Breast cancer is one of the most common causes of death in women worldwide. Early diagnosis helps to reduce the number of early deaths. The data examines the medical images of breast cancer using ultrasound scan.

Dataset:

Data collected at baseline included breast ultrasound images among women aged 25 to 75 years. This data was collected in 2018. The number of patients is 600 female patients. The dataset contains 780 images with an average image size of 500 x 500 pixels. Images are in PNG format. Background images (masks) are provided with original images. Images are classified into three categories: normal, benign and malignant.

Data examines medical images of breast cancer using ultrasound scan. Breast ultrasound datasets are classified into three categories: normal, benign and malignant. Breast ultrasound images can produce very good results in the classification, diagnosis and division of breast cancer if combined with artificial intelligence methods.

This dataset is called Breast Ultrasound Images Dataset or Dataset_BUSI_with_GT, which can be accessed from the following link:

https://www.kaggle.com/datasets/aryashah2k/breast-ultrasound-images-dataset

GT = Ground Truth

Breast ultrasound images can produce very good results in the classification, diagnosis and segmentation of breast cancer when combined with machine learning. The purpose of medical image segmentation is to identify important or suspicious areas in medical images. However, we usually face many challenges when developing networks for this type of analysis. First, it is important to preserve the original image resolution for this task, as the detection of subtle features or anomalies can significantly affect the detection accuracy. Recently, the widespread adoption of data-driven models has led to a significant increase in the exploration and development of AI (artificial intelligence)-assisted computer-aided diagnosis

(CAD) systems. Using artificial intelligence-assisted CAD systems, radiologists may be able to combine computer insights with their expertise and enable more accurate and faster evaluation. Such intelligent systems often work with biomedical imaging techniques, such as X-rays, CT scans, and MRI specimens. One of the most popular networks used in deep learning is convolutional neural network.

In addition, the article approved by Dr. Eslami for conducting this research can be found in the following:

Google Drive link Article :

https://drive.google.com/file/d/19BKddd8Q8RwJoTYg6xsHZTg034NCwxq3/view?usp=drivesdk

Google Drive link: Video explaining the article:

https://drive.google.com/file/d/1A7aqsWWJX88XZKhOXF4dgi6MQhhmBlRT/view?usp=drivesdk

Google Drive link: Basic code implementation:

https://drive.google.com/drive/folders/1yXS4Tp2Iywd1It-dAo7hXEJXH3GxCSuo

Google Drive link: full implementation of the code:

https://drive.google.com/drive/folders/1fIrbMz1RrFdSqoRfO8dHKZexkc_MRkSy

Google Drive link: code implementation video:

https://drive.google.com/file/d/1TvlkfqxsOvhaLnuqhe00WpUD8yfXzOUh/view?usp=drivesdk

The simulation results are as follows:

1) preprocessing

Then the preprocessing operation is performed using the following:

- Change the size of images and masks.
- Convert images to RGB format (if not already).

- Normalize the pixel values of images and masks.
- Convert images and masks to numpy arrays.

Then the data is read using the path defined in the drive of each file. In addition, the mask and each image are stored in a dedicated list. Then, the dimensions of the list related to the images and mask are checked, and the order and position of the data in the list is changed so that it does not overfitting in the model training.

In this Project 90% of data are selected as training data, 7.5% as validation data and 2.5% as test data.

Then the dimensions and data of training and validation and testing are determined as follows:

Training data:

(702, 256, 256, 3)

(702, 256, 256)

Validation data:

(58, 256, 256, 3)

(58, 256, 256)

