semantic segmentation," in *Proc. 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology*, 2020, pp. 1–7.

[21] N. Huynh. (2023). Understanding evaluation metrics in medical image segmentation. medium. [Online]. Available: https://medium.com/@nghihuynh_37300/understanding-evaluation-metrics-in-medical-image-segmentation-d289a373a3f

[22] P. Schratz *et al*., "Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data," *Ecological Modelling*, 109, 2019.

[23] S. Jadon, "A survey of loss functions for semantic segmentation," in *Proc. 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology*, 2020, pp. 1–7.

[24] B. M. Tummala and S. S. Barpanda, "Curriculum learning based overcomplete U-Net for liver tumor segmentation from computed tomography images," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 3, pp. 1620–1629, 2023.