Tech Science Press

Multi-Scale Network for Thoracic Organs Segmentation

**Muhammad Ibrahim Khalil[1], Samabia Tehsin[1], Mamoona Humayun[2],
N.Z Jhanjhi[3,4,*] and Mohammed A. AlZain[5]**

[1]Department of Computer Science, Bahria University, Islamabad, Pakistan
[2]Department of Information systems, College of Computer and Information Sciences, Jouf University, KSA
[3]School of Computer Science and Engineering (SCE), Taylor's University, Malaysia
[4]Center for Smart Society 5.0 (CSS5), Faculty of Innovation and Technology (FIT), Taylor's University, Malaysia
[5]Department of Information Technology, College of Computers and Information Technology, Taif University,
Taif, 21944, Saudi Arabia
*Corresponding Author: N.Z Jhanjhi. Email: noorzaman.jhanjhi@taylors.edu.my
Received: 29 May 2021; Accepted: 01 July 2021

Abstract: Medical Imaging Segmentation is an essential technique for modern medical applications. It is the foundation of many aspects of clinical diagnosis, oncology, and computer-integrated surgical intervention. Although significant successes have been achieved in the segmentation of medical images, DL (deep learning) approaches. Manual delineation of OARs (organs at risk) is vastly dominant but it is prone to errors given the complex irregularities in shape, low texture diversity between tissues and adjacent blood area, patient-wide location of organisms, and weak soft tissue contrast across adjacent organs in CT images. Till now several models have been implemented on multi organs segmentation but not caters to the problem of imbalanced classes some organs have relatively small pixels as compared to others. To segment OARs in thoracic CT images, we proposed the model based on the encoder-decoder approach using transfer learning with the efficientnetB7 DL model. We have built a fully connected CNN (Convolutional Neural network) having 5 layers of encoding and 5 layers of decoding with efficientnetB7 specifically to tackle imbalance class pixels in an accurate way for the segmentation of OARs. Proposed methodology achieves 0.93405 IOU score, 0.95138 F1 score and class-wise dice score for esophagus 0.92466, trachea 0.94257, heart 0.95038, aorta 0.9351 and background 0.99891. The results showed that our proposed framework can be segmented organs accurately.

Keywords: Deep learning; convolutional neural network; computed tomography; organs at risk; computer-aided diagnostic

1 Introduction

Imaging is essential in biomedical research, both in hospitals and in clinical studies. Imaging helps in diagnostics, prognosis, clinical preparation, and consultation. Medical image segmentation is of great significance in medical imaging, as various areas of image processing, object detection,

and visual observation are intersecting. In the image processing field, segmentation, the isolation of the objects of interest from their background and with each other seems to be a central research mechanism by which new approaches are being formed [1]. In diagnostic imaging, automatic delineation of various image attributes is used for the analysis of morphological characters and forms of tissues, the spatial pattern of structure and action, as well as biological aspects, are addressed. Segmentation is often used as an intermediary method for representation and compression. Typically, an object's segmentation is accomplished by either defining all pixels or vertices associated with the object or by labeling those that form its borders. The latter is primarily dependent on pixel strength, but other characteristics, such as form, which can be contrasted with each pixel, are also employed for feature extraction [2].

Since segmentation involves pixel classification, it is also used as an issue in pattern detection and is approached through associated methods. Special interest in medical the field of imaging diagnostic, the data uncertainty can be high, is provided to patterns image enhancement that offers flexibility and easy automation as shown in Fig. 1. In certain applications for medical image processing, the appearance of numerous structures of varying properties indicates the use of a specially constructed series of several segmentation methods [3]. Primary steps may use simple data-reduction practices, and subsequent steps may be supplemented by more complex methods that are efficient however time-taking. Usually, the accurate selection of methods and their direction relies on both the challenge and the computational assets. Several segmentation ways may have complexities on the borderline between various tissue categories in CT scan images. This may be attributed to the circumstance that a voxel may contain a combination of constituents. Segmentation with anomaly detection offers a single approach based on the membership features. Another option is to model each feature space as an area and to measure the proportion of material properties in each input image [4].
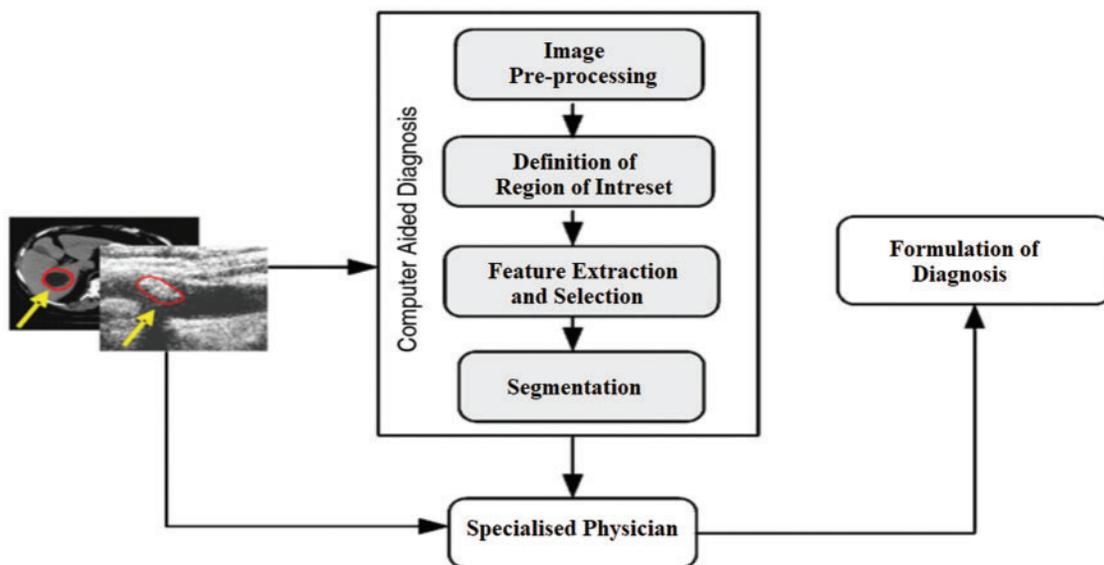
Figure 1: Medical image segmentation process flow

Clinical image segmentation plays a crucial part in computer-aided diagnostic (CAD) methods with various uses. The enormous speculation and advancement of imaging aspects such as tomography, X-ray, ultrasound, positron emission tomography, computed tomography (CT), and magnetic resonance imaging (MRI) draw researchers to introduce new process control processes. Image segmentation is known to be the utmost important health imaging procedure since it removes the region of interest (ROI) using semi-automatic and autonomous methods. It distinguishes the image into parts dependent on a basic classification, like the body tissues and organ segmentation in boundary detection, disease recognition, and object detection applications. Segmentation segments of the image into cohesive regions, clustering techniques may be extended to subdivision by removing the overall features of the object to efficiently distinguish ROI from the object context [5].

