

# FROM U-NET TO (U-NET) +, WHAT INNOVATIONS HAVE WE MADE FOR THE TREATMENT AND DISCOVERY OF BREAST CANCER?

Minuo Qing<sup>1,2</sup>, Emily X. Ding<sup>1</sup>(*Corresponding author*), Robert J. Hou<sup>1</sup>  
1 Vineyards AI Lab, Auckland, New Zealand

2 Westlake Girls High School, Auckland, New Zealand

## **ABSTRACT**

*Artificial Intelligence (AI) has achieved remarkable performance in the field of medical image analysis, particularly in tasks such as object detection, segmentation, and classification. In this paper, we introduce a solution for automatic breast cancer diagnosis based on the U-Net architecture, which we call (U-Net)+. The novel (U-Net)+ is designed to handle both segmentation and classification tasks within a signal framework. We retained the original U-Net architecture due to its strong learning capabilities and its advantages in semantic segmentation. Notable, we incorporated fully connected layers into the bottleneck layers, serving as a multi-functional classifier for both initial diagnoses based on raw images and further diagnoses for segmented images. The (U-Net)+ model is trained using a joint loss function. We conducted the experiments on breast ultrasound images, demonstrating that the (U-Net) performs well in both classification and segmentation tasks.*

## **KEYWORDS**

*Breast Cancer; Automatic Diagnose; Classification; Semantic Segmentation; U-Net;*

## **1. INTRODUCTION**

At the beginning of 2020, I experienced the loss of a loved one to cancer for the first time. My aunt passed away after a two-year battle with breast cancer, and it was the first occasion I felt that cancer's proximity to my life. It was no longer a pathological concept in textbooks but a disease that could affect those around us, or even ourselves. Moreover, since 2022, I have been involved in advocating for affirmative action for Afghan women, including women's right to education and voice. During this journey, I learned that breast cancer is the most common cancer among Afghan women, especially in Kabul City. Using this fact as a starting point, I began to focus on events related to breast cancer. According to the World Health Organization (WHO) Global Female Breast Disease Questionnaire, more than 2.3 million women worldwide suffer from breast cancer every year. In more than 90% of countries, breast cancer is the first or second leading cause of female cancer death[1].

However, in some cases, breast cancer is detected only after symptoms appear, yet many women with breast cancer remain asymptomatic in the early stages. Early detection or diagnosis is important, as it

can significantly enhance survival rates and drastically reduce overall treatment costs[2]. Usually, the most direct and efficient method for detecting breast tumours is through medical imaging tests. The precision of this diagnosis relies on the physician's experiences and expertise in image analysis. This raises the question: can we develop a computer-aided system to assist doctors in achieving more accurate and efficient breast cancer diagnoses?

With the rapid development of AI technologies, they have been applied in medical image analysis, including image classification, target detection, image segmentation, and image retrieval, etc. These applications have achieved remarkable performances[3,4,5,6]. AI technologies were used to develop tools to assist doctors in detecting and segmenting tumours with lower possible errors. During my research, I explored various online resources and discovered U-Net as a deep-learning model that has shown great success in the field of medical image segmentation[7,8]. The availability of open-source code implementations and extensive academic research attracted me. I try my best to understand its details and implement it on breast cancer ultrasound images. Additionally, we designed an enhanced U-Net model and named it (U-Net)+, which is proficient in both segmentation and classification. The model offers the physicians the segmented tumour and automatic diagnosis results, categorizing them as normal, benign, or malignant. The main contributions of our work include:

1. We presented a comprehensive view of U-Net in an easily understandable manner. To understand it well, we likened U-Net to the process of seed growth for the contracting path and blossoming for the expansive path, where the skip connections can be thought of as the source of nourishment. In addition, we displayed the 2D convolution, and 2D transpose convolution operations, and introduced the concepts of encoder and decoder.
2. We designed a system that incorporates multiple functions, including classification and segmentation based on the U-Net architecture. We named it as (U-Net)+. We embedded a classifier that can classify the normal, benign, and malignant samples. It performs effectively on both ultrasound images and segmented images. During the training process, we employed Sigmoid and Softmax activation functions as per their respective tasks and also applied different weights in the total loss functions.

## **2. BACKGROUND**

### **2.1 Basic information about Breast Cancer**

Breast cancer is the most frequent cancer among women worldwide, accounting for 1 in 4 cancer cases. It occurs when the damaged cells grow in an uncontrolled way and a tumour is formed. This disease is typically categorized into four stages: the appearance of abnormal cells, pre-invasive and invasive cancer, and cancer spread. When breast cancer is detected and treated at an early stage, the chances of survival are significantly higher. Figure 1 illustrates the division of different stages and the corresponding survival rates.

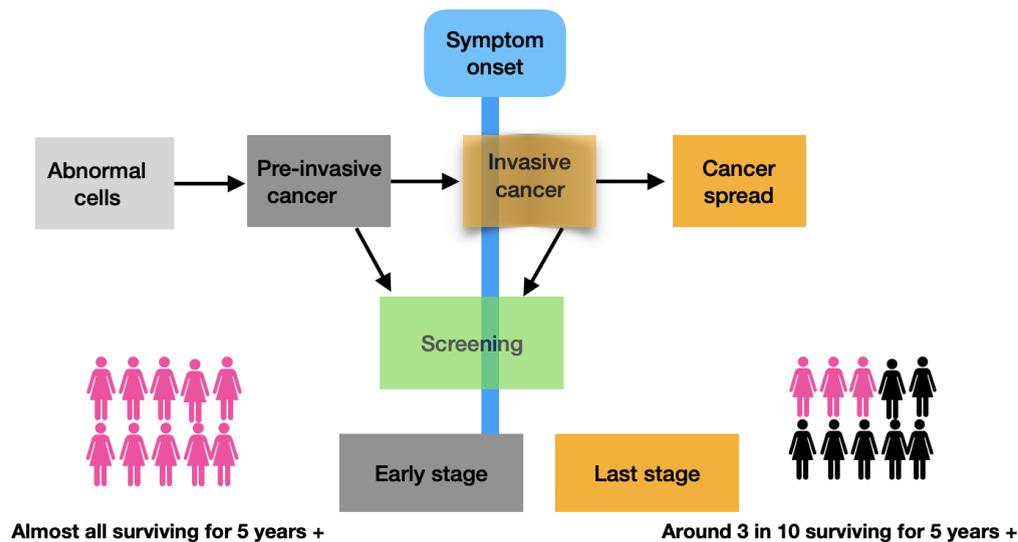


Figure 1. Breast Cancer Diagnosis Stages and Survival Rates

According to statistical data, we see that nearly all women survive breast cancer for 5 years or more if breast cancer is diagnosed at the early stage, but this rate decreases to approximately 3 in 10 women when the cancer is in its advanced stage[9]. However, early diagnosis can be challenging because many women with breast cancer don't exhibit symptoms in the initial stages. Therefore, regular breast cancer screening and precise medical imaging analysis are critical for preventing death. Currently, a range of medical imaging methods is employed, such as breast ultrasound, breast MRI, mammograms, newer and experimental Breast Imaging Tests, etc. In the next, advanced image analysis technologies gradually appeared to improve the accuracy of diagnosis.

## 2.2 AI in medical image analysis

Medical imaging plays an important role in disease diagnosis and surgical treatment. Through these images, doctors can locate tumours, segment specific areas, and make decisions regarding staging[10]. However, manual detection and segmentation processes are time-consuming and susceptible to various subjective and objective factors that include a physician's expertise, emotions, prejudice, and diagnostic methods, resulting in a relatively high rate of misdiagnosis. To understand and interpret images well, many computer-aided diagnosis (CAD) schemes have been developed. These schemes aim to assist doctors more efficiently and objectively, to achieve a higher diagnostic accuracy[11,12]. In recent years, AI has experienced rapid development, leading to its widespread applications in the medical field. At present, AI is employed in medical image processing, including image classification, target detection, image segmentation, and image retrieval.