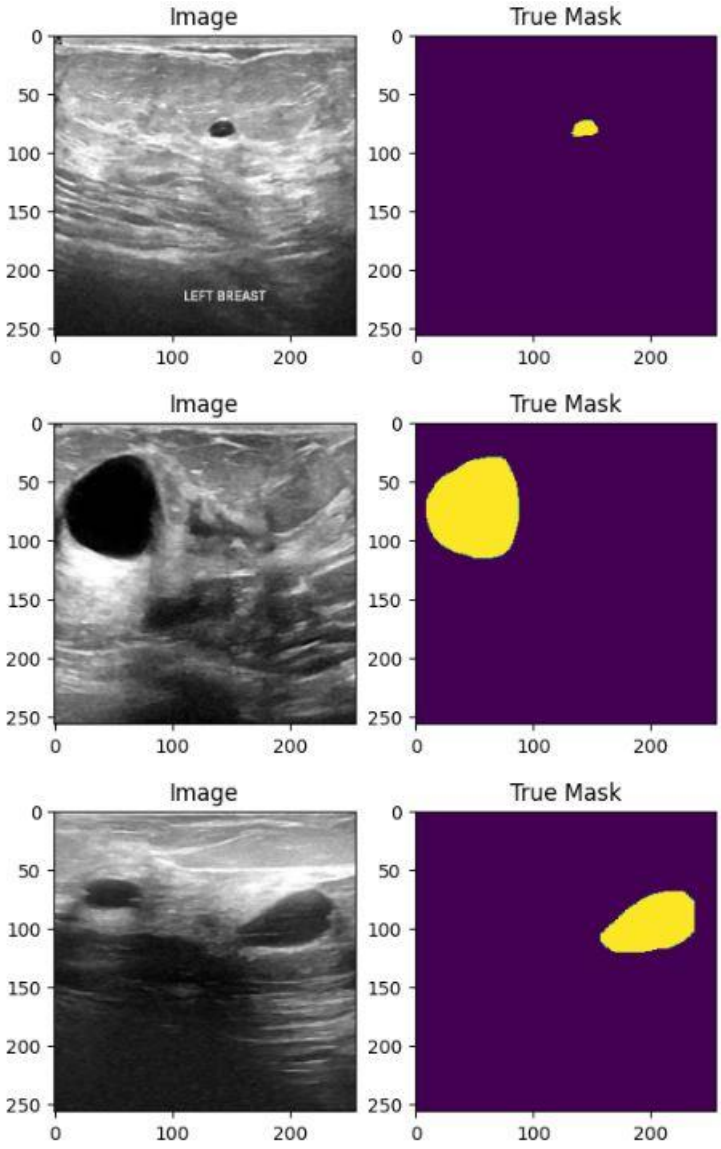
Test data:

(20, 256, 256, 3)

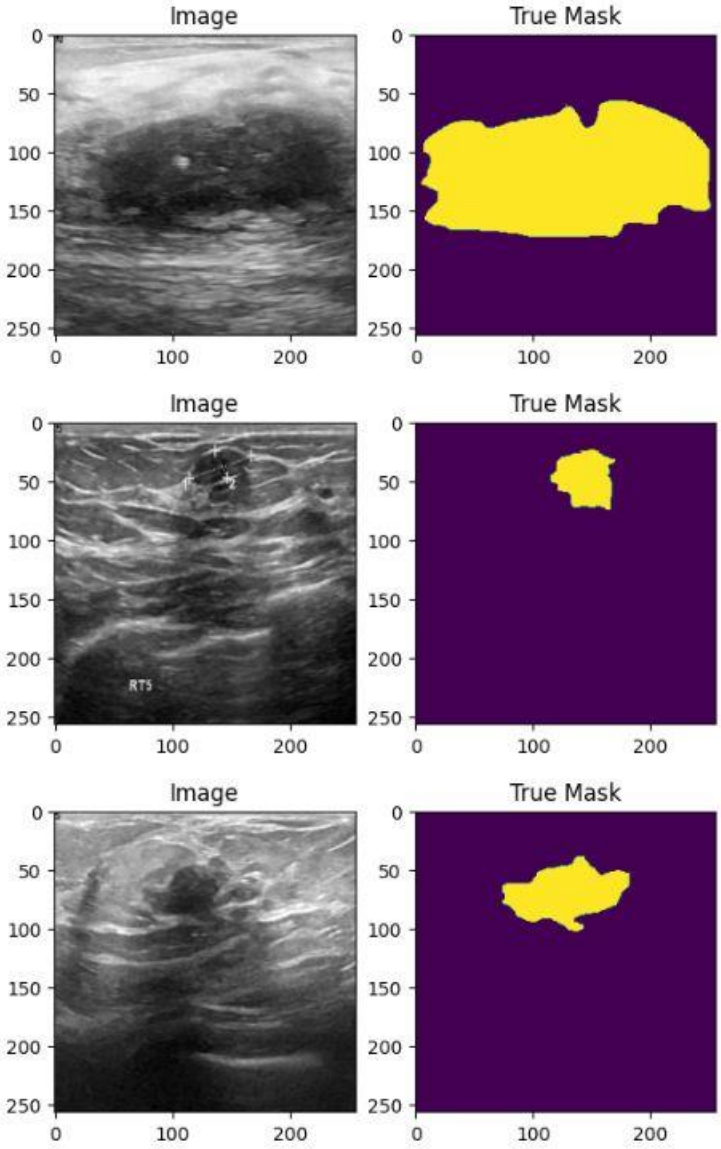
(20, 256, 256)

Display data related to each category:

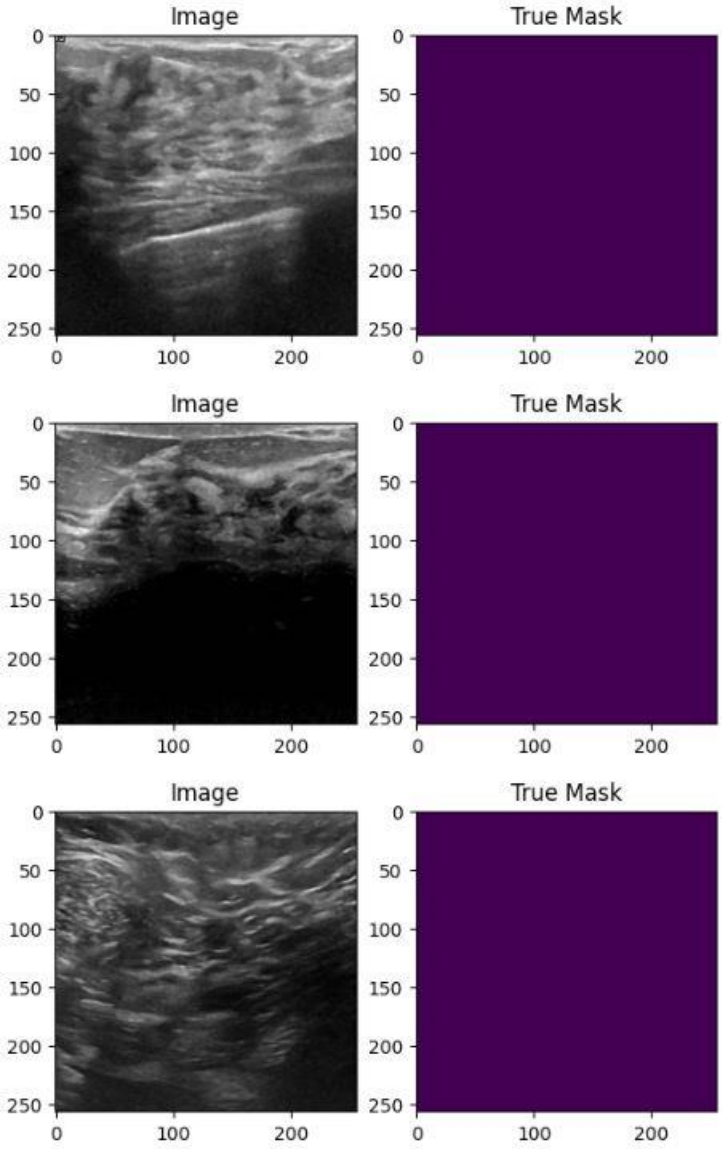
Images of breast cancer in the benign category with the corresponding mask:



Images of breast cancer in the malignant category with the corresponding mask:



