Breast Cancer Segmentation from Ultrasound Images Using ResNext-based U-Net Model

Oğuzhan KATAR[1*], Özal YILDIRIM[1]

[1]Department of Software Engineering, Faculty of Technology, Firat University, Elazig, Turkey
(ORCID: 0000-0002-5628-3543) (ORCID: 0000-0001-5375-3012)

Keywords: Breast cancer, Ultrasound images, Semantic segmentation, U-Net, Disease detection, Deep learning.

Abstract

Breast cancer is a type of cancer caused by the uncontrolled growth and proliferation of cells in the breast tissue. Differentiating between benign and malignant tumors is critical in the detection and treatment of breast cancer. Traditional methods of cancer detection by manual analysis of radiological images are time-consuming and error-prone due to human factors. Modern approaches based on image classifier deep learning models provide significant results in disease detection, but are not suitable for clinical use due to their black-box structure. This paper presents a semantic segmentation method for breast cancer detection from ultrasound images. First, an ultrasound image of any resolution is divided into 256×256 pixel patches by passing it through an image cropping function. These patches are sequentially numbered and given as input to the model. Features are extracted from the 256×256 pixel patches with pre-trained ResNext models placed in the encoder network of the U-Net model. These features are processed in the default decoder network of the U-Net model and estimated at the output with three different pixel values: benign tumor areas (1), malignant tumor areas (2) and background areas (0). The prediction masks obtained at the output of the decoder network are combined sequentially to obtain the final prediction mask. The proposed method is validated on a publicly available dataset of 780 ultrasound images of female patients. The ResNext-based U-Net model achieved 73.17% intersection over union (IoU) and 83.42% dice coefficient (DC) on the test images. ResNext-based U-Net models perform better than the default U-Net model. Experts could use the proposed pixel-based segmentation method for breast cancer diagnosis and monitoring.

1. Introduction

Cancer is a disease resulting from the accumulation of genetic mutations or aberrant alterations in cellular genes [1]. In a healthy body, cells follow a regulated cycle, but mutations can disrupt this balance, leading to uncontrolled cell growth [2]. These mutations can be inherited or occur due to environmental factors like radiation, chemicals, or viruses [3].

When a gene loses its normal function or gains a new abnormal function due to a mutation, it can disrupt the complex signaling pathways that regulate cell growth and division. As a result, the affected cells gain the ability to divide uncontrollably and form a mass of abnormal cells called a tumor [4]. Tumors can be broadly classified into two main categories: benign tumors and malignant tumors [5]. Benign tumors are usually not cancerous and do not pose an immediate threat to health. The cells in benign tumors are very similar to normal cells and their growth rate tends to be relatively slow. Importantly, they do not invade nearby tissues or spread to other parts of the body [6]. Benign tumors are therefore considered localized and do not usually lead to life-threatening conditions. In contrast, malignant tumors become cancerous and have the potential to cause

---

serious harm to an individual's health. If left untreated, malignant tumors can invade surrounding tissues and organs and have the ability to metastasize, meaning they can spread to distant parts of the body through the bloodstream or lymphatic system [7].

Breast cancer is the most frequently diagnosed cancer among women worldwide [8]. It originates in the cells of the breasts, particularly in the ducts or lobules. As the disease progresses, malignant cells derived from these areas possess the ability to invade neighboring healthy tissues. Moreover, if cancer cells infiltrate the lymph nodes, they can easily spread via the lymphatic system to distant organs such as the bones, lungs, liver, or brain [9]. Detecting breast cancer in its early stages is therefore crucial for successful treatment and improved outcomes [10]. Specialists use radiological images to assess the structure of breast tissue and possible tumor findings. However, manual evaluation of radiologic images is time-consuming and involves subjective decision depending on the experience and training of the specialist. To overcome these challenges, it is a popular research topic to use artificial intelligence methods to detect breast cancer findings from radiological images. Michael et al. [11] proposed a LightGBM-based method for breast cancer detection from ultrasound images. The proposed method is validated on a dataset of 912 samples. To classify malignant and benign tumors, 185 features were extracted manually. Using these features, LightGBM achieved 99.86% accuracy. González-Luna et al. [12] proposed a method for classifying breast ultrasound images as benign or malignant. The dataset used in the study contains 2032 samples, of which 1341 are benign lesions and 691 are malignant lesions. Linear discriminant analysis (LDA) outperformed other classifiers by classifying the test samples with 89.00% accuracy, 82.00% sensitivity and 93.00% specificity. Wei et al. [13] proposed a method for automatic classification of cancer from breast ultrasound images. The proposed method is based on benign and malignant lesion features. Validated on a total of 1061 different ultrasound images, the support vector machine (SVM) classifier achieved 75.94% accuracy, 66.37% sensitivity, 86.87% specificity and 85.23% precision.

The traditional machine learning approach to breast cancer detection involving feature extraction from radiological images requires feature engineering and performs poorly on complex datasets. Modern deep learning approaches based on convolutional neural networks (CNN) provide significant results for image classification studies. Atrey et al. [14] proposed a hybrid method (CNN+LSTM) for the classification of benign and malignant breast cancer

tumors. They used 43 mammogram images and 43 ultrasound images. Dataset samples were increased by using data augmentation techniques. The proposed model achieved a classification accuracy of 97.16% for mammography images and 98.84% for ultrasound images. Raza et al. [15] proposed the DeepBraestCancerNet model for breast cancer detection and classification. In the training and validation of the model, 1030 samples obtained from two different publicly available ultrasound datasets were used. The proposed model achieved 99.35% classification accuracy. Gupta et al. [16] proposed a modified ResNet-50 model that can classify breast ultrasound as normal, benign or malignant. Validated on a publicly available dataset, the model achieved 97.8% accuracy, 97.68% recall, 99.21% precision and 98.44% F1-score.

CNN models have demonstrated promising results in classification studies. However, their black box structure renders them unsuitable for clinical implementation. The primary challenge lies in the inability to identify which pixel areas in the image are crucial for predictions. To overcome this limitation, CNN models can be made explainable using class activation mapping (CAM) algorithms that visualize the feature maps from the last convolutional layer as heat maps. Nonetheless, CAM algorithms employ a coarser approach when locating objects at the region level. Therefore, the output of the CAM algorithm is less accurate compared to sharp and precise pixel-based segmentation results. Byra et al. [17] proposed a deep learning-based method for automatic segmentation of breast cancer tumor areas in ultrasound images. The proposed method was evaluated on 882 different ultrasound images. Of the 882 images, 632 images were used for training, 100 images for validation and 150 images for testing. The trained U-Net and SK-U-Net models made mask predictions on the 150 images allocated for testing. As a result of the predictions, the SK-U-Net model achieved a Dice Score (DS) of 82.60%. Sannasi and Rajaguru [18] proposed a deep learning-based approach to segment tumor areas from breast cancer ultrasound images. They used a publicly available dataset containing 780 ultrasound images collected from 600 different female patients. An adaptive median filtering algorithm was used to remove the noise in the images. FCN16s, FCN8s, SegNet, SegNet, U-Net, SK-U-Net and SKMAT-U-Net models built specifically for this study were trained for 60 epochs with 224×224 input resolution and batch size of 12. As a result of the completed training stages, the SKMAT-U-Net model is the most successful model among the tested models, reaching a DS value of 92.90%.

Segmentation studies have shown high precision and high accuracy. However, since segmentation models can process input images at a fixed resolution, the proposed methods are usually based on image resizing. Information loss is inevitable in resized images. In addition, given the availability of different radiological devices and technologies today, the scope of methods that expect a fixed input size is quite limited.

This paper proposes a ResNext-based U-Net model for segmenting benign and malignant breast cancer tumors from ultrasound images. The proposed model first cropping the input image into 256×256 pixel patches and then extracts features with the encoder blocks (ResNext) of the U-Net model. The model processes the output of the encoder network in the default decoder network of the U-Net model to obtain the prediction mask image. The sequentially processed patches are merged again in the same order to reach the original size. The approach proposed in this study has the potential to be applied clinically for computer-aided automatic breast cancer detection due to its flexible structure. The main contributions of this study can be summarized as follows:

- Semantic segmentation was performed with the ResNext-based U-Net model using a public dataset.

- A method is proposed to process different and high resolution images in the U-Net model.

- Hardware requirements are reduced by using input images divided into 256×256 pixel patches.

- The proposed model outperforms the default U-Net model.

2. Material and Method

A segmentation system was developed to identify benign and malignant tumor regions in ultrasound images using a deep learning model based on ResNext and U-Net. The block diagram illustrating the system is given in Figure 1.

The use of images with high and non-constant resolution poses challenges that can hamper the efficiency and practicality of the model. In order to overcome these challenges, a method is proposed that involves cropping the original images into smaller and more manageable patches. In this way, high-resolution images can be utilized without imposing the memory constraints of the hardware.

2.1. Ultrasound Dataset

In this study, we used a publicly available dataset [19] of ultrasound images of female patients, with corresponding mask images showing tumor areas marked by an expert. The dataset was compiled in 2018 and consists of samples from 600 different patients ranging from 25 to 75 years old. The dataset consists of a total of 780 different images, all stored in PNG file format. The images have different resolutions with non-fixed sizes ranging from 190×335 pixels to 1048×578 pixels. The images in the dataset are categorized into three different classes: normal, benign and malignant.
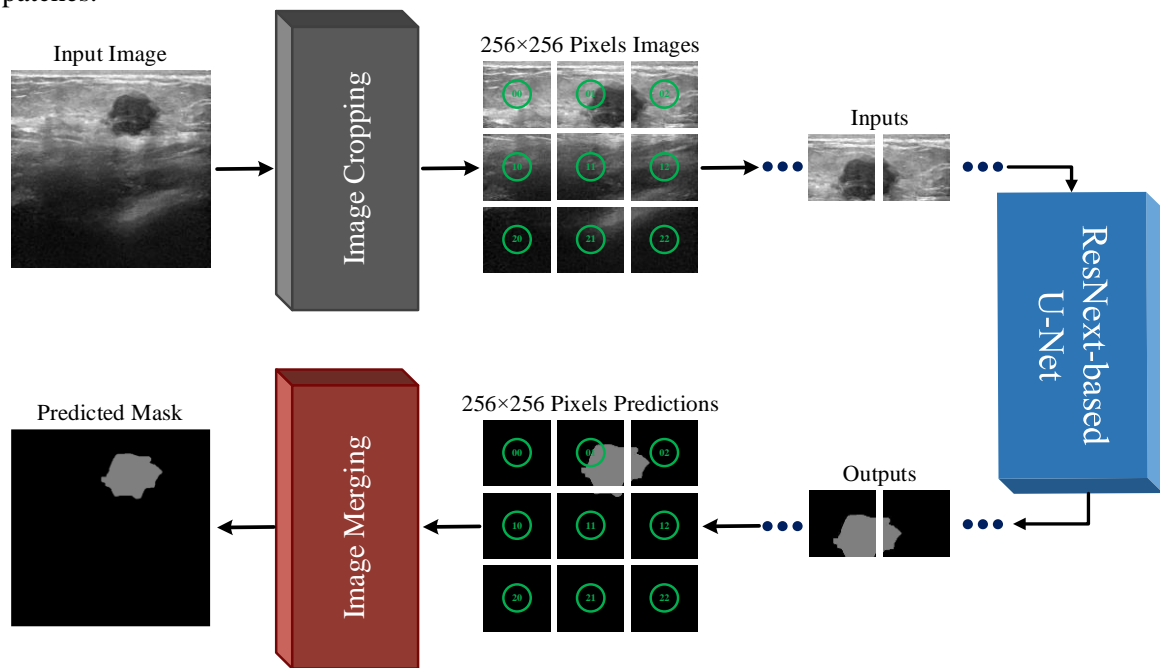

Figure 1. Block diagram of the proposed method

These classes serve to classify ultrasound images according to the presence and nature of tumors. To provide a better understanding of the structure of the dataset, the distribution of images into the three classes is summarized in Table 1.

Table 1. Classes and distributions of dataset samples

Class Name	Number Of Samples
Benign	437
Malignant	210
Normal	133
Total	780

Typically, in medical imaging, different tumor or tissue types are assigned different pixel values or color codes to facilitate easier identification and analysis [20]. In this particular dataset, however, mask images do not use different pixel values to distinguish between benign and malignant tumors. Instead, mask images mark the regions where tumors are present in ultrasound images, regardless of their specific structure. Figure 2 shows a randomly selected set of dataset examples and their corresponding mask images.
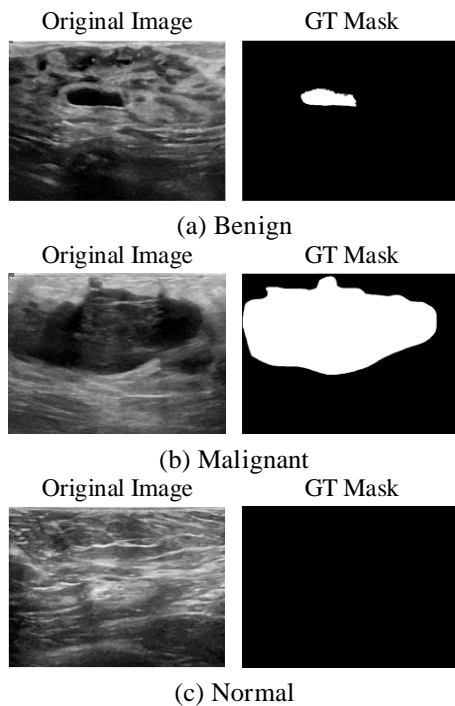

(a) Benign

(b) Malignant

(c) Normal

Figure 2. Dataset samples

2.2. Pre-processing

In the dataset samples, some images contain multiple tumor areas. To accurately represent these multiple tumor regions, separate mask images were initially created for each different area in an image. However, having multiple mask images for a single ultrasound image can lead to challenges and may not be ideal for certain analyses or applications. To address this issue, a preprocessing step was applied to the dataset samples. This pre-processing involves merging the individual mask images associated with multiple tumor areas within a single ultrasound image. This results in a merged mask image that covers all tumor regions in the image. The goal of this pre-processing step is to ensure that each ultrasound image in the dataset is associated with a single mask image, eliminating the complexity and potential complications of having multiple mask images for a single image. Figure 3 shows an example of a dataset with multiple mask images and the composite mask image generated after the pre-processing step.

Pixel values are numerical representations of the intensity or color of each pixel in an image. If benign and malignant tumors have the same pixel values, it becomes problematic to create a reliable segmentation model that can accurately classify and distinguish between these two classes. To overcome this issue, a technique called pixel-based value substitution was employed. This technique involves replacing the original pixel values with new values that correspond to specific classes. In this case, a value of '1' was assigned to denote regions representing benign tumor areas, a value of '2' was assigned for malignant tumor areas, and a value of '0' was designated for areas defined as the background that do not contain tumor regions. The result of the pixel-based value substitution process is shown in Figure 4.
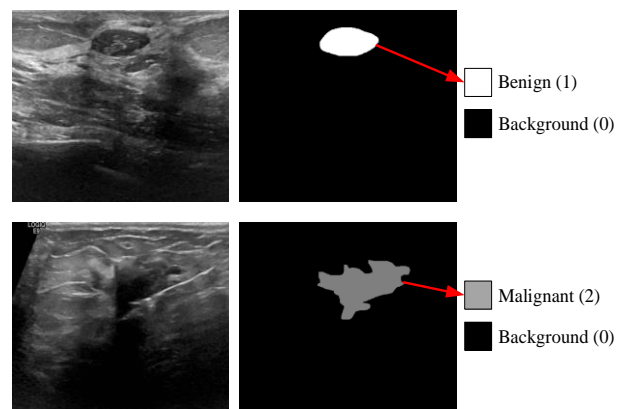

Figure 4. Class-based pixel value assignment

Original Image

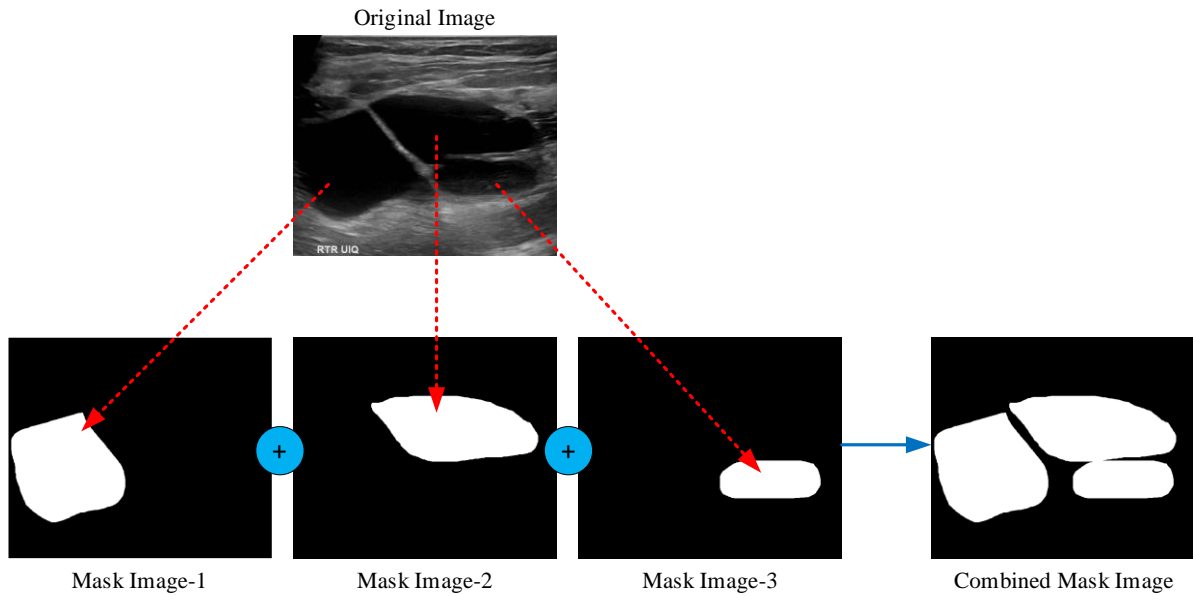Mask Image-1   Mask Image-2   Mask Image-3   Combined Mask Image

Figure 3. Combining multiple mask images into a single mask image

The use of deep learning models for image segmentation often requires fixed input sizes to ensure consistent processing. However, in the case of the dataset used in this study, the samples did not have a fixed resolution, which posed a challenge for training a deep learning model. To address this issue, a decision was made to employ a deep learning model with an input size of 256×256 pixels, which is a commonly chosen size for various segmentation tasks. In order to adapt the original images and corresponding mask images to the desired input size, a cropping process was applied. This involved dividing the images into smaller sub-images, each

with a size of 256×256 pixels. By cropping the images, it became feasible to train a deep learning model that expects fixed-size inputs. During the cropping process, some of the original images might not perfectly fit the 256×256 pixel size due to their inherent dimensions. In such cases, a padding method was utilized to address this discrepancy. The details of the image cropping and padding process are shown in Figure 5. After the padding and cropping operations are completed for the original images, the same operations are repeated for the mask images. In this way, there is no loss in the matching of the original image and the mask image.
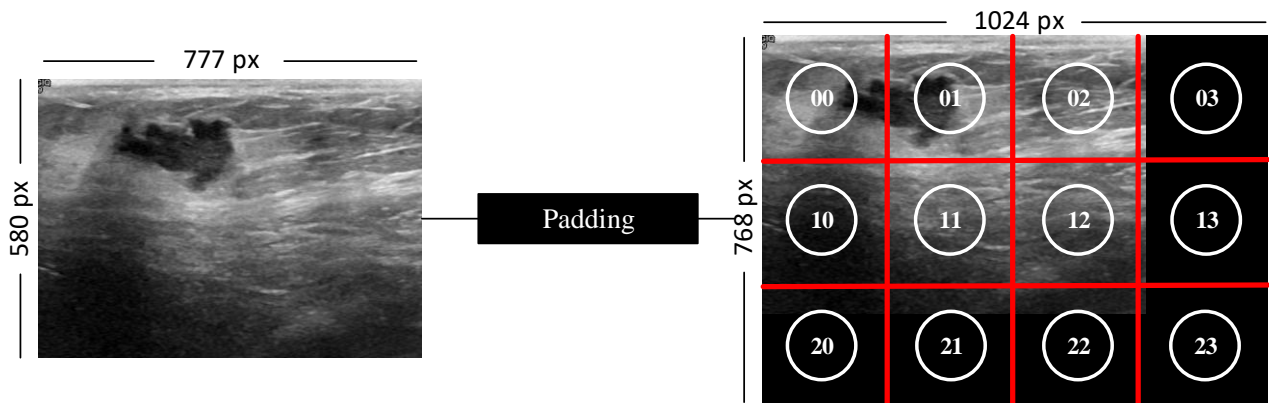

Figure 5. The proposed padding process

2.3. Segmentation Model

The U-Net architecture is a popular CNN architecture that is widely used for image segmentation tasks, particularly in the field of biomedical image analysis [21]. The U-Net architecture derives its name from its

shape, which resembles the letter "U". A block diagram of the U-Net architecture is given in Figure 6. It is designed to address the challenges of pixel-wise image segmentation, where the goal is to assign a label to each pixel in an image, indicating the class or category it belongs to.
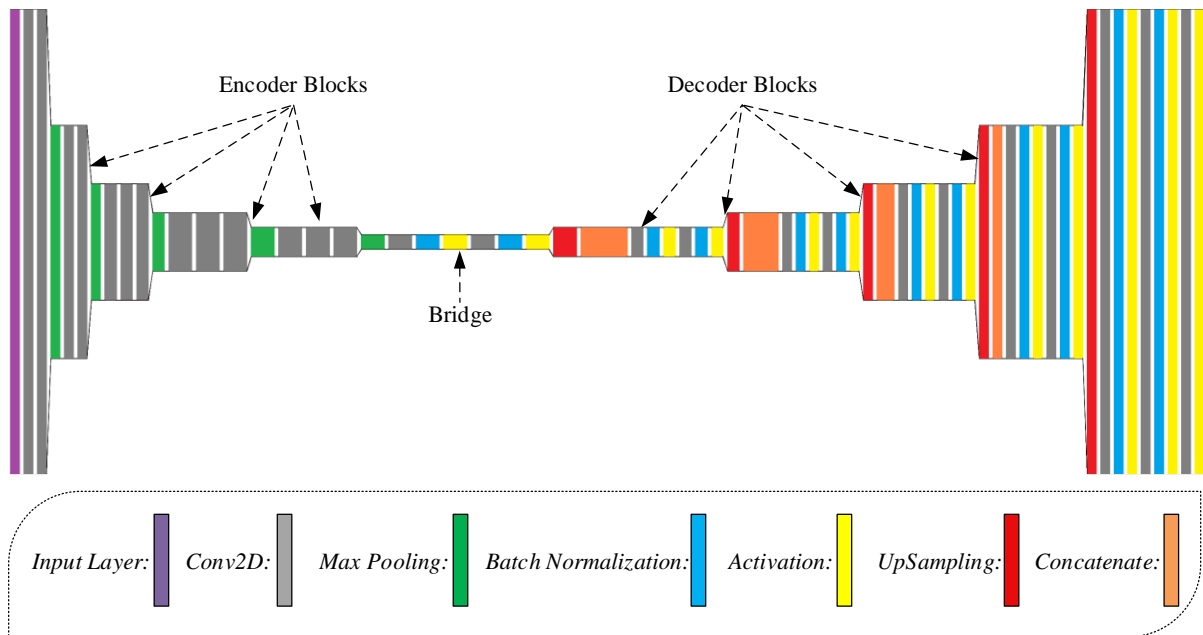
875

Figure 6. The default U-Net architecture

This is particularly useful in applications such as medical image segmentation, where identifying and delineating specific structures or regions within an image is crucial [22]. The key characteristic of the U-Net architecture is its use of an encoder-decoder structure combined with skip connections.

To enhance the performance and efficiency of the U-Net model, pre-trained CNN networks can be utilized in the encoder layer [23]. These networks are models trained on large-scale datasets, such as ImageNet [24], and have learned to recognize a wide range of low-level and high-level features from images. The main advantage of using pre-trained CNN networks in the encoder layer is transfer learning. It allows us to leverage the knowledge and feature representations learned by the pre-trained models on a large dataset and apply it to a different task or domain with limited data. By using pre-trained CNN networks, we can benefit from their generalization capabilities and feature extraction power [25]. In many studies, these networks are used as feature extractors in the encoder layer of the U-Net model [26], [27]. The input images are passed through the pre-trained CNN network, and the output feature maps are then fed into the subsequent layers of the U-Net model. The pre-trained CNN networks can capture generic features like edges, corners, textures, and object parts, which are beneficial for a wide range of computer vision tasks. By utilizing these pre-trained networks, the U-Net model can achieve better performance, even when training data is limited. Moreover, using pre-trained networks in the encoder layer can help speed up the training process since the initial layers are already well-initialized with meaningful filters.

In this study, pre-trained ResNext-50 and ResNext-101 models are integrated into the encoder network of the U-Net models. The ResNeXt architecture is an extension of the ResNet architecture. It introduces a cardinality parameter to ResNet blocks, enabling increased model capacity and performance without significantly increasing computational complexity. The block diagram of the first encoder block of the ResNext-50 based U-Net model is given in Figure 7.

In this study, three different models are compiled to study the effect of CNN architecture on the performance of the encoder network. These models are the default U-Net, ResNext-50 based U-Net and ResNext-101 based U-Net (Models are available at (github.com/oguzhankatar/BCSeg).

2.4. Performance Metrics

In order to assess the performance of a segmentation algorithm, various metrics, including true positive (TP), true negative (TN), false positive (FP), and false negative (FN), are employed. To comprehensively evaluate segmentation algorithms, pixel-based evaluation techniques have been developed. These methods are pivotal in determining the algorithm's effectiveness in identifying and distinguishing malignant, background, and benign regions. They quantify the reate of overlap between the predicted segmentation and the ground truth. Several commonly utilized pixel-based evaluation methods can be summarized as follows.
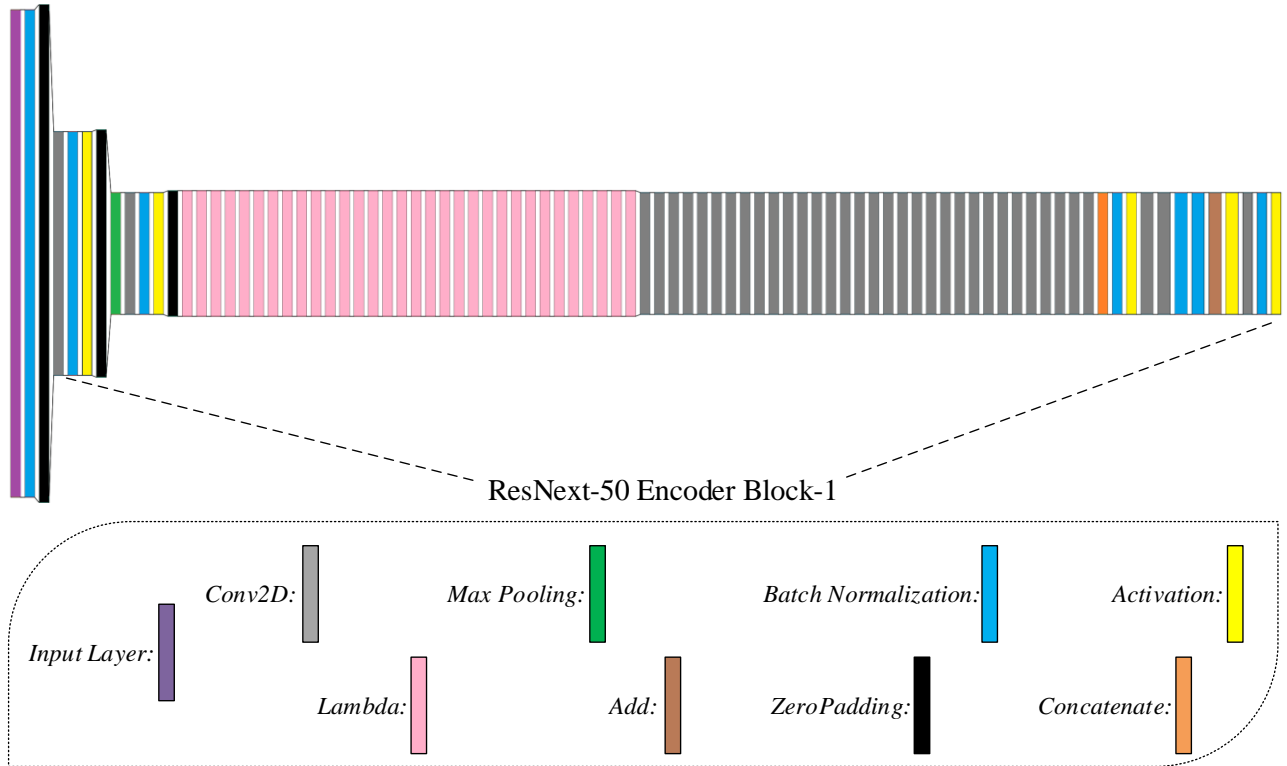
Figure 7. The first encoder block of the ResNext-50 based U-Net

- Pixel accuracy (PA) measures the proportion of correctly classified pixels over the total number of pixels in the image [28].
- Intersection over Union (IoU) measures the overlap between the predicted and ground truth segmentation masks for each class [29].

- Dice coefficient (DC) is another metric that evaluates the overlap between the predicted and ground truth masks. The DC ranges from 0 to 1, with 1 indicating a perfect match [30].

The equations used to calculate these methods and the advantages and disadvantages of each method are given in Table 2.

Table 2. Performance metrics and details

Method	Equation	Advantages	Disadvantages
PA	$\dfrac{(TP + TN)}{(TP + FP + FN + TN)}$	A simple and straightforward metric. Effective for similar distributions of two classes.	It can be misleading in unbalanced class distributions.
IoU	$\dfrac{TP}{(TP + FP + FN)}$	Resilient to class imbalance. Clearly shows the difference between correct and incorrect predictions.	Poor performance on objects with sharp edges.
DC	$\dfrac{2 \times TP}{(2 \times TP + FP + FN)}$	Resilient to class imbalance. Performs well in object segmentation.	Can be complex when there are multiple classes.

3. Results

The results of the U-Net models trained for segmentation of benign, malignant and background areas in ultrasound images are presented in this section. In addition, various evaluation metric values of the experimental findings are shown in the following sections.

3.1. Experimental Setups

After the preprocessing steps on the images were completed, 70% of the images in the dataset were divided for training, 20% for validation and 10% for testing. The dividing process was randomized in such a way that the original image and the mask image were kept together. The setup of the data dividing process is given in Figure 8.
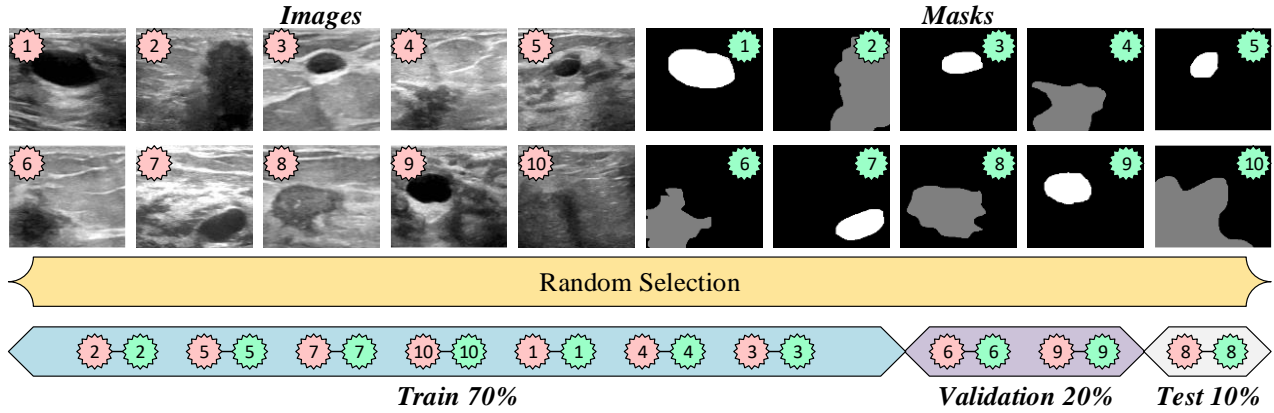

Figure 8. Dividing of training, validation and test subsets

U-Net models with 256×256×3 input size were included in the training thanks to the 'Segmentation Models' library. Models were created by integrating pre-trained ResNext-50 and pre-trained ResNext-101 models into the encoder network of the default U-Net model. In order to accurately compare the performance of the default U-Net, ResNext-50 based U-Net and ResNext-101 based U-Net models, the parameters in the training and validation processes were kept identical. Each model was compiled with Adam optimization and a learning rate of 0.0001. In Adam optimization, the weights are updated by the mathematical process given in Equation 1. w_t represents the updated weights at step t, w_{t-1} represents the weights at step $t-1$, η is the constant known as the learning rate, $\hat{m}_t$ represents the moving average of the gradients, $\hat{v}_t$ represents the moving average of the squares of the gradients, $\in$ represents a small positive value added to avoid division by zero error.

$$w_t = w_{t-1} - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \in} \tag{1}$$

The batch size was 16 and the models were trained for 100 epochs with the Dice Loss function. The mathematical equation used to calculate the Dice Loss value is given in Equation 2. In this equation, p_{true} represents the ground truth mask, p_{pred} represents the model's prediction mask, and $\in$ represents a small positive value added to avoid a division by zero error.

$$Loss_{Dice} = \frac{2 * \sum p_{true} * p_{pred}}{\sum p^2_{true} + \sum p^2_{pred} + \in} \tag{2}$$

All experimental phases were carried out on the A100 GPU in the Google Colab environment using the Python programming language.

3.2. Experimental Results

The number of trainable parameters in a model is a crucial factor that greatly influences the training and validation process. For the proposed models aimed at segmenting benign, malignant, and background areas from ultrasound images, the number of trainable parameters is as follows: 23,748,531 for the default U-Net, 31,993,270 for the ResNext-50 based U-Net, and 51,142,070 for the ResNext-101 based U-Net. The duration of each epoch in training the models directly correlates with the number of trainable parameters. Throughout the training process, the IoU and loss values were recorded at each epoch on both the train and validation sets. The IoU value represents the model's accuracy in segmenting correctly, while the loss value quantifies the discrepancy between the model's predictions and the true values. Figure 9 illustrates the IoU and loss graphs for each model during the training and validation phases.

The performance discrepancy observed between the default U-Net model and the models incorporating pre-trained ResNext-50 and pre-trained ResNext-101 architectures can be attributed to the differences in the feature extraction capabilities of these architectures. The default U-Net model typically consists of simple convolutional layers, while the pre-trained ResNext-50 and ResNext-101 architectures are based on deeper and more complex networks, pre-trained on large-scale datasets.
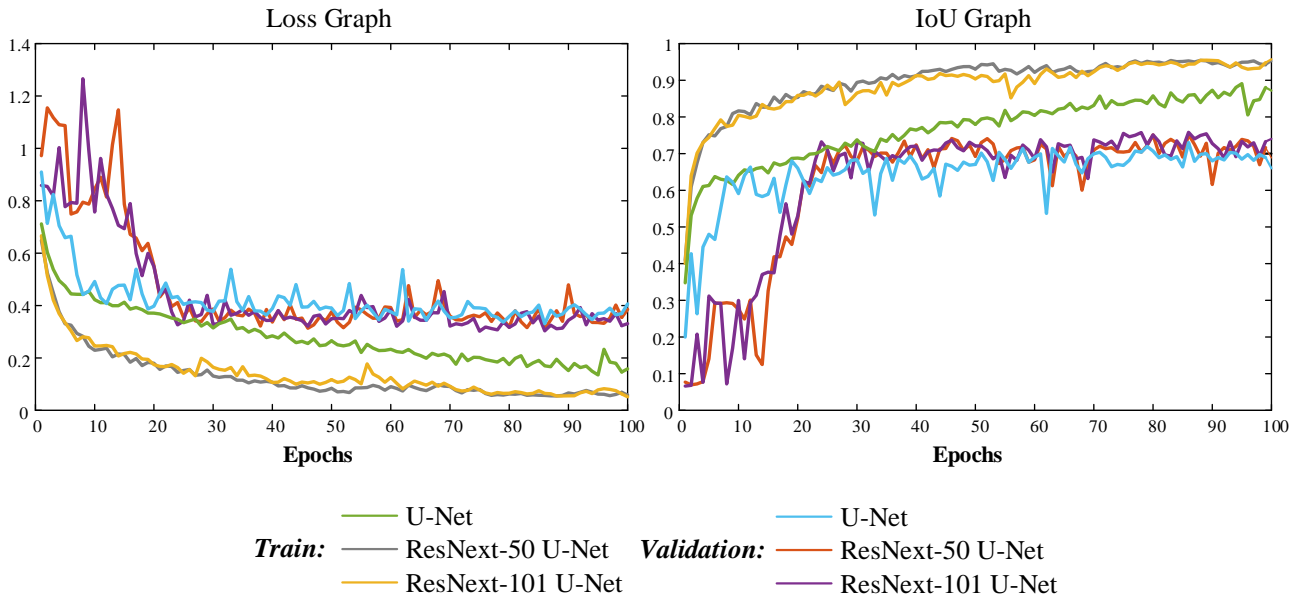
Figure 9. Train and validation graphs

By leveraging the knowledge encoded in the pre-trained ResNext-50 and ResNext-101 architectures, the U-Net models with these architectures as encoders benefit from transfer learning. The pre-trained models have already learned generalizable features from a large and diverse dataset, allowing them to extract relevant information for the segmentation task more effectively. This transfer of knowledge gives the models a head start during training, resulting in improved segmentation performance compared to the default U-Net model.

Therefore, the lower performance of the default U-Net model can be attributed to its relative simplicity and limited ability to capture intricate image features compared to the more advanced pre-trained ResNext-50 and ResNext-101 architectures, which good at feature extraction tasks. The weights obtained by the models after training were saved in 'hdf5' format. Using these weights, the models were tested on the test set samples. The samples randomly selected among the predictions made by the models in the test phase are given in Figure 10.
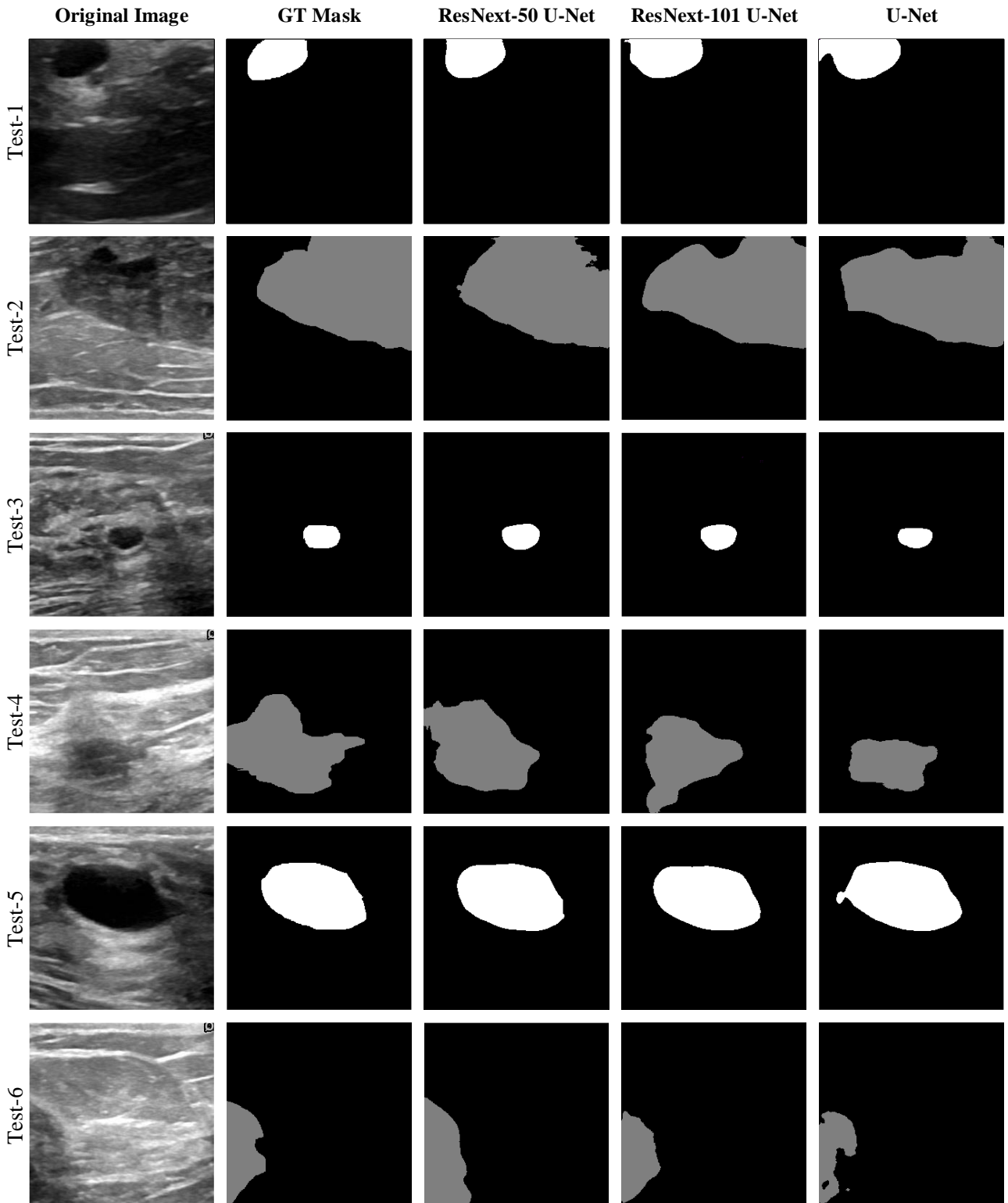
Figure 10. Test samples and predicted masks

Performance metrics were calculated by comparing the predictions with the ground truth mask images on a pixel basis. The bar graph of these metrics is given in Figure 11.
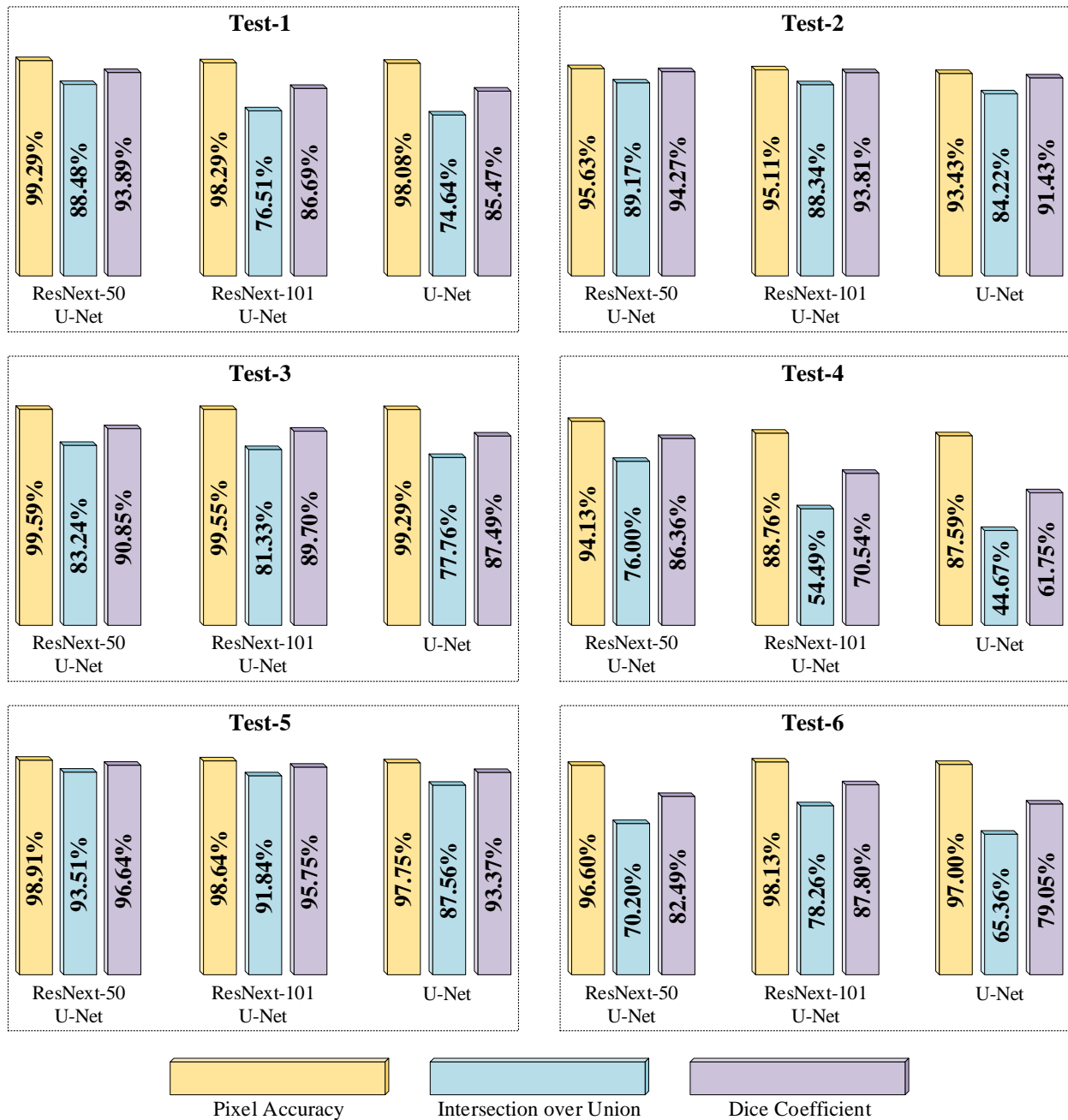
Figure 11. The performance values of test predictions

When the performance of the models on the random test images is analyzed, it is seen that the models achieve high pixel accuracy rates. However, it is not sufficient to make a comparison between the models based on pixel accuracy alone. For this reason, the class-based IoU values of the models were analyzed on each test sample. The class-based IoU values obtained by the models in their predictions on the test images are presented in Table 3.

Table 3. The IoU values obtained on test images

Classes	ResNext-50 U-Net	ResNext-101 U-Net	U-Net
Background (0)	89.06%	87.27%	84.10%
Benign (1)	71.44%	68.34%	65.33%
Malignant (2)	59.01%	61.11%	54.79%
Mean IoU	73.17%	72.24%	68.07%

The models performed worse than the other classes in classifying malignant areas in the test images. The main reasons for this are the visual similarity of malignant areas to background areas and the insufficient number of samples in the dataset. Even though the ResNext-50 based U-Net model classified malignant areas with an IoU rate of 59.01%, it is the model with the highest mean IoU value due to its superior performance over other classes. The ResNext-50 based U-Net model proposed in this study can only predict on 256×256 pixel images. However, thanks to the image cropping and image merging functions, it can produce predictions for images of any size without being bound to any resolution value. The proposed method for the model to produce predictions on all images without depending on any resolution value is given in Figure 12.
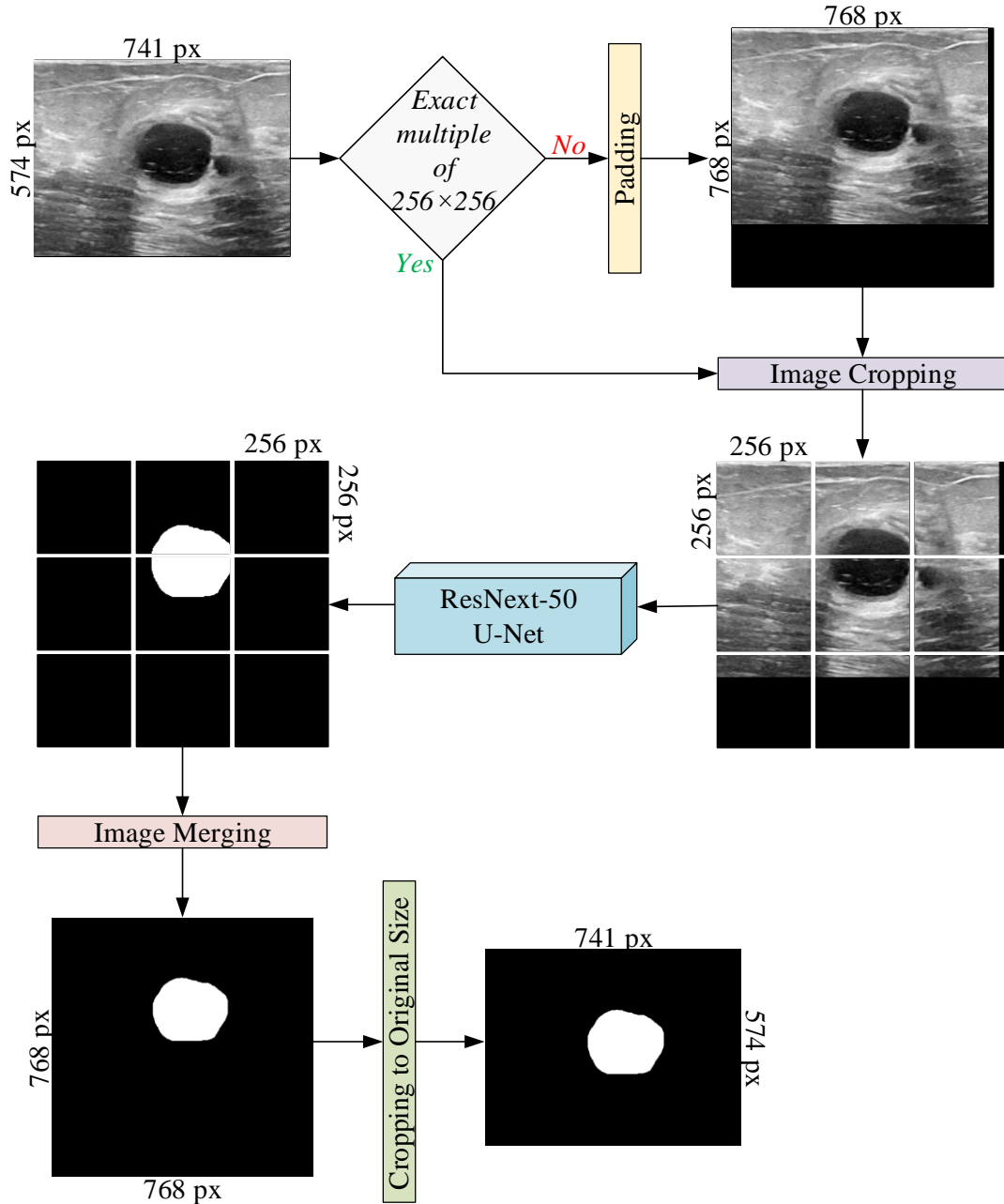

Figure 12. The proposed method for input images of different resolutions

Different radiological devices used today produce images of various resolutions for medical imaging. With the proposed method, it is possible to directly feed images of different resolutions obtained from radiological devices into the model. This provides more flexibility to the experts, allowing them to evaluate images from various devices on the same model. The predictions of the ResNext-50-based U-Net on ultrasound images with different resolutions are presented in Figure 13.
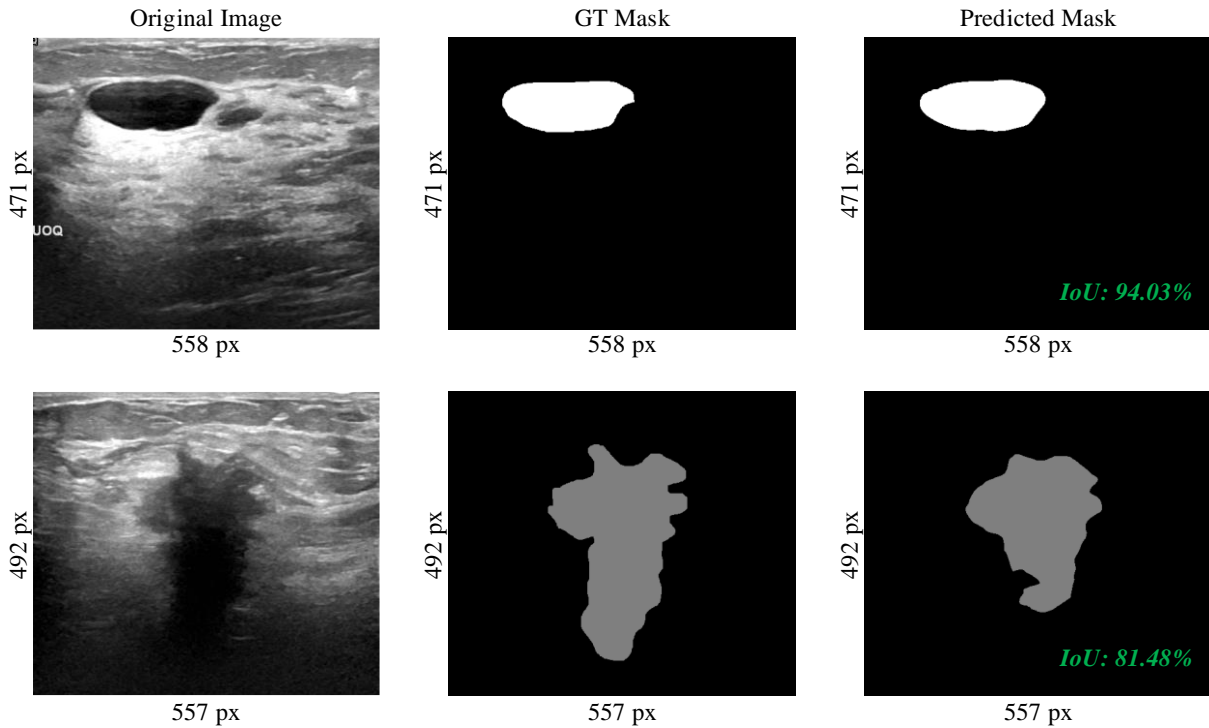
Original Image	GT Mask	Predicted Mask

Figure 13. The predictions on images of different resolutions

4. Discussion

Breast cancer segmentation using radiologic images is an active research topic. Table 4 provides details of research studies developed using the same dataset. Kai et al. [31] proposed the swin-transformerv2-UNet (S2UNet) model. The S2Unet model achieved 73.26% DC and 58.30% IoU on the test set. Inan et al. [32] proposed a U-Net and VGG-16 based method for classification and segmentation of benign and malignant tumor areas on ultrasound images. Simple Linear Iterative Clustering (SLIC) algorithm was applied to the dataset samples in the preprocessing stages. The proposed model was trained for 300 epochs and achieved a DC of 63.40% in the test phase. Bal-Ghaoui et al. [33] proposed a DenseNet-121 based U-Net model for automatic segmentation of breast cancer lesions. This model achieved a DC of 73.70%. Azam et al. [34] proposed the Efficient-Net based Atrous Spatial Pyramid Pooling (Efficient-Net ASPP) model for breast cancer segmentation. Only 210 malignant samples were used in the training and testing processes of the model consisting of EfficientNetV2B3 backbone. The proposed model achieved 62.00% DC on 42 test images.

Table 4. Comparison of studies using the same dataset

Study	Year	Model	Fixed Input Size	Number of Classes	Performance
Kai et al. [31]	2022	S2UNet	Yes	3 (benign, malignant, background)	IoU=58.30% DC=73.26%
Inan et al. [32]	2022	U-Net	Yes	3 (benign, malignant, background)	DC=63.40%
Bal-Ghaoui et al. [33]	2023	DenseNet-121 U-Net	Yes	3 (benign, malignant, background)	DC=73.70%
Azam et al. [34]	2023	Efficient-Net ASPP	Yes	2 (malignant, background)	DC=62.00%
The proposed study	**2023**	**ResNext-50 U-Net**	**No**	**3 (benign, malignant, background)**	**IoU=73.17% DC=83.42%**

In this study, a pre-trained ResNext-50 based U-Net model is used for automatic segmentation of benign and malignant tumor areas from ultrasound images. The model was able to predict three different classes of pixel values in its output and achieved 73.17% test IoU. Compared to the studies listed in Table 4, we obtained higher performance rates. In addition, the constraint of processing on fixed resolution images, which is included in other studies, is not included in our study. Thus, our model is capable of predicting on input images of any resolution and has a wider scope compared to other studies. The advantages of our explainable model can be summarized as follows:

- Since the ResNext-50 model was pre-trained on large datasets, it performed well on the tumor segmentation problem with a low training cost.
- The proposed method has an end-to-end CNN architecture and can perform segmentation without any manual feature extraction.
- In clinical settings, experts can effortlessly utilize the proposed model, facilitating easy interpretation and understanding of the results.
- Due to its flexible structure, it can be fine-tuned with datasets of varying resolutions and effectively employed in diverse tasks.

The limitations of our study are outlined as follows. The dataset used in the study is a publicly available dataset with unbalanced distribution. The performance of the study on images in different formats for integration into real-time applications has not been verified. In future studies, the number of dataset samples will be increased by collaborating with clinical radiologists. Optimization studies will be carried out on the hyper parameters of the model.

5. Conclusions

Detection of benign and malignant tumors is a vital step in the accurate diagnosis of breast cancer. Current detection methods are time-consuming and error-prone due to human factors. Detection of breast cancer cells using artificial intelligence techniques is an active area of research. Studies using CNN-based classifier models achieve high accuracy. However, due to the black box structure of the models, the pixels they focus on in their predictions are not explained. This creates a lack of confidence in the clinical use of classifier models. This study proposes to perform pixel-based segmentation of breast cancer tumors using ResNext-50 based U-Net model. The proposed method includes image cropping and image merging functions that segment the size of the input images into 256×256 pixel patches and reassemble them after performing model predictions on these images. A mean IoU value of 73.17% was obtained on the test images. Although the proposed method is only applied on ultrasound images, it can be easily fine-tuned for different types of radiological imaging such as mammography. In the future, we plan to expand the dataset and process data from different devices with the support of an expert radiologist. The proposed method can be tested to reduce the hardware requirements for segmentation of high-resolution radiological images.

Contributions of the Authors

The authors confirm that the contribution is equally for this paper.

Conflict of Interest Statement

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

References

[1]   J. S. You and P. A. Jones, "Cancer genetics and epigenetics: two sides of the same coin?," *Cancer Cell*, vol. 22, no. 1, pp. 9–20, 2012.

[2]   S. Gómez-López, R. G. Lerner, and C. Petritsch, "Asymmetric cell division of stem and progenitor cells during homeostasis and cancer," *Cellular and Molecular Life Sciences*, vol. 71, pp. 575–597, 2014.

[3]   N. Parsa, "Environmental factors inducing human cancers," *Iran J Public Health*, vol. 41, no. 11, p. 1, 2012.

[4]   M. Hejmadi, Introduction to cancer biology. *Bookboon*, 2014.

[5]   M. Amrane, S. Oukid, I. Gagaoua, and T. Ensari, "Breast cancer classification using machine learning," in *2018 electric electronics, computer science, biomedical engineerings' meeting (EBBT)*, IEEE, 2018, pp. 1–4.

[6]   J. Boutry et al., "The evolution and ecology of benign tumors," *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*, vol. 1877, no. 1, p. 188643, 2022.

[7]   K. Soda, "The mechanisms by which polyamines accelerate tumor spread," *Journal of Experimental & Clinical Cancer Research*, vol. 30, pp. 1–9, 2011.

[8]   L. Wilkinson and T. Gathani, "Understanding breast cancer as a global health concern," *Br J Radiol*, vol. 95, no. 1130, p. 20211033, 2022.

[9]   K. Mortezaee, "Organ tropism in solid tumor metastasis: an updated review," *Future Oncology*, vol. 17, no. 15, pp. 1943–1961, 2021.

[10]   O. Ginsburg et al., "Breast cancer early detection: A phased approach to implementation," *Cancer*, vol. 126, pp. 2379–2393, 2020.

[11]   E. Michael, H. Ma, H. Li, and S. Qi, "An optimized framework for breast cancer classification using machine learning," *Biomed Res Int*, vol. 2022, 2022.

[12]   F. A. González-Luna, J. Hernández-López, and W. Gomez-Flores, "A performance evaluation of machine learning techniques for breast ultrasound classification," in *2019 16th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE)*, IEEE, 2019, pp. 1–5.

[13]   M. Wei, Y. Du, X. Wu, and J. Zhu, "Automatic classification of benign and malignant breast tumors in ultrasound image with texture and morphological features," in *2019 IEEE 13th International Conference on Anti-counterfeiting, Security, and Identification (ASID)*, IEEE, 2019, pp. 126–130.

[14]   K. Atrey, B. K. Singh, N. K. Bodhey, and R. B. Pachori, "Mammography and ultrasound based dual modality classification of breast cancer using a hybrid deep learning approach," *Biomed Signal Process Control*, vol. 86, p. 104919, 2023.

[15]   A. Raza, N. Ullah, J. A. Khan, M. Assam, A. Guzzo, and H. Aljuaid, "DeepBreastCancerNet: A Novel Deep Learning Model for Breast Cancer Detection Using Ultrasound Images," *Applied Sciences*, vol. 13, no. 4, p. 2082, 2023.

[16]   S. Gupta, S. Agrawal, S. K. Singh, and S. Kumar, "A Novel Transfer Learning-Based Model for Ultrasound Breast Cancer Image Classification," in *Computational Vision and Bio-Inspired Computing: Proceedings of ICCVBIC 2022*, Springer, 2023, pp. 511–523.

[17]   M. Byra et al., "Breast mass segmentation in ultrasound with selective kernel U-Net convolutional neural network," *Biomed Signal Process Control*, vol. 61, p. 102027, 2020.

[18]   S. R. Sannasi Chakravarthy and H. Rajaguru, "SKMAT-U-Net architecture for breast mass segmentation," *Int J Imaging Syst Technol*, vol. 32, no. 6, pp. 1880–1888, 2022.

[19]   W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, "Dataset of breast ultrasound images," *Data Brief*, vol. 28, p. 104863, 2020.

[20]   F. Oztekin et al., "Automatic semantic segmentation for dental restorations in panoramic radiography images using U-Net model," *Int J Imaging Syst Technol*, vol. 32, no. 6, pp. 1990–2001, 2022.

[21]   O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference*, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, Springer, 2015, pp. 234–241.

[22]   J. Zhang, Y. Zhang, Y. Jin, J. Xu, and X. Xu, "MDU-Net: multi-scale densely connected U-Net for biomedical image segmentation," *Health Inf Sci Syst*, vol. 11, no. 1, p. 13, 2023.

[23]   A. Abedalla, M. Abdullah, M. Al-Ayyoub, and E. Benkhelifa, "Chest X-ray pneumothorax segmentation using U-Net with EfficientNet and ResNet architectures," *PeerJ Comput Sci*, vol. 7, p. e607, 2021.

[24]   J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*, IEEE, 2009, pp. 248–255.

[25]   X. Zhong and H. Ban, "Pre-trained network-based transfer learning: A small-sample machine learning approach to nuclear power plant classification problem," *Ann Nucl Energy*, vol. 175, p. 109201, 2022.

[26]   M. H. BENDIABDALLAH and N. SETTOUTI, "A comparison of U-net backbone architectures for the automatic white blood cells segmentation," *WAS Science Nature*, vol. 4, no. 1, 2021.

[27] A. Abedalla, M. Abdullah, M. Al-Ayyoub, and E. Benkhelifa, "The 2ST-UNet for pneumothorax segmentation in chest X-Rays using ResNet34 as a backbone for U-Net," *arXiv preprint* arXiv:2009.02805, 2020.

[28] M. Xi, J. Li, Z. He, M. Yu, and F. Qin, "NRN-RSSEG: A deep neural network model for combating label noise in semantic segmentation of remote sensing images," *Remote Sens*, vol. 15, no. 1, p. 108, 2022.

[29] A. N. Gajjar and J. Jethva, "Intersection over Union based analysis of Image detection/segmentation using CNN model," in *2022 Second International Conference on Power, Control and Computing Technologies (ICPC2T)*, IEEE, 2022, pp. 1–6.

[30] A. M. H. Mahran, W. Hussein, and S. E. D. M. Saber, "Automatic Teeth Segmentation Using Attention U-Net," *Preprint*, 2023.

[31] H. Kai, Z. Y. Feng, H. Meng, F. Y. Baoping, and Y. R. Han, "Ultrasound Image Segmentation of Breast Tumors Based on Swin-transformerv2," in *Proceedings of the 2022 10th International Conference on Information Technology: IoT and Smart City*, 2022, pp. 106–111.

[32] M. S. K. Inan, F. I. Alam, and R. Hasan, "Deep integrated pipeline of segmentation guided classification of breast cancer from ultrasound images," *Biomed Signal Process Control*, vol. 75, p. 103553, 2022.

[33] M. Bal-Ghaoui, M. H. E. Y. Alaoui, A. Jilbab, and A. Bourouhou, "U-Net transfer learning backbones for lesions segmentation in breast ultrasound images," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 5, pp. 5747–5754, 2023.

[34] K. Azam, M. A. Azam, M. A. Qureshi, K. B. Khan, and M. A. Azam, "Efficient-Net ASPP Deep Network for Malignant Ultrasound Breast Cancer Segmentation," in *2023 IEEE International Conference on Emerging Trends in Engineering, Sciences and Technology (ICES&T)*, IEEE, 2023, pp. 1–6.