To complete these tasks, extensive models have been developed, and deep learning models have been remarkably successful. As one of the most popular deep learning-based models, U-Net was proposed by Ronneberger et al in 2015, which is for Biomedical Image Segmentation[13]. It can capture both the context and the localisation feature. In our work, we'll train a smart U-Net called (U-Net)+ for breast cancer diagnosis.

## 3. BREAST CANCER DIAGNOSIS SYSTEM

We aim to design an intelligent system capable of performing both segmentation and classification for breast cancer diagnosis automatically, as illustrated in Figure 2. Usually, input samples in the form of ultrasound images are processed through the system. Firstly, the system provides an initial classification result as either normal, benign, or malignant. Then, the images are directed to the

semantic segmentation module, and based on the segmentation results, the system generates further and final diagnostic results. Next, we'll describe this process from an AI perspective. Enhanced segmentation is crucial, as it not only facilitates more accurate diagnoses but also plays a significant role in planning surgery and determining appropriate therapeutic strategies.

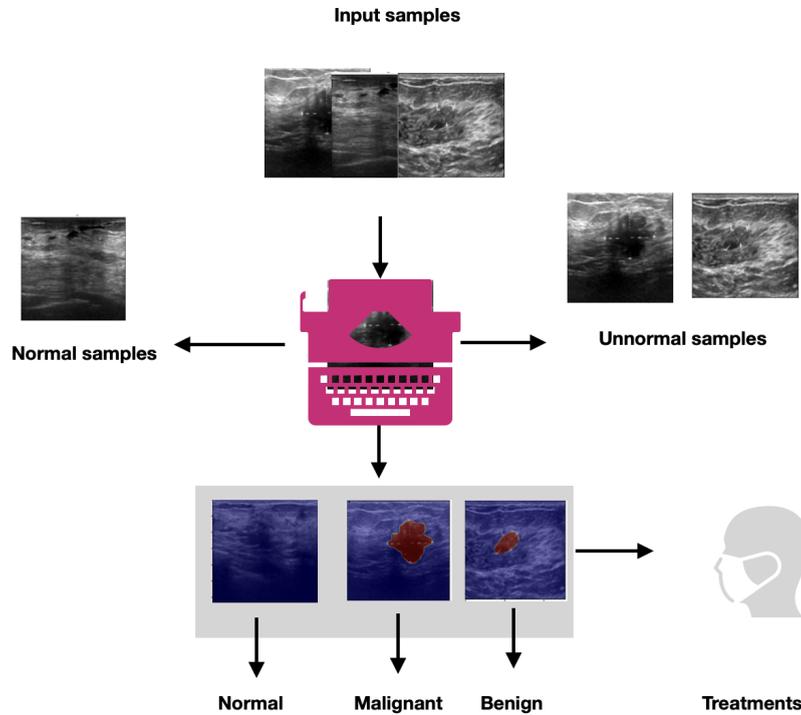


Figure 2. Breast Cancer Diagnose

First, we organised the dataset into three categories: normal, benign and malignant. The dataset includes ultrasound image sets  $X$  and their corresponding labels  $Y$ , 0 for benign, 1 for malignant, and 2 for normal. Consequently, we need to train a multi-class classifier  $C_1$ . The goal of the classifier  $C_1$  is to establish a rule for predicting the label  $Y_{pred}$  for a given sample in  $X$ , which can be represented as

$$C_1: X_{test} \rightarrow Y_{pred} \{0, 1, 2\} \quad (1)$$

Then the ultrasound images undergo processing in the semantic segmentation module, which is a technology that aims to categorise each pixel in an image which means each pixel is assigned to a specific class. Semantic segmentation can also be seen as a classifier to predict the categories for each pixel.

In our work, we defined the two categories in the breast cancer image, one is a normal pixel (such as tissue) and the other is abnormal pixels representing the tumour. Unlike the first classifier  $C_1$ , the labels generated by segmentation form a matrix of the same size as the image. In this matrix, 0 represents the background, and 1 denotes the tumour. The goal of semantic segmentation is to highlight the tumour's shape, edges, and positions while eliminating noise or interference.

$$S: X_i \in X \rightarrow M_i \{0,1\}, W_i \in M \quad (2)$$

We call the  $M$  mask image, which can be represented as  $M = S(X)$  semantic segmentation seems

like a labelling operator  $S$ .

Finally, the classifier is needed again to recognise the normal, malignant or benign samples based on the segmentation image  $M$ . Therefore, the diagnosis classifier is defined as

$$C_2: M_{test} \rightarrow Y_{pred} \{0, 1, 2\} \quad (3)$$

Our motivation is to design a smart U-Net system that can facilitate the diagnosis process. How to design this system and organise the relevant training data? We'll display our solution in section 4.

## 4. OUR SOLUTION: TRAIN A (U- NET)+

As we discussed in the previous section, we'd like to design a system that can perform both classification and semantic segmentation tasks. With the rapid development of neural networks, image recognition and segmentation achieved outperforming performance. In this section, we'll display our solution (U-Net)+.

### 4.1 U-Net introduction

U-Net is a popular deep-learning architecture for semantic segmentation. It has been widely applied in various medical image segmentation tasks, including brain tumour segmentation, lung segmentation, cell segmentation, etc. [14]. This architecture is based on U-shape type encoder-decoder networks with skip connections, which was named U-Net. Research has shown that U-Net has demonstrated successful results due to its representation learning capabilities and the ability to recover fine-grained details[15]. While existing U-Net has demonstrated powerful performance, they can only handle one task at a time. Figure 3 displays the U-Net and its corresponding analogies.

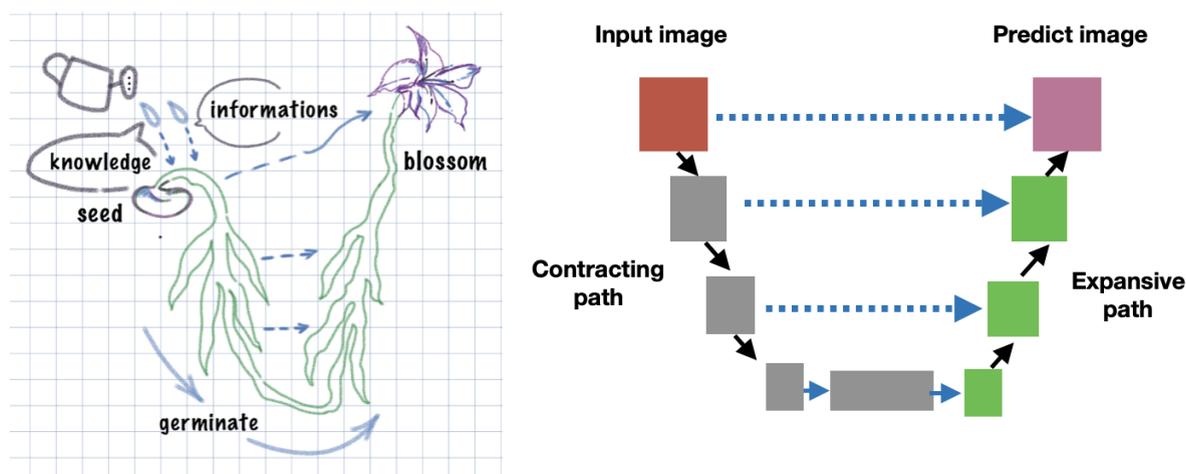


Figure. 3 U-Net architecture and its analogy

To understand the U-net architecture easily, an analogy has been drawn between its architecture and the growth process of a seed. The network's architecture comprises three parts: a contracting path (left side), a bottleneck layer (bottom), and an expansive path (right side). The contracting path seems like a seed was planted in the soil, where it takes root and absorbs the nutrients. In contrast, the expansive path mirrors the process of germination and growth, while the bottleneck layer acts as the bridge connecting these two paths. Moreover, the transfer of essential information between these paths is facilitated by the skip connections, which are analogous to the transport of nutrients.

In addition, there are two important concepts in this symmetrical architecture, encoder and decoder, where the left side takes root down to extract the nutrients just like an encoder and grows in an

expansive path that represents the function of a decoder. In the next section, we'll introduce them in detail.

#### 4.2 The contracting path: absorb and extract the nutrients (learning features)

In this section, we are confronted with a challenging question: how to extract the essential features (resembling nutrients) required for our tasks? The answer is hidden in the contracting path, which is constructed by several encoder blocks. We explained them from a high school student's view, making them understandable to younger learners and inspiring their curiosity. Figure 3 shows the modules in the contracting path.

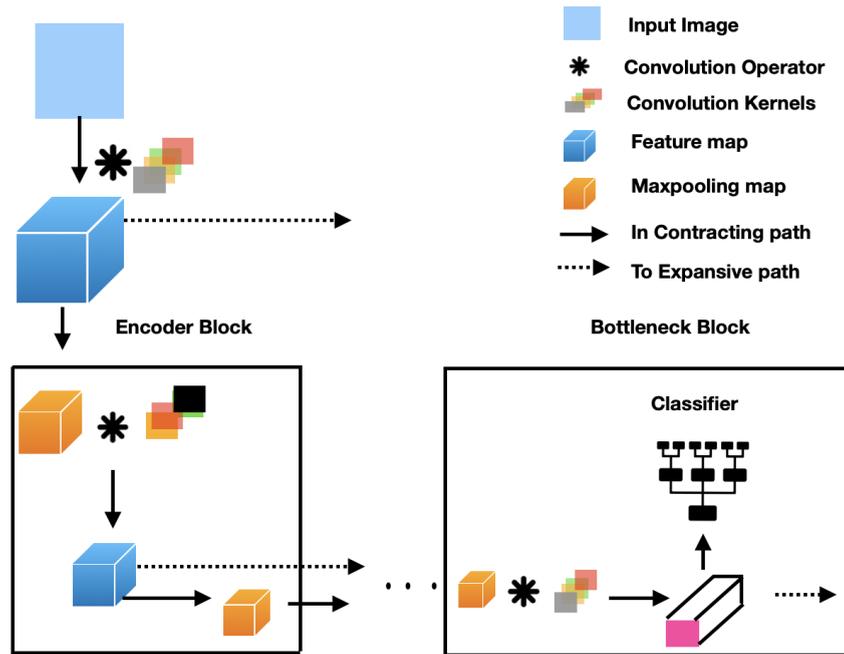


Figure. 4 Contracting Path and Bottleneck Block

First, let's introduce what the encoder in an AI algorithm is. Encoder is a terminology and an unfamiliar word for us. It serves as an operator that extracts the most significant features from the original data and represents it in a latent space. Consequently, the data was compressed after passing through each encoder block.

To provide a simpler analogy for the encoder, think of it as the process of transforming a seed into a root, with the essential nutrition being the hidden feature.

In our model, the encoder block comprises two components: 2D convolution and max-pooling operations. The 2D convolution is a technique in image processing that generates a feature map. While max-pooling can reduce the size/shape of the feature map. The 2D convolution equation is as follows:

$$F(m, n) = \sum_j \sum_k K(i, j) I(m - i)(n - j) \quad (4)$$

Where  $K$  is the kernel or filter, pass it over the image  $I$ . The indexes of rows and columns of the results feature map are marked with  $m$  and  $n$  respectively. In Figure 5, we show what are the operations in the encoder block, including 2D convolution, max pooling and related technologies.

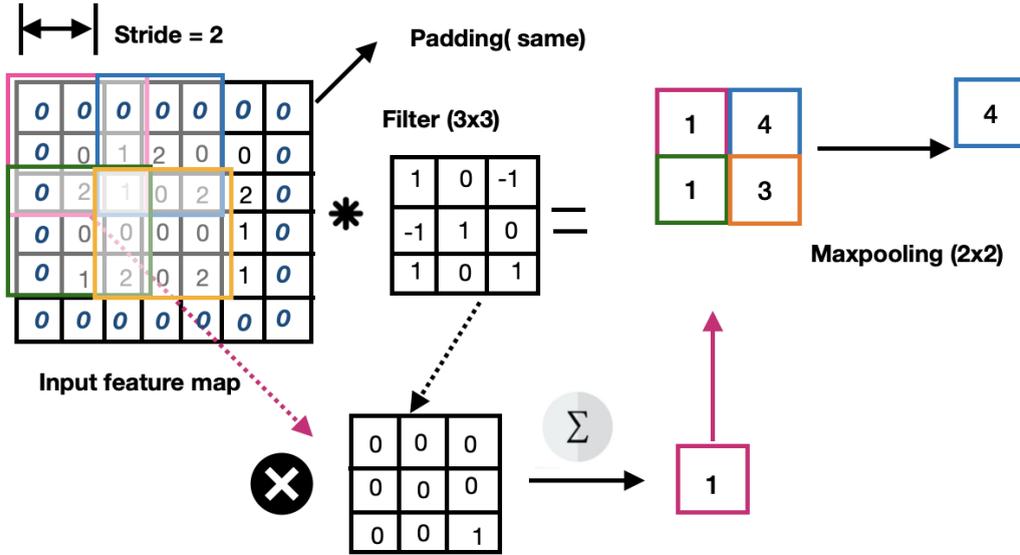


Figure 5. Operations in the encoder block

In Figure 5, we add a layer with the number “0” called padding, which is used to maintain the size of the original input. Then the filter  $K$  shaped as a 3 by 3 matrix performs an element-wise multiplication with the input image  $I$ . The values are summed to generate the feature map. This operation is achieved by sliding the filter over the input image, and the movement happens in steps known as strides. Finally, the max-pooling operation is employed to calculate the maximum value within each patch of the feature map. The purpose of max-pooling is to reduce the dimensionality of the feature map.

Therefore, there are two outputs for each encoder block, feature map and max pooling map. The max pooling map  $S_{map}^l$  will be the input of the next encoder block, and the feature map  $FC_{map}^l$  will be passed to the other side with a skip connection. We describe the  $l$ th encoder block as follows:

$$FC_{map}^l, S_{map}^l = Encoder(S_{map}^{l-1}) \quad (5)$$

$$FC_{map}^l = Conv2D(S_{map}^{l-1}, Filters^l) \quad (6)$$

$$S_{map}^l = Max Pooling(FC_{map}^l) \quad (7)$$

Then we'll introduce the bottleneck layer and skip connection, which are important parts of our (U-Net)+.

### 4.3 Bottleneck layer and skip connection: Connect and transport nutrients (features)

The bottleneck layer is located between the final encoder block of the contracting path and the first decoder block of the expansive path. It comprises a convolution layer with 1024 filters without max-pooling. To make class predictions, a fully connected layer and a softmax function need to be employed at this stage. Figure. 6 displays the details of this part.

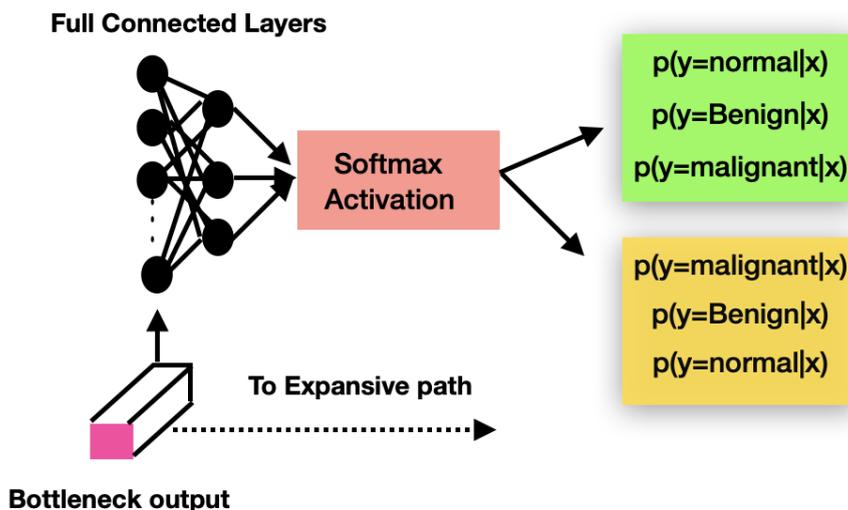


Figure 6. Details of Bottleneck layers in (U-Net)+

The output of the bottleneck layer serves as both the input for the expansive path and connects to a fully connected layer. The objective of the fully connected layer is to make class predictions. This layer takes an input that has been flattened into a vector and processes it through three dense layers. Then an activation function is applied before being sent to the output layer. It is described as follows:

$$y = f\left(\sum_{i=1} w_i x_i + b\right) \tag{8}$$

Where  $w$  is the weights and  $b$  bias (a constant value),  $f(\cdot)$  is the activation function. The choice of activation function depends on the type of classification problem. The “Sigmoid” for binary classification and “Softmax” for multi-class classification [16,17]. Here, the task is a multi-class classification for 3 categories, therefore, softmax is used. The final output vector’s size should correspond to the number of classes for prediction.

Another important element in the transfer of nutrients (features) is known as skip connection. We can imagine that the deep networks might “forget” or “lose” certain features as information passes through successive layers. To overcome this weakness, skip connections were introduced. Every decoder incorporates the feature map from its corresponding encoder, helping to preserve and reintroduce important features into the network.

#### 4.4 The expansive path: How to germinate and grow

In the expansive path, it takes the extracted features and reconstructs a segmentation mask. There are several Decoder blocks in the expansive path. In each decoder block, two fundamental operations are performed: transposed convolution and convolution. In addition, there is a bridge (skip connection) that connects the two paths and completes the flow of information. Specifically, the feature map from the contracting path is concatenated with the transposed map and they are passed through the other convolution together.

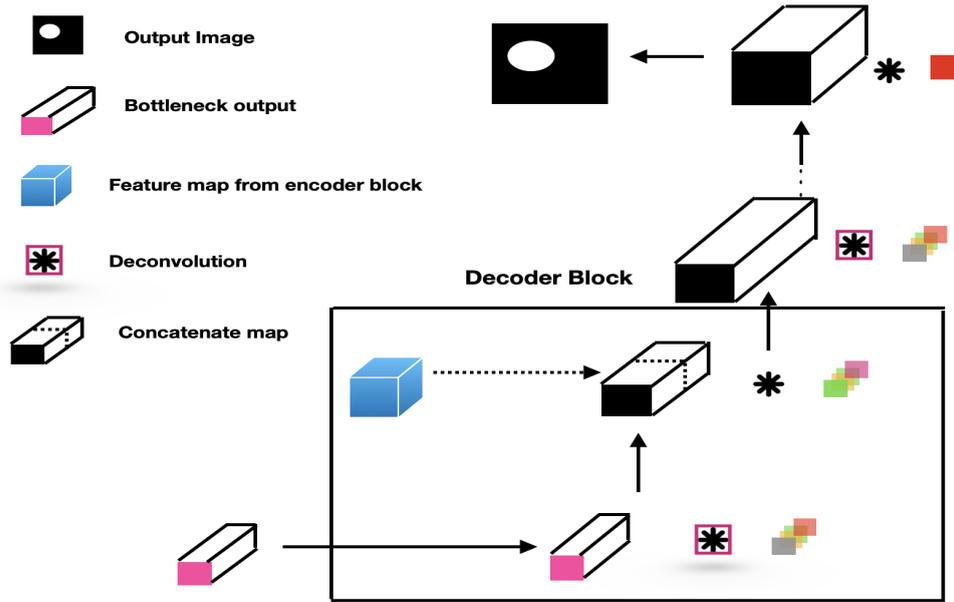


Figure 7. The expansive path

It consists of two 3x3 convolutions and a transposed convolution operation in each decoder block. We have already introduced the convolution in section 4.2. Here, we'll explain what's the transposed convolution, which is another kind of convolution, but in a backward direction. It aims to increase the dimensions of each element pixel in an image. Figure 8 displays the transposed convolution operation.

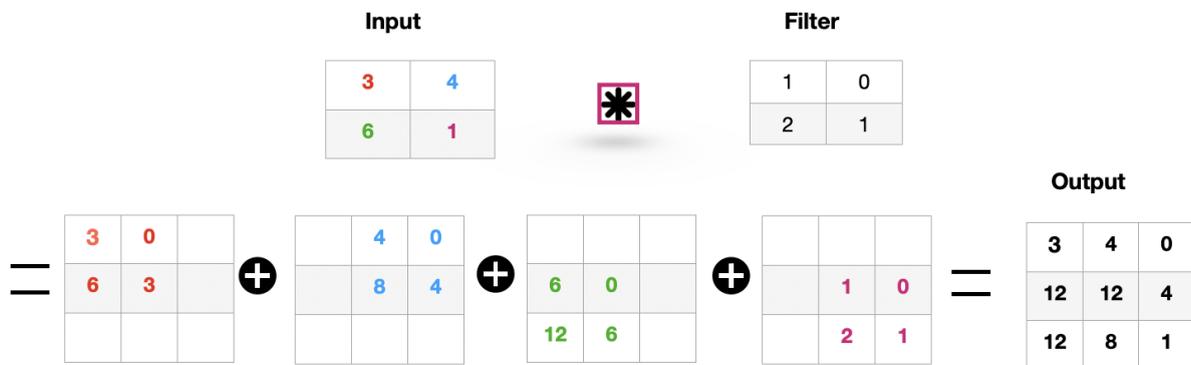


Figure 8. Transposed Convolution Example

We represent the  $l$ th decoder block as follows:

$$DT^l = Conv2DTranspose(D^{l-1}, Filters) \tag{9}$$

$$L^l = [FC_{map}^l, DT^l] \tag{10}$$

$$D^l = Conv2D(L^l, Filters) \tag{11}$$

Where  $FC_{map}^l$  is the feature map which connects to the  $l$ th layer in the decoder. At the final layer, there is a 2D Convolution with 1 filter and the activation function sigmoid, which can be seen as a binary classification for each pixel.

## 4.4 Training

The original U-Net architecture was modified in our solution by employing classifier networks to the bottleneck layer. Then, we'll discuss how to train the proposed (U-Net)+. Three loss functions are combined to train the model.

The first loss function is shown in equation (12) for the semantic segmentation. The input is the original breast cancer ultrasound image and the output is for the segmentation masks.

$$L_{seg} = - \sum_{i=1}^N \mathbf{y}_{true}^{(i)} \log \widehat{\mathbf{y}}_{pred}^{(i)} + (1 - \mathbf{y}_{true}^{(i)}) \log(1 - \widehat{\mathbf{y}}_{pred}^{(i)}) \quad (12)$$

Where  $\mathbf{y}^{(i)}$  is the  $i$ th sample's label, ( $\mathbf{y}^{(i)}=1$  if it's a tumour, and  $\mathbf{y}^{(i)}=0$  if it's background), and  $\widehat{\mathbf{y}}^{(i)} = P(\mathbf{y}^{(i)} = l|x)$ ,  $l = \{0, 1\}$ . The  $P(\mathbf{y} = l|x)$  can be computed by the pixel-wise sigmoid as follows:

$$P(\mathbf{y}^{(i)}) = 1 / (1 + \exp(-a_k(\mathbf{y}^{(i)}))) \quad (13)$$

$a_k(\mathbf{y}^{(i)})$  represents the activation in the feature channel  $k$  at the pixel position,  $P(\mathbf{y}^{(i)})$  and is the approximated maximum function.

The second loss function is for the classification of three categories, i.e. normal, benign and malignant. While the ultrasound images are the input and the output is their labels.

$$L_{UL} = \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^C q(\mathbf{X}^{(ij)}) \log(q(\widehat{\mathbf{X}}^{(ij)})) \quad (14)$$

$$q(\widehat{\mathbf{Y}}^{(ij)} | \mathbf{x}_i) = \frac{e^{x_i}}{\sum_{j=1}^C e^{x_j}} \text{ for } j = 1, 2, \dots, C \quad (15)$$

In equation(14),  $m$  is the number of samples for the ultrasound image for training.  $C$  is the number of categories.  $\mathbf{X}$  is the ultrasound images. In equation (15), Where  $x_i$  is the output of the dense layers before the classifiers, which is decided by the input of the contracting path.

The third one is also used for the classification of the three categories while the input is the segmentations (masks) images denoted as  $\mathbf{S}$ . Also  $p(\widehat{\mathbf{S}}^{(ij)})$  can be computed in the equation (13). But the  $\mathbf{x}_i$  is derived from the input images  $S$ .

$$L_S = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^C p(\mathbf{S}^{(ij)}) \log(p(\widehat{\mathbf{S}}^{(ij)})) \quad (16)$$

Therefore, the total loss functions are

$$L = aL_{seg} + bL_{UL} + cL_S \quad (17)$$

During training, different weights  $a$ ,  $b$ ,  $c$  are set for the three loss functions.

## 5. EXPERIMENTS

### 5.1 Data introduction

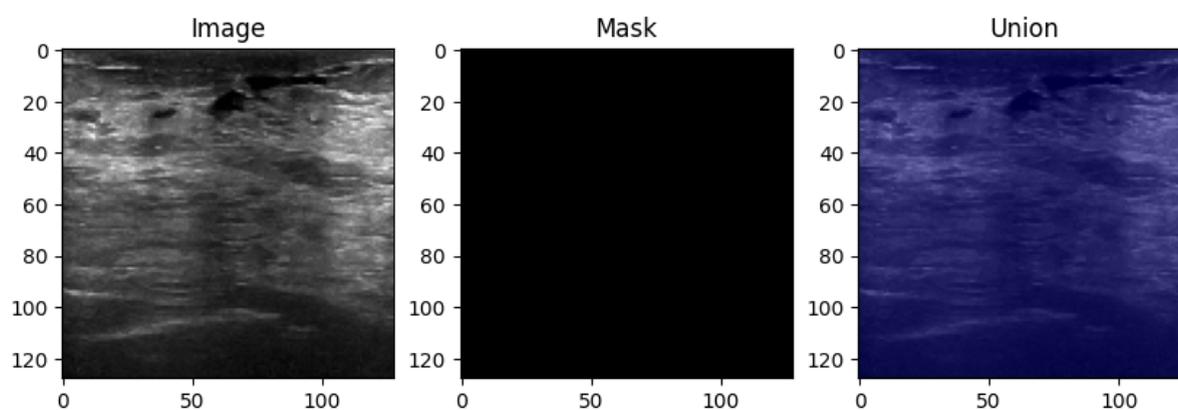
To prove the efficiency of our solution, we evaluated its performance on a public dataset of breast Ultrasound presented in [4]. The images are obtained through ultrasound scans, a safe and commonly used technique for breast cancer examination and early detection, especially when compared to other radiological imaging methods. This dataset was collected in 2018 including 780 images taken from women aged between 25 and 75. Among the 780 images, there were 600 female patients, and the images have an average size of 500x500 pixels, stored in PNG format.

These ultrasound images are categorized into three classes: normal, benign, and malignant, making

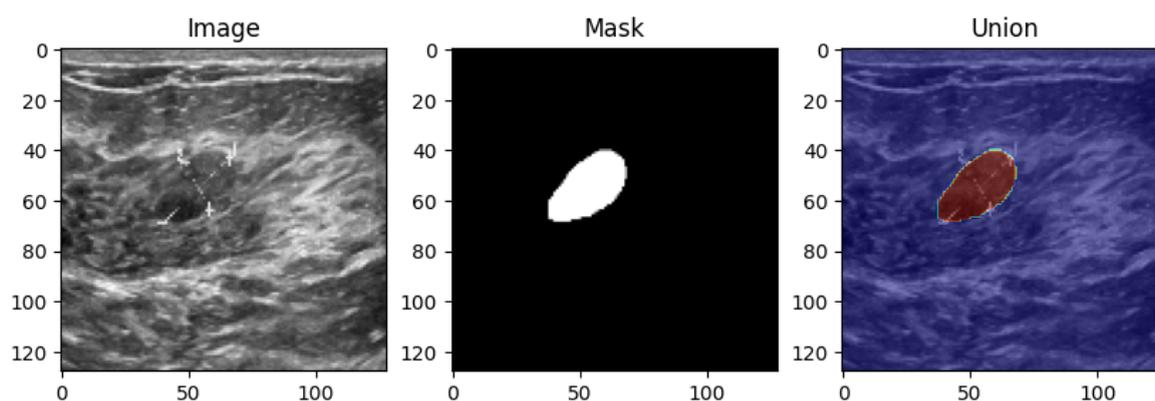
them suitable for classification and detection. In addition, corresponding masks are provided for the original images. In some cases, multiple masks are available, which may be due to the presence of multiple tumours in certain samples. Table 1 shows the distribution of these categories and Figure 9 displays a selection of sample images from the dataset.

Table 1. Dataset distribution according to categories

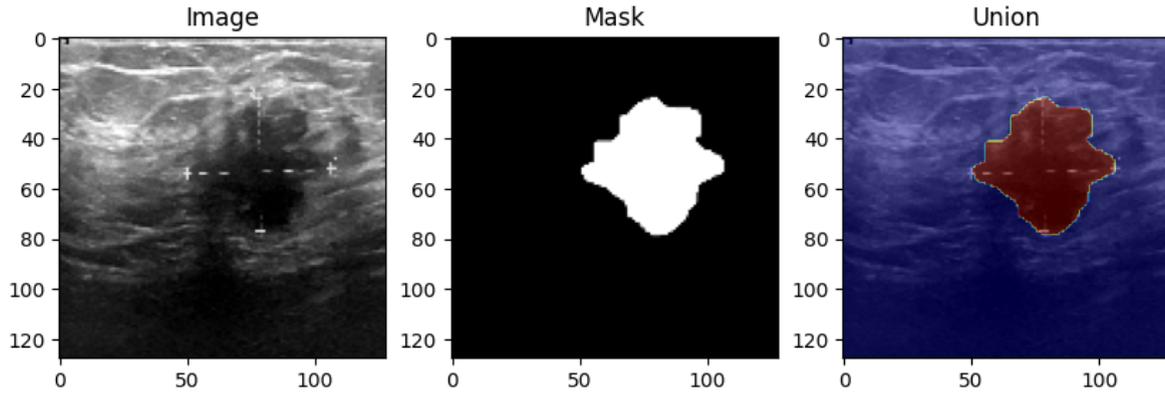
Category	Number of images	Number of masks
normal	133	133
Benign	437	454
malignant	210	211
Total	780	798



(a)Normal Class



(b)Benign Class



(c) Malignant Class

Figure 9. Some samples from the dataset

### 5.2 The Network settings in our experiments

We implemented an encoder-decoder network with a bottleneck layer as our fundamental architecture. The contracting path incorporates several encoder blocks to generate feature maps, with the bottleneck layers serving as input for the classifier. Moreover, the bottleneck layer also provides input to the expansive path. The settings and parameters for our (U-Net)+ are displayed in Table 2.

Table 2. The (U-Net)+ Settings

Parts	Contracting path	Bottleneck layer	Expansive path
Number of subblocks	4	1	4
filters	[64,128,256,512]	1024	[512,256,128,64]
Activation function	ReLU	Softmax	Sigmoid
parameters	Filters size(3,3) strides:2 Padding: same	----	Filters size (2,2) Strides =2 Padding: same
----	Optimizer: Adam; Loss: Binary_crossentropy; metrics: accuracy		

### 5.3 Performance of classification on ultrasound images

In this section, we evaluate the performance of the first classifier, which is designed to identify the three categories of ultrasound samples. Early detection serves as an important step in providing evidence for further diagnosis. This classifier can be seen as a rapid initial diagnostic tool. The training history and classification results (displayed with a confusion matrix) are shown in Figure 10.

A confusion matrix presents a table layout of the different outcomes between the true labels and the predicted ones. It helps to evaluate and visualise the classification performance easily.

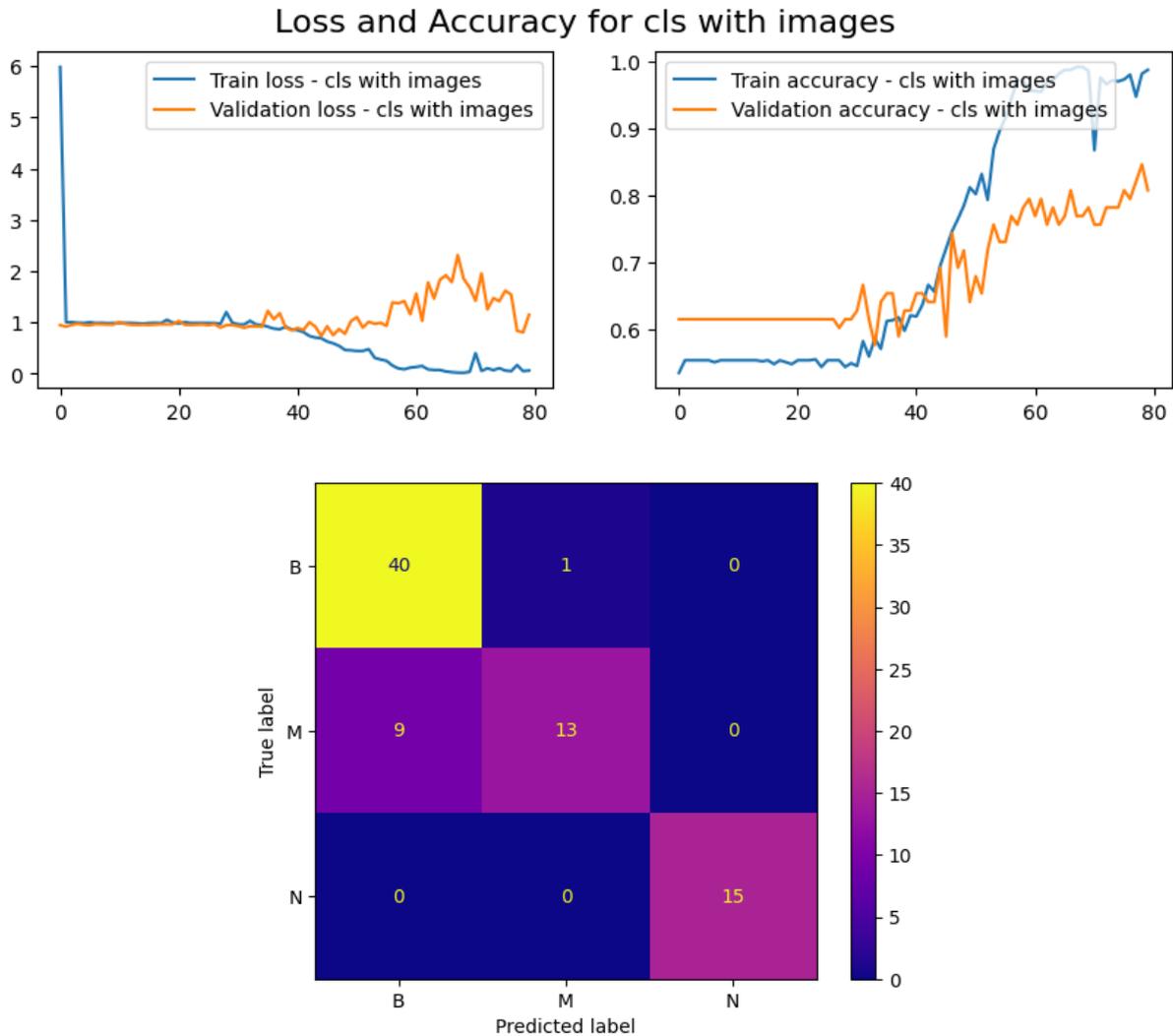


Figure. 10 The initial classification performance on Ultrasound images

From the training history lines in Figure 10, we observe that the loss consistently exhibits a decreasing trend, indicating that the model is learning and improving over epochs. However, it's important to acknowledge that, in the case of validation samples (depicted by the yellow line), the loss doesn't stabilize until around the 80th epoch. This implies that the model's performance on the validation set takes some time to reach a steady state. The classification accuracy for the ultrasound images exceeds 80% for the validation images, suggesting that the model is performing well. The confusion matrix displayed more details. For the samples labelled as "normal", the model correctly predicts all of them as "normal". This demonstrates the model's proficiency in identifying normal cases. In the case of samples labelled as "Benign", 40 samples are classified correctly, and there is one sample that is incorrectly recognized as "Malignant". When dealing with samples labelled as "Malignant", the model correctly identifies 13 of them as "Malignant". Unfortunately, 9 samples are classified as "Benign" incorrectly. It is very risky for further treatments.

#### 5.4 Performance of Semantic Segmentation

Semantic segmentation is a critical task in medical image analysis that aims to label each pixel in the image. This process is greatly helpful in aiding more accurate diagnoses and treatments for various medical conditions. In this section, we assess the segmentation performance both visually and quantitatively. Figure 11 displays the training history for the segmentation module and Figure 12

presents the visual segmentation results. Table Table 3 displays the quantitative results for the two categories. They are segmentation accuracy (ACC), Specificity (SP) and Sensitivity (SE) [18].

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

$$Sp = \frac{TN}{TN+FP}$$

$$Acc = \frac{TP+TN}{TP+FN+TN+FP}$$

Where TP, true positives, are the tumour pixels predicted correctly. FP(false positive), are the background pixels predicted as tumour pixels incorrectly. Therefore, TN (true negative) means the background pixels are predicted correctly, and FN (false negatives) are the pixels incorrectly predicted as background.

Table 3. Quantitative results for two categorical variables

Model	Se	Sp	Acc
Benign	0.927	0.997	0.992
Malignant	0.906	0.995	0.986
Benign+Malignant	0.920	0.997	0.990

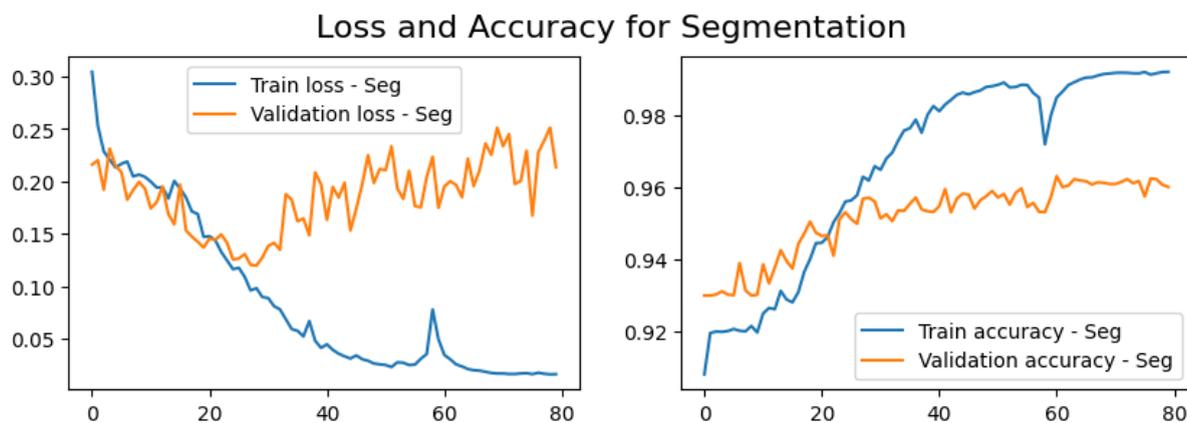


Figure. 11 Training history for the segmentation module

From the loss trends, we observe fluctuations in the validation loss when compared to the training loss, with a slight upward trend. However, it's worth noting that the model's accuracy remains consistently high, hovering around 96%. This accuracy stabilizes after the 20th epoch, indicating that the model has reached a stationary state in its learning process.

Our model displayed a good performance in the segmentation task as shown in Figure 12. It effectively preserves crucial information, including tumour shapes and edges, which serve as vital evidence for subsequent classification tasks. However, a classification error is shown in Figure 12, where a "Benign" case was incorrectly identified as "Malignant". This misclassification may be attributed to irregular or non-smooth edges in the segmentation. To address this issue, extending the training could potentially lead to more accurate classifications.

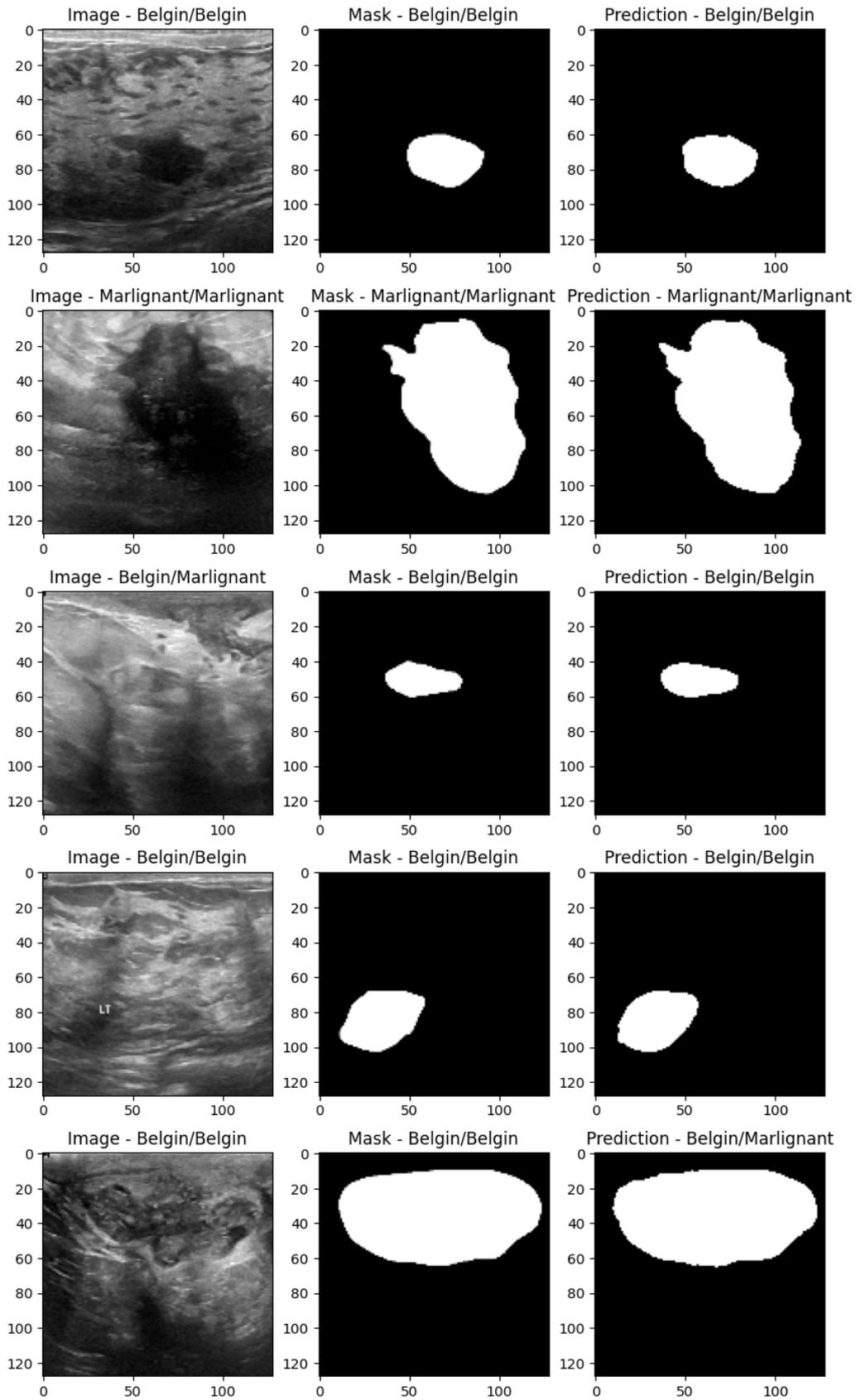


Figure 12. Segmentation and classification results

### 5.5 Performance of the second classifier on segmented images

We expect the classifier’s accuracy will be improved when applied to the segmentations, as the noise or disturbance is removed. Figure 13 displays the training history and presents the confusion matrix for the ACC.

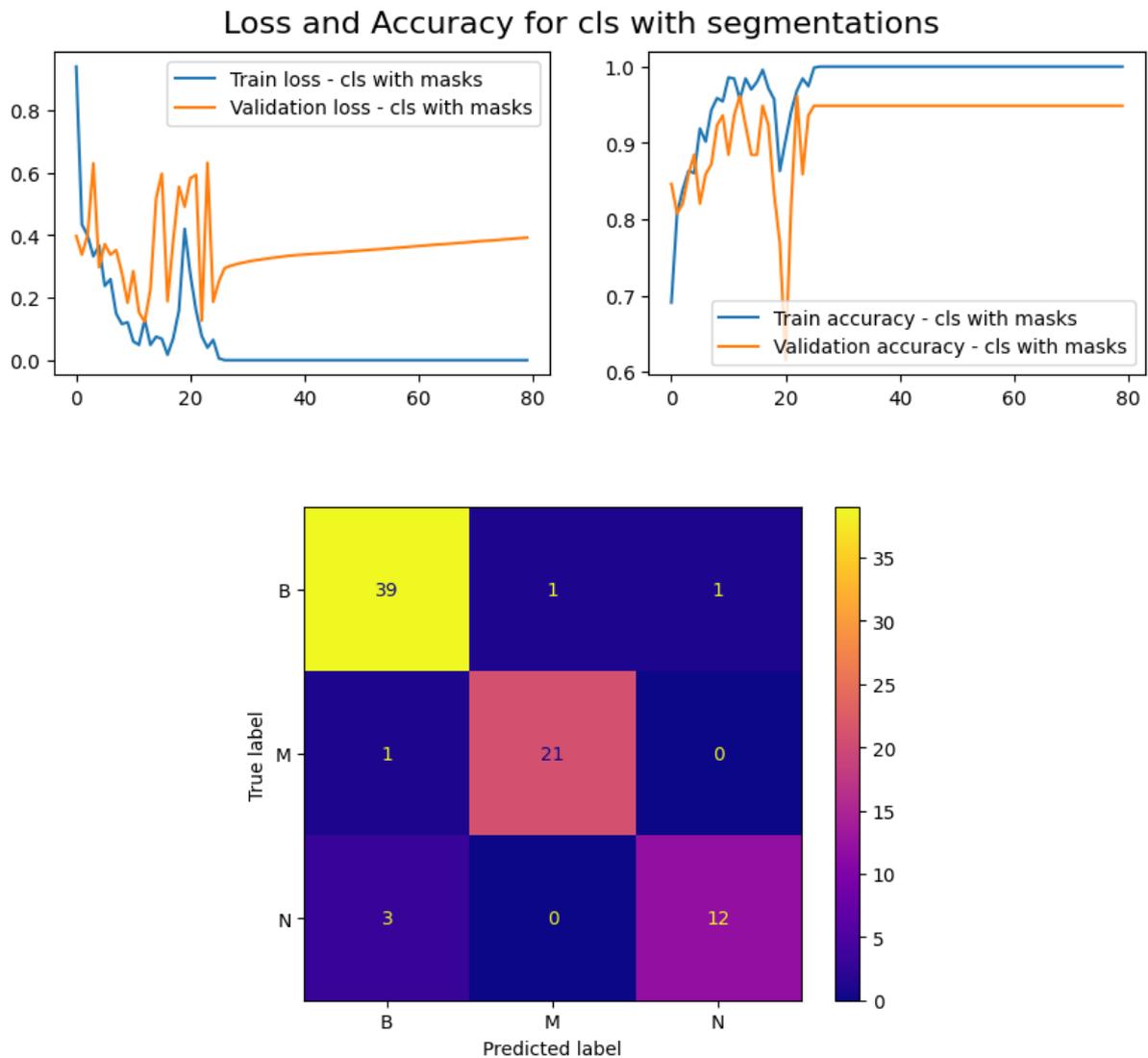


Figure 13. Classification performance based on segmentation images

Figure 13 shows that the convergence performance with notable stability observed in the loss and accuracy lines, reaching a stationary state around the 20th epoch. The classification accuracy stands at approximately 96%, representing an improvement over the initial classification. The confusion matrix provides a more detailed breakdown of the results. For samples labelled as “normal”, 12 of them were accurately predicted, and 3 were incorrectly classified as “benign”. This indicates that all 15 samples were diagnosed as “Normal”, and these misclassifications are unlikely to significantly impact further treatments.

In contrast, for samples labelled as “Malignant”, 21 were correctly predicted, and 1 was misclassified as “benign”. The accuracy for this category of samples is notably high at about 95.5%. However, the signal incorrect prediction could potentially lead to a delay in treatment. Comparing this classification to the initial assessment of ultrasound images, the error rate has significantly decreased substantial progress in the model’s accuracy and the potential to improve patient care.

## CONCLUSIONS

In this paper, we introduce a novel system, denoted as (U-Net)+, which has been trained to handle both classification and semantic segmentation tasks for breast cancer diagnosis. To understand the U-Net well, we draw an analogy between the U-Net architecture and the process of seed germination and growth, visually illustrating its key components. In our approach, we've made modifications to the original U-Net architecture by incorporating fully connected layers, thereby enabling it to perform classification tasks alongside segmentation. Our training process involves the use of three distinct loss functions, which are integrated into a total loss function. Experimental results validate the efficiency and effectiveness of our novel (U-Net)+ model.

## REFERENCES

- [1]World Health Organization: WHO. (2023, February 3). WHO launches new roadmap on breast cancer. *World Health Organization: WHO*. <https://www.who.int/news/item/03-02-2023-who-launches-new-roadmap-on-breast-cancer>
- [2]*Why is early cancer diagnosis important?* (2015, April 2). Cancer Research UK. <https://www.cancerresearchuk.org/https%3A/www.cancerresearchuk.org/about-cancer/spot-cancer-early/why-is-early-diagnosis-important>.
- [3]Wang, E. K., Chen, C.-M., Hassan, M. M., & Almogren, A. (2020). A deep learning based medical image segmentation technique in Internet-of-Medical-Things domain. *Future Generation Computer Systems*, 108, 135–144. <https://doi.org/10.1016/j.future.2020.02.054>
- [4]Khalili, N. (n.d.). *Machine learning for automatic segmentation of neonatal and fetal MR brain images* [Utrecht University Library]. Retrieved October 7, 2023, from <http://dx.doi.org/10.33540/230>
- [5]Islam, M. M. U., & Indiramma, M. (2020, September). Retinal Vessel Segmentation using Deep Learning – A Study. *2020 International Conference on Smart Electronics and Communication (ICOSEC)*. <http://dx.doi.org/10.1109/icosec49089.2020.9215378>
- [6]- Incremental engineering of lung segmentation systems. (2016). In *Lung Imaging and Computer Aided Diagnosis* (pp. 30–63). CRC Press. <http://dx.doi.org/10.1201/b11106-5>
- [7]Cai, Y., & Yuan, J. (2022, September 23). A review of u-net network medical image segmentation applications. *Proceedings of the 2022 5th International Conference on Artificial Intelligence and Pattern Recognition*. <http://dx.doi.org/10.1145/3573942.3574048>
- [8]Mohit Angurala. (2023). *Review for “Medical image segmentation method based on multi-scale feature and U-net network.”* <https://doi.org/10.1002/itl2.451/v2/review1>
- [9]*Why is early cancer diagnosis important?* (2015, April 2). Cancer Research UK. <https://www.cancerresearchuk.org/https%3A/www.cancerresearchuk.org/about-cancer/spot-cancer-early/why-is-early-diagnosis-important>
- [10] Razmjooy, N., & Rajinikanth, V. (2022). Medical-imaging-supported disease diagnosis. In *Frontiers of Artificial Intelligence in Medical Imaging* (pp. 2-1-2–32). IOP Publishing. <http://dx.doi.org/10.1088/978-0-7503-4012-0ch2>
- [11]Wu, Q., & Markey, M. (n.d.). Computer-Aided diagnosis of breast cancer on MR imaging. In *Recent Advances in Breast Imaging, Mammography, and Computer-Aided Diagnosis of Breast Cancer* (pp. 739–762). SPIE. Retrieved October 7, 2023, from <http://dx.doi.org/10.1117/3.651880.ch22>
- [12]Berglin, S., Shin, E., Raicu, D., & Furst, J. (2019, March 13). Efficient learning in computer-aided diagnosis through label propagation. *Medical Imaging 2019: Computer-Aided Diagnosis*. <http://dx.doi.org/10.1117/12.2512803>

- [13]Olaf R, Philipp F,Thomas B. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention, pages 234–241. Springer, 2015.
- [14]Su, R., Zhang, D., Liu, J., & Cheng, C. (2021). MSU-Net: Multi-Scale U-net for 2D medical image segmentation. *Frontiers in Genetics*, 12. <https://doi.org/10.3389/fgene.2021.639930>
- [15]Qin, X., Zhang, Z., Huang, C., Dehghan, M., Zaiane, O. R., & Jagersand, M. (2020). U2-Net: Going deeper with nested U-structure for salient object detection. *Pattern Recognition*, 106, 107404. <https://doi.org/10.1016/j.patcog.2020.107404>
- [16] Contributors to Wikimedia projects. (2023a, September 20). *Sigmoid function*. Wikipedia. [https://en.wikipedia.org/wiki/Sigmoid\\_function](https://en.wikipedia.org/wiki/Sigmoid_function)
- [17]Contributors to Wikimedia projects. (2023, October 3). *Softmax function*. Wikipedia. [https://en.wikipedia.org/wiki/Softmax\\_function](https://en.wikipedia.org/wiki/Softmax_function)
- [18]Jinyong Hou, [Xuejie Ding](#), [Jeremiah D. Deng](#): Semi-Supervised Semantic Segmentation of Vessel Images using Leaking Perturbations. [WACV 2022](#): 1769-1778.

## AUTHORS

Minuo Qing. She is a student from Westlake Girls High School in Year 12. She is passionate about promoting gender equality in Afghanistan and fundraising for autistic children. Throughout these endeavours, she has discovered the threat that breast cancer poses to women’s lives. She is attempting to enhance diagnostic efficiency using AI while also hoping to raise awareness among more peers about this important issue.