In addition, intensity information and variation are being used for image segmentation. Various border-focused segmentation methods, such as the deformable model, could be used, whereas other frameworks involve regional strategies such as regional mergers and regional extension. The development of smart advanced medical image segmentation techniques is becoming a hotspot that leads to research methods by using image segmentation information from both regions and borders for efficient borderline-based and ROI-centered segmentation. These strategies involve graph-based methods, such as graph-cut segmentation, which is iterated by selecting seed points containing certain pixels that belong to the target and other pixels from the background to separate the target and the background in the image by explicit segmentation restrictions [6].

One of the most complex uses for medical image segmentation is the segmentation of skin lesions. However, dermoscopic images include varying types of lesions of a variety of objects, such as blurry lesion boundaries and features of their irregularity, surface folds, feathers, bubbles, within multi-colored zones, and low comparison between both the lesion and the skin regions around them. Picture segmentation is a complex, dynamic process affected by a wide range of aspects, including noise, poor contrast, illumination, and irregularity of the target boundary. Moreover, in the development of electronic clinical CADs to encourage dermatologists, the segmentation of skin cancer in dermoscopic images plays a major role. However, the segmentation of skin cancer is complicated due to differences in lesion sizes, patterns, dimensions, outlines, and color [7].

In diagnostic imaging, the segmentation approach for the extraction of features, the calculation of images, and the viewing of images. The most widely used segmentation approaches are classified into two different types of regional segmentation techniques that search regions that follow certain homogeneity criteria, and edge-based image segmentation which searches for edges between regions with different properties.

The article is organized as: Section 1 describes the detailed overview of medical image segmentation. In Section 2 is about the existing work on multi-organ segmentation. Section 3, presents the detailed operation of the proposed technique that segments the organs at OAR from abdominal CT images. Section 4, presents the analysis and result work. The last section concludes this work and provides future directions.

2 Literature Review

Through the use of diagnostic methods as biomarkers is now an increasingly important area of study. Specifically, for predictive tracking of disease development, individualized therapy preparation, and drug reaction monitoring, the computer-assisted evaluation of biological image data has changed to the focus of concern as a matter of moderate research program. Major groups,

such as the healthcare and medical device industry, contract testing agencies, and institutions, are designing methods to engage in a targeted way in providing effective assurance for prescription drug or surgical trials focused on clinical imaging. Segmentation by automation is ideal. However, this is challenging to do, since the nuanced reasoning skills that have just been described can hardly be translated to computer programmers [8].

A successful approach to fix this issue is to present demand image data as inputs to automated image segmentation applications, thus compensating for the lot of quality computer image processing capability. The extraction and development of "multispectral" datasets, which form the base of the segmentation method, is a potential application of this method. Data processing can be done by two separate methods [9]. The first attempts to describe the features of the multidimensional data distribution of unlabeled feature vectors, i.e., without a clear description of the data concerning the segmentation groups. Refers to this method as unsupervised clustering (UC). The second strategy includes the labeled data, i.e., the learning method employs both the feature vector itself and the goal function to determine its representation of the segmentation attribute values. This method is like studying with a tutor. Research considers this a supervised classification (SC) [10].

More complex models of partitioning cannot generalize to real-world datasets from clinical parameters. Researchers in the medical imaging community have been aggressively pursuing approaches in response to this problem, resulting in a broad and productive range of tools with clear capabilities to manage unique and poor annotations for the role of segmenting medical images [11,12]. Recent approaches to the growing shortcomings of the medical image analysis repositories allow the research team to look at solutions to less researched datasets' challenges.

The literature on medical imaging has seen steady improvements in the design and implementation of deep convolutional methods for segmentation are shown in Tab. 1. Popular shortcomings of the image segmentation data sets include inconsistent annotations and there is only partial data on learning annotations, the training set either contains sparse annotations, noisy annotations with error rates.

Table 1: Review of existing work on multi-class segmentation

Paper	Year	Methodology	Dataset	Results
[13]	2020	UNet	CT scans hospital dataset	Dice Scores for bladder 0.924, femoral head left 0.906, bone marrow 0.854, small intestine 0.833, femoral head right 0.900, rectum 0.791, spinal cord 0.827
[14]	2018	U-Net combo, V-Net combo	MRI prostate scans, Ultrasound echocardiography	Dice score on MRI 0.90, dice score on Ultrasound 0.92
[15]	2019	U-Net-based MTL	GLIOMA NUCLEI	IOU 0.59, F1 Score 0.79
[16]	2019	Organ localization with boundary sensitive MTL segmentation	Prostate cancer patients	Prostate 0.89, Bladder 0.94, Rectum 0.89

(Continued)

Table 1: Continued

Paper	Year	Methodology	Dataset	Results
[17]	2019	2D dilated residual U-Net	SegThor2019	Dice rate for Esophagus 0.858, Trachea 0.926, Heart 0.941, Aorta 0.938
[18]	2019	VNet	SegThor2019	Esophagus 0.8597, Trachea 0.9217, Heart 0.9459, Aorta 0.9433
[19]	2020	MTL with ResNet and DenseNet	SegThor2019	Esophagus 0.8651, Trachea 0.9276, Heart 0.9536, Aorta 0.9464
[20]	2020	UNET + with EfficientnetB1	EAD2020	F1 Score 0.60

The networks that containing residual block has the main features that have enabled it to perform better. The feature maintains a steady data flow between both the incoming and outgoing images [21,22]. These methods are among the most effective architectures in this medical field [23]. Due to the existence of broad annotated libraries, significant efforts have been made to resolve the segmentation of natural features in photographs. For example, ImageNet with over 14 million manually annotated images, including over a million framed annotations. Due in part to certain challenges the distortion of the clinical image becomes harder. Medical images, on the other hand, contain complexity on segmentation, due to low contrast and noise. The task of segmentation is difficult for the following reasons the shape and location of each member on the CT sections vary considerably between patients and CT images have poor contrast properties and can be missing.

Many researchers have been investigating the application of CNN to medical image analysis with the exciting potential of CNN to perform image detection and pattern classification. The basic principle is to do segmentation using input data and adding filters to it. Images get transferred to the CNN input layer on different spatial channels. Their results showed greater efficiency than many of those that used a single data modality. Through the comprehensive study, it is found that most of the existing techniques that work on the SegTHOR dataset not cater to the problem of the unbalanced classes do not work. Four classes are the esophagus, heart, trachea, and aorta. Some members have relatively small pixels compared to other devices, there is a high gap in results towards accurate results. A better DL model is required explicitly for the section OARs into thoracic CT images to tackle unbalanced class pixels in a precise way to segment OARs.

3 Proposed Methodology

The proposed methodology consists of the four modules in the first module converts the dataset raw images, the second module performs the dataset pre-processing to normalize the data, in the third module includes the proposed model based on the encoder-decoder that contains 5 encoding layer and 5 decoding layers using transfer learning with efficientnetb7 as a backbone model to extract the powerful features to speed up the training process as shown in Fig. 2, the activation RELU is used and Adam as an optimizer, for multi-class the softmax activation function added at the output layer. Lastly, the evaluation performance is calculated using IOU, F1 score and dice score is calculated for all classes. The proposed DL model consists of 74,736,118 trainable parameters and 312,704 non-trainable parameters the total parameters are 75,048,822.
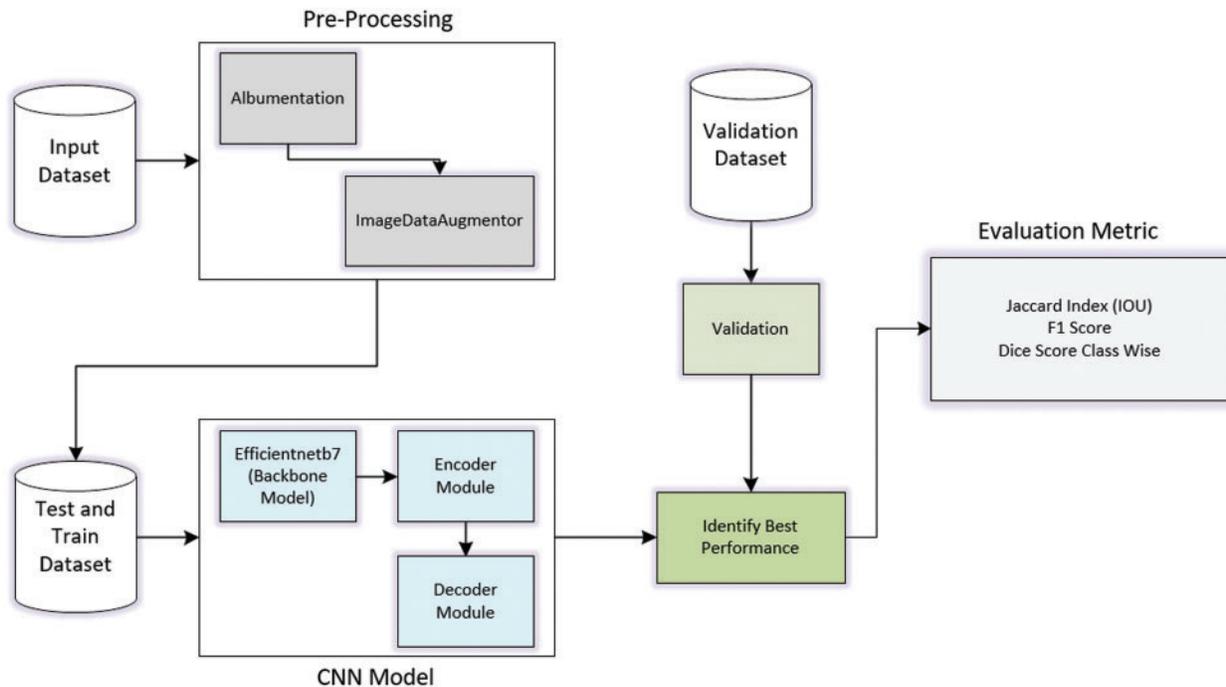
Figure 2: Proposed model process flow

This Medical Image Segmentation uses the Anaconda platform of Jupyter Notebook IDE for pre-processing works and GPU-based training using NVIDIA Tesla K80 GPU with 25GB RAM and all the open-source libraries are being used provided by Google. The libraries included Efficient, Albumentation, and ImageDataAugmentor modules. The purpose to convert into png format because, in this format, the background does exist it only contains foreground information.

The total images have 11084 of 60 patients, now it has been split into training and testing in the ratio of 80%, 20% respectively. The image size has width, height, and depth $512 \times 512 \times 3$ respectively from the input image and randomly mirrors it along the first two axes. Then apply Data Normalization so that the features can be extracted and finally apply intensity shift augmentation. The augmentation for visualization of the data include Horizontal Flip = 0.5, Perspective Rotation = 0.1, Gaussian Noise = 0.2, Gamma = 1 and Hue Saturation = 1. For image sharpness, a Gaussian filter has been applied to minimize the blur images to verify the correctness of the features and orientation in the images. The augmentation is performed to enhance the diversity of the training data and to address the problem of class imbalance. It increases the robustness of the model.

Propose a method based on the Encoder and Decoder technique for the segmentation of CT images uses transfer learning with efficientnetB7 DL model. Designed a fully connected CNN having 5 layers of encoding and 5 layers of decoding. The backbone model EfficientNet is designed with 1000 class labels for the ImageNet classification. The network is built on the latest EfficientNet, which has reached state-of-the-art high accuracy on ImageNet. EfficientNet is very smaller in comparison with other models that reach comparable ImageNet accuracy. For example,

the ResNet50 structure has a total of 23,534,592 parameters and yet performs less as compared to the smallest EfficientNet, which only takes a total of 5,330,564 parameters as shown in Fig. 3.
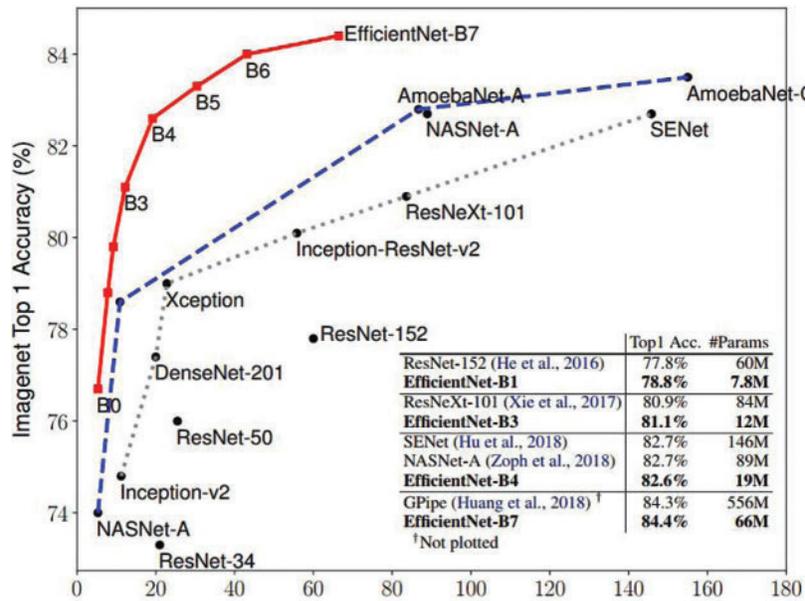
Figure 3: DL models parameter comparison [24]

The EfficientNet models are built on basic and extremely efficient compounded scaling approaches. This approach allows you to scale up the ConvNet baseline to any target limited resources while retaining the model utility used for the transfer of learning datasets. In general, EfficientNet versions achieve both higher accuracy and improved performance over current CNNs such as AlexNet, GoogleNet, and MobileNetV2. EfficientNet contains versions from B0 to B7, each with various parameters from 5.3 M to 66 M. Using transfer learning may prove useful for segmentation problems, although as all the other new problems require entirely separate classes as shown in Fig. 4.

$$Width \ w = \beta^{\varnothing}, \quad Depth \ d = \alpha^{\varnothing} \quad and \quad Resolution \ R = \gamma^{\varnothing} \tag{1}$$

$$\alpha \geq 1, \quad \beta \geq 1, \quad \gamma \geq 1 \tag{2}$$

The technique of compound scaling is premised on the concept of balancing with a fixed value to balance the proportions of depth, width, and resolution. Where α, β and γ are defined by employing the algorithm of the grid search and $\varnothing$ identifies the network computational resources.

The encoding and decoding layer having different kernel sizes in which for Encoding layers kernel size is 3 and filter size are 16, 32, 64, 128, and 256 with channel 3, and for decoding layers kernel size is 4 and filter size is 256, 128, 64, 32 and 16 with channel 3, the number of channels increases due to the growing number of channels across consecutive spatial resolution levels. Encoding and decoding blocks are used to make the model robust as shown in Fig. 5. The number of slices in the range of 150 to 220.
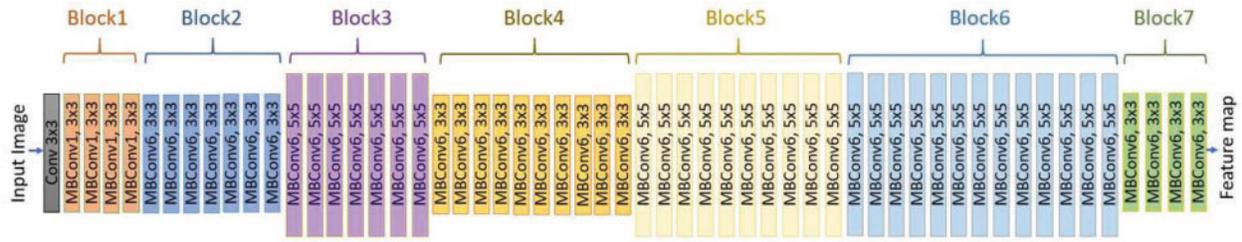
Figure 4: Efficientnetb7 model architecture [25]
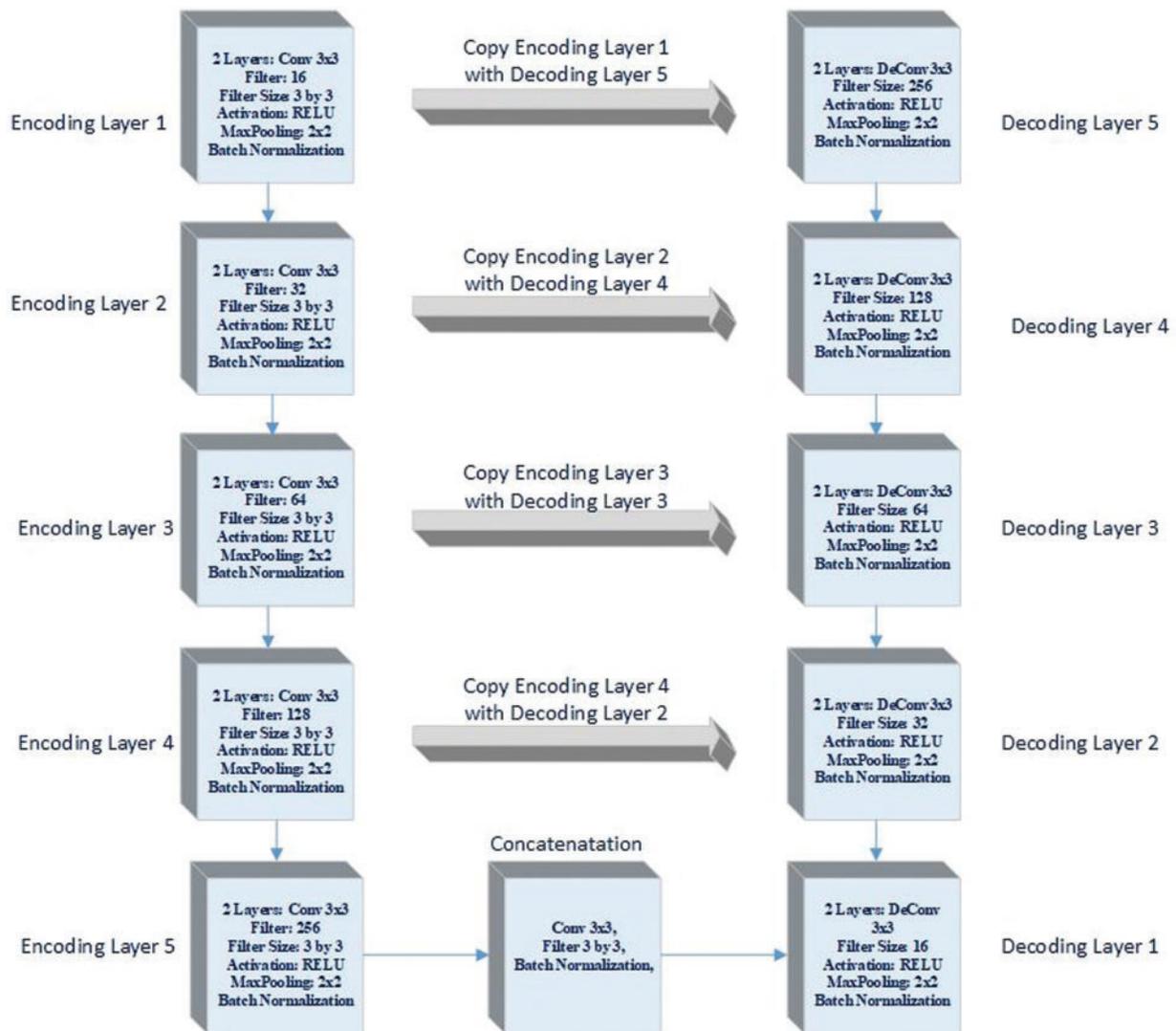
Figure 5: Encoder and decoder layers architecture

The efficient-net models are productive with substantially fewer numbers of parameters and give better results. The selection of Optimizer Learning rate was 0.0001 and decay rate was 10–6 "Adam optimizer" for improving the performance of the model and the activation function "softmax" has been selected due to problem multi-class segmentation. The set of the number of epochs is 60 and batch size 2. During training, the evaluation metric dice score and IOU score, and F1 score are calculated. In this medical image segmentation, the technique used the custom model of Convolution Neural Network with "Encoding and Decoding layers" having activation function "Relu" and same padding with transfer learning using "efficientnetB7" DL. In this model for compiling, used the "softmax" activation function and "Adam" for optimizer and for retrieving the dice score, uses the Keras segmentation dice score functions.

4 Experiment and Results

The experiments were carried out using the SegTHor dataset using the proposed DL model based on the encoder-decoder approach with transfer learning using the efficientnetb7 model. The model works efficiently to segment OAR classes' esophagus, heart, trachea, and aorta. The image size is $512 \times 512 \times 3$ width, height, and depth, respectively, from the input image and mirrors it uniformly around the first two axes. The dataset is normalized using data augmentation. The proposed model trained for the 60 epochs and each epoch has 2271 iterations. The performance metric measured IOU score achieved is 0.93405 and the overall F1 score is 0.95138. Our proposed methodology archives the better dice score for the less predicated classes like Esophagus and Trachea. Also added the background as a class and measured the dice score for the background that is not considered in the previous studies on the SegThor dataset. The organ having smaller sizes were difficult to segment.

Dice Coefficient and IOU metrics are the most well-known measures for segmentation evaluation and the researchers use this evaluation metric on the SegTHor dataset for experimentation work. The DC mostly used computer scientists to check the similarity between predicted segmentation map and ground truth segmentation map and widely used in the medical segmentation community.

The Jaccard index is an evaluation metric for quantifying the percentage overlap between the targeted mask and the forecast results. The mean IOU has been used to evaluates the overall similarities between the items that show how the organs are overlapping between the targeted and predicted regions.

$$\text{Jaccard Index} = \frac{Area\ of\ Overlap}{Area\ of\ Union}$$

$$IoU = \frac{|target \cap prediction|}{|target \cup prediction|}$$

$$IoU = \frac{TP}{TP + FP + FN} \tag{3}$$

where TP refers to true positive, FP represents false positives, TN represents true negative and FN represents false negative. The dice coefficient is similar to the Jaccard index and is commonly used as a segmentation evaluation metric.

$$Dice\ Cofficient\ (F1\ Score) = \frac{2 * Area\ of\ Overlap}{Sum\ of\ Areas}$$

$$Dice\ Cofficient = \frac{2*|target \cap prediction|}{|target|+|prediction|}$$

$$Dice\ Cofficient = \frac{2\,TP}{2\,TP+FP+FN} \tag{4}$$

It annotates the ground truth area in your picture and then creates an automatic algorithm to use it. Validate the methodology by measuring the Dice score, which is an indicator of how close the items are. So that is the amount of overlap between the two segmentation variables separated by the overall similarities between the two items.

The ISBI SegTHOR challenge the director to issue computed CT photographs of 60 patients in the medical reports. The size of the CT scans is 512 × 512 pixels, which lies in the plane resolution changing from 0.90 mm per image to 1.37 mm. The volume of slices with a z-resolution between 2 and 3.7 mm ranges between 150 and 284. The most commonly used resolution is $0.98 \times 0.98 \times 2.5$ mm^3. The dataset SegTHOR consists of (60 patients) were divided randomly into the training sample of 40 patients that includes (7390 slices) and 20 patients (3694 slices) tested as shown in Tab. 2. An experienced radiation oncologist has delineated the ground truth for OARs [26].

Table 2: Dataset distribution statistics

Dataset	CT Images			Image Slices		
	No. of patients	CT scan size	Resolution	No. of slices per patient	Resolution	Total no. of slices
Training	40 patients	512 × 512 pixels	0.90–1.37 mm per pixel	from 150 to 284	The z-resolution interpolated 2 and 3.7 mm	7390
Testing	20 patients	512 × 512 pixels	0.90–1.37 mm per pixel	from 150 to 284	The z-resolution interpolated 2 and 3.7 mm	3694
Total	60 patients			11084 slices		

In the ISBI SegTHOR dataset, the heart is represented by green, the esophagus with red, the trachea with blue, and the aorta with yellow. The challenge was indeed a multi-class segmentation that every participant in the dataset involved all classes (trachea, heart esophagus, and aorta) as shown in Fig. 6. Some organs have small in size and the CT images are difficult challenges due to the poor quality of pixel values that represent other tissues. The proposed model works efficiently and predicts OARs more accurately.

The proposed methodology is based on the encoder and decoder approach having five up-sampling and down-sampling layers the performance analyzed on multiple DL models as a feature extractor the performance is measured by IOU and F1 scores are shown in Tab. 3 and Fig. 7.

The efficacy of the proposed framework is analyzed in the SegTHOR challenge dataset using transfer learning with efficientnetb7 model, dice score for all classes is calculated because in medical image segmentation dice score is the commonly used evaluation metric the score for Esophagus, Trachea, Aorta, Heart and. Background. The proposed methodology results compared with the previous approaches that are used in the multi-class image segmentation are shown in Tab. 4.
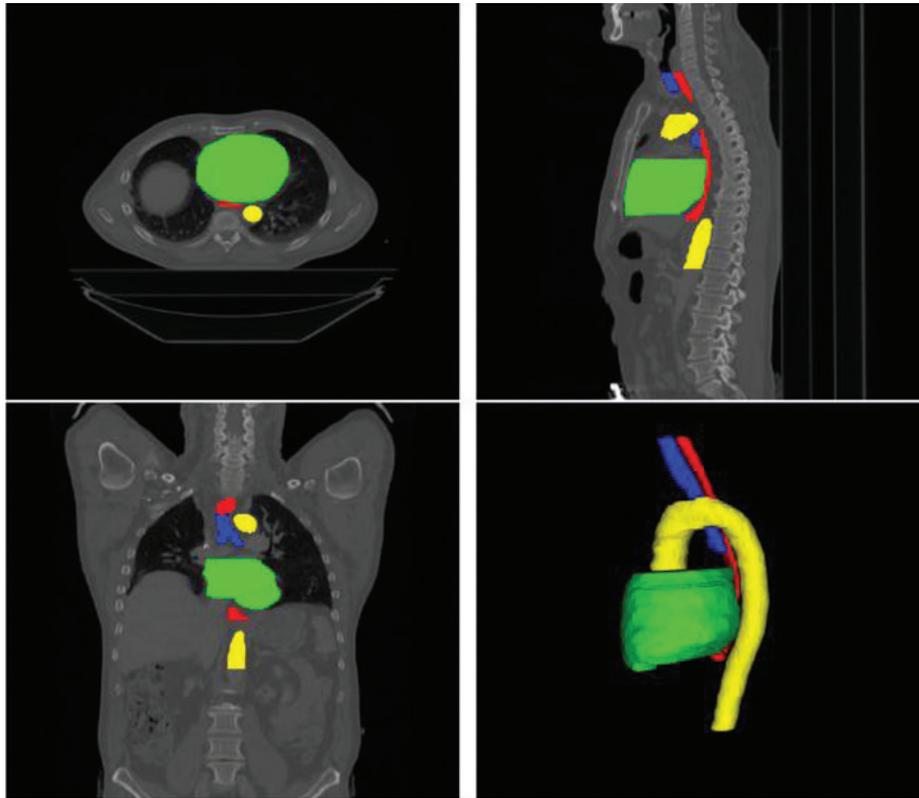
Figure 6: CT images with axial, sagittal, and coronal views

Table 3: Results on SegThor multi-class image segmentation

Approach	Jaccard index (IOU)	F1 score
Encoding decoding with inceptionresnetv2	0.7083	0.7347
Encoding decoding with resnet50	0.6176	0.6309
Encoding decoding with mobilenetv2	0.8841	0.9037
Encoding decoding with efficientnetb6	0.8945	0.9275
Proposed methodology	0.93405	0.95138

The results of the proposed DL model for automated multi-class segmentation of OAR have a significant role in medical imaging. The proposed methodology based on the encoder-decoder

approach with transfer learning using the efficientnetb7 model achieves the 0.93405 IOU score and 0.95138 F1 scores. Classes with better pixels already archive good results. In the proposed methodology the dice score for the esophagus, trachea, heart, aorta, and background are 0.92466, 0.94257, 0.95038, 0.9351and 0.99891 as shown in Fig. 8. The proposed model achieves better accuracy on small classes as compared to the previous methods and caters to the problem of imbalanced classes.
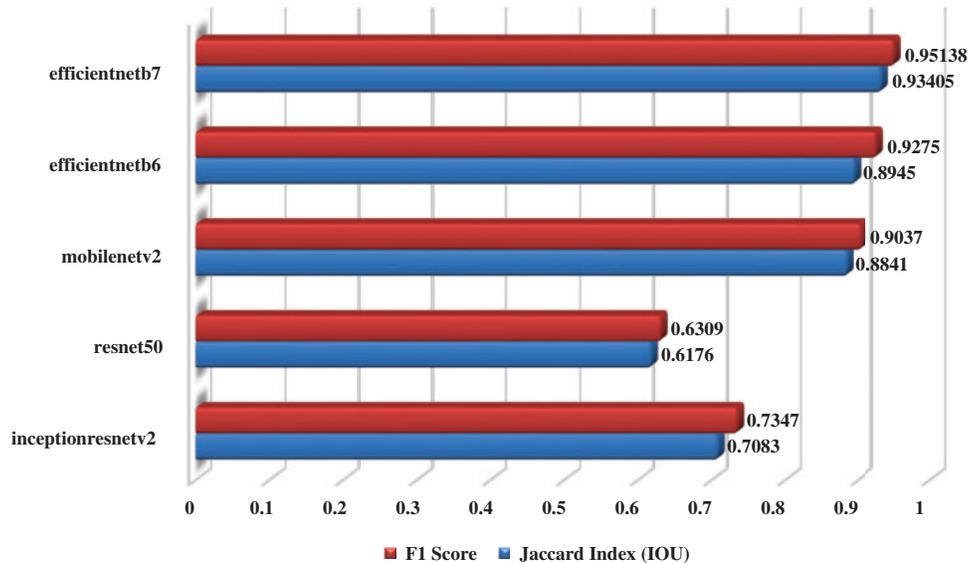
Figure 7: Results comparision on SegThor multi-class image segmentation

Table 4: Dice score comparison with existing approaches

Paper	Methodology	Dataset	Esophagus	Heart	Trachea	Aorta	Background
Vesal et al. [17]	2D dilated residual U-Net	SegTHOR	0.858	0.941	0.926	0.938	–
Wang et al. [18]	VNet	SegTHOR	0.8597	0.9459	0.9217	0.9433	–
He et al. [19]	MTL with ResNet and DenseNet	SegTHOR	0.8651	0.9536	0.9276	0.9464	–
Proposed methodology	Enocder decoder with efficientnetb7	SegTHOR	0.92466	0.95038	0.94257	0.9351	0.99891

5 Discussion

In this research, we introduced a DL approach using transfer learning that is capable to cater the problem of imbalanced datasets and segment organs accurately. The proposed methodology is based on the encoder-decoder approach using transfer learning with efficientnetb7.

The segmentation of multi-classes regarded as difficult due to issues with interclass function similarities and class imbalance issues. Binary techniques, as opposed to multiclass approaches,

are often more stable and attain better accuracy. In contrast to binary segmentation, multiclass segmentation does not produce conflicting labels. The key purpose of the method of segmentation is to segment the image into regions that are identical to one or more features and characteristics. Rapid developments in the world of medical imaging are revolutionizing medicines. Determination of the presence or seriousness of the condition will affect the health treatment of the patient or the status of the results of the research.

	Dice Score
◼ Esophagus	0.92466
◼ Heart	0.95038
◼ Trachea	0.94257
◼ Aorta	0.9351
◼ Background	0.99891

◼ Esophagus ◼ Heart ◼ Trachea ◼ Aorta ◼ Background
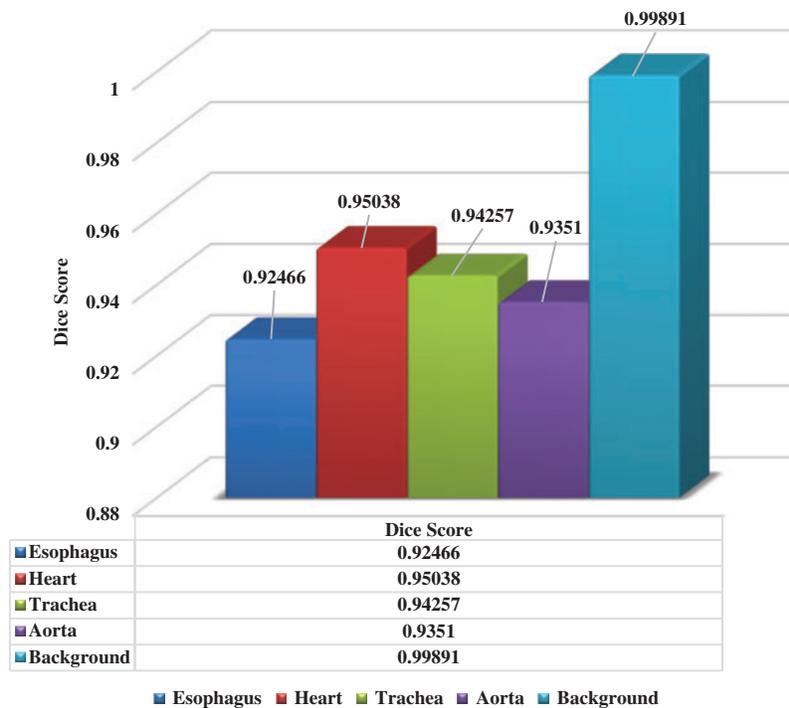
Figure 8: Class wise deice score

DL approaches are increasingly being adapted for disease identification and diagnosis in diverse clinical environments using computer-assisted prediction and identification. In other biomedical databases, our proposed network is capable of multi-class segmentation.

Because of the computing complexity and inadequate amount of GPU memory, the model was trained with a batch size of two and uses the encoder-decoder blocks up to level five because the deeper structure's feature maps are relatively low in resolution. To reduce computing costs and get better results, pre-processed input is supplied into the network.

Moreover, from a clinical point of view segmentation have a great role in the medical images for diagnosing disease. Clinical segmentation of organs around tumors tends to compensate for the unavoidable variation of patient location and anatomy, allowing for integrated and computer-assisted radiotherapy. The most major element is to enhance the distinction between regular and abnormal.

6 Conclusion

The segmentation of medical images is of enormous significance in the analysis of medical images, in the combination of many factors of image processing and computer vision. A medical practitioner shall manually carry out the delineation process. Such a time-consuming strategy is often vulnerable to an inexpensive degree of inaccuracy that may lead to missed tumor areas. The proposed solution segments the organs-at-risk (OAR) centered on the encoding and decoding method for extracting distinguishing characteristics with transfer learning with the efficientnetb7 model. The results demonstrated that our framework can be segmented organs accurately. Dice score is evaluated for the heart, trachea, esophagus, and aorta. The key purpose of this research is to address the problem of imbalanced classes and increase the dice score of the less predicted classes that have a lesser number of pixels like the trachea and esophagus as compared to other classes like the heart and aorta. The proposed methodology performs segmentation accurately on OAR.

7 Future Work

In the future, complex approaches will be investigated with various parameters that will be integrated into the proposed system to increase the further accuracy of all classes. Also, refer to the other multi-class segmentation activities. Delineation strategies may help to refine the borders of organs and to generalize the model, train on other medical datasets in the domain of biomedical.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] S. Devunooru, A. Alsadoon, P. Chandana and A. Beg, "Deep learning neural networks for medical image segmentation of brain tumours for diagnosis: A recent review and taxonomy," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 1, pp. 455–483, 2021.

[2] A. Sinha and J. Dolz, "Multi-scale self-guided attention for medical image segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 1, pp. 121–130, 2021.

[3] M. H. Hesamian, W. Jia, X. He and P. Kennedy, "Deep learning techniques for medical image segmentation: Achievements and challenges," *Journal of Digital Imaging*, vol. 32, no. 4, pp. 582–596, 2019.

[4] S. Feng, H. Zhao, F. Shi, X. Cheng, M. Wang *et al.,* "CPFNnet: Context pyramid fusion network for medical image segmentation," *IEEE Transactions on Medical Imaging*, vol. 39, no. 10, pp. 3008–3018, 2020.

[5] Y. Weng, T. Zhou, Y. Li and X. Qiu, "Nas-unet: Neural architecture search for medical image segmentation," *IEEE Access*, vol. 7, no. 1, pp. 44247–44257, 2019.

[6] Z. Guo, X. Li, H. Huang, N. Guo and Q. Li, "Deep learning-based image segmentation on multimodal medical imaging," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 3, no. 2, pp. 162–169, 2019.

[7] G. Wang, W. Li, M. A. Zuluaga, R. Pratt, P. A. Patel *et al.,* "Interactive medical image segmentation using deep learning with image-specific fine tuning," *IEEE Transactions on Medical Imaging*, vol. 37, no. 7, pp. 1562–1573, 2018.

[8]   S. A. Taghanaki, K. Abhishek, J. P. Cohen, J. Cohen-Adad and G. Hamarneh, "Deep semantic segmentation of natural and medical images: A review," *Artificial Intelligence Review*, vol. 54, no. 1, pp. 137–178, 2021.

[9]   R. Gu, G. Wang, T. Song, R. Huang, M. Aertsen *et al.,* "CA-Net: Comprehensive attention convolutional neural networks for explainable medical image segmentation," *IEEE Transactions on Medical Imaging*, vol. 40, no. 2, pp. 699–711, 2021.

[10] S. Randhawa, A. Alsadoon, P. Prasad, T. Al-Dala'in, A. Dawoud *et al.,* "Deep learning for liver tumour classification: Enhanced loss function," *Multimedia Tools and Applications*, vol. 80, no. 3, pp. 4729–4750, 2021.

[11] F. Ahmad, S. N. Almuayqil, M. Humayun, S. Naseem, W. A. Khan *et al.,* "Prediction of covid-19 cases using machine learning for effective publichealth management," *Computers, Materials and Continua*, vol. 66, no. 3, pp. 2261–2282, 2021.

[12] D. Muller and F. Kramer, "Miscnn: A framework for medical image segmentation with convolutional neural networks and deep learning," *BMC Medical Imaging*, vol. 21, no. 1, pp. 1–11, 2021.

[13] Z. Liu, X. Liu, B. Xiao, S. Wang, Z. Miao *et al.,* "Segmentation of organs-at-risk in cervical cancer CT images with a convolutional neural network," *Physica Medica*, vol. 69, no. 1, pp. 184–191, 2020.

[14] S. A. Taghanaki, Y. Zheng, S. K. Zhou, B. Georgescu, P. Sharma *et al.,* "Combo loss: Handling input and output imbalance in multi-organ segmentation," *Computerized Medical Imaging and Graphics*, vol. 75, no. 1, pp. 24–33, 2019.

[15] X. Li, Y. Wang, Q. Tang, Z. Fan and J. Yu, "Dual U-net for the segmentation of overlapping glioma nuclei," *IEEE Access*, vol. 7, no. 1, pp. 84040–84052, 2019.

[16] S. Wang, K. He, D. Nie, S. Zhou, Y. Gao *et al.,* "CT male pelvic organ segmentation using fully convolutional networks with boundary sensitive representation," *Medical Image Analysis*, vol. 54, no. 1, pp. 168–178, 2019.

[17] S. Vesal, N. Ravikumar and A. Maier, "A 2D dilated residual u-net for multi-organ segmentation in thoracic CT," arXiv preprint arXiv:1905.07710, 2019.

[18] Q. Wang, W. Zhao, C. H. Zhang, L. Zhang, C. Wang *et al.,* "3D enhanced multi-scale network for thoracic organs segmentation," *SegTHOR@ ISBI*, vol. 3, no. 1, pp. 1–5, 2019.

[19] T. He, J. Hu, Y. Song, J. Guo and Z. Yi, "Multi-task learning for the segmentation of organs at risk with label dependence," *Medical Image Analysis*, vol. 61, no. 1, pp. 1–11, 2020.

[20] N. B. Le and Duy Huynh, "A U-net++ with pre-trained efficientnet backbone for segmentation of diseases and artifacts in endoscopy images and videos," *EPITA Research and Development Laboratory*, vol. 2595, no. 1, pp. 14–16, 2020.

[21] L. Tao, J. Fisher, E. Anaya, X. Li and C. S. Levin, "Pseudo CT image synthesis and bone segmentation from MR images using adversarial networks with residual blocks for MR-based attenuation correction of brain pet data," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 5, no. 2, pp. 193–201, 2021.

[22] D. Wang, G. Hu and C. Lyu, "Frnet: An end-to-end feature refinement neural network for medical image segmentation," *The Visual Computer*, vol. 37, no. 1, pp. 1101–1112, 2021.

[23] K.-K. Tseng, R. Zhang, C.-M. Chena and M. M. Hassan, "DNetUnet: A semi-supervised CNN of medical image segmentation for super-computing AI service," *The Journal of Supercomputing*, vol. 77, no. 4, pp. 3594–3615, 2021.

[24] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *Proc. PMLR*, California, USA, pp. 6105–6114, 2019.

[25] B. Baheti, S. Innani, S. Gajre and S. Talbar, "Eff-unet: A novel architecture for semantic segmentation in unstructured environment," in *Proc. CVPR*, Seattle, WA, USA, pp. 358–359, 2020.

[26] Z. Lambert, C. Petitjean, B. Dubray and S. Kuan, "Segthor: Segmentation of thoracic organs at risk in CT images," in *Proc. IPTA*, Paris, France, pp. 1–6, 2020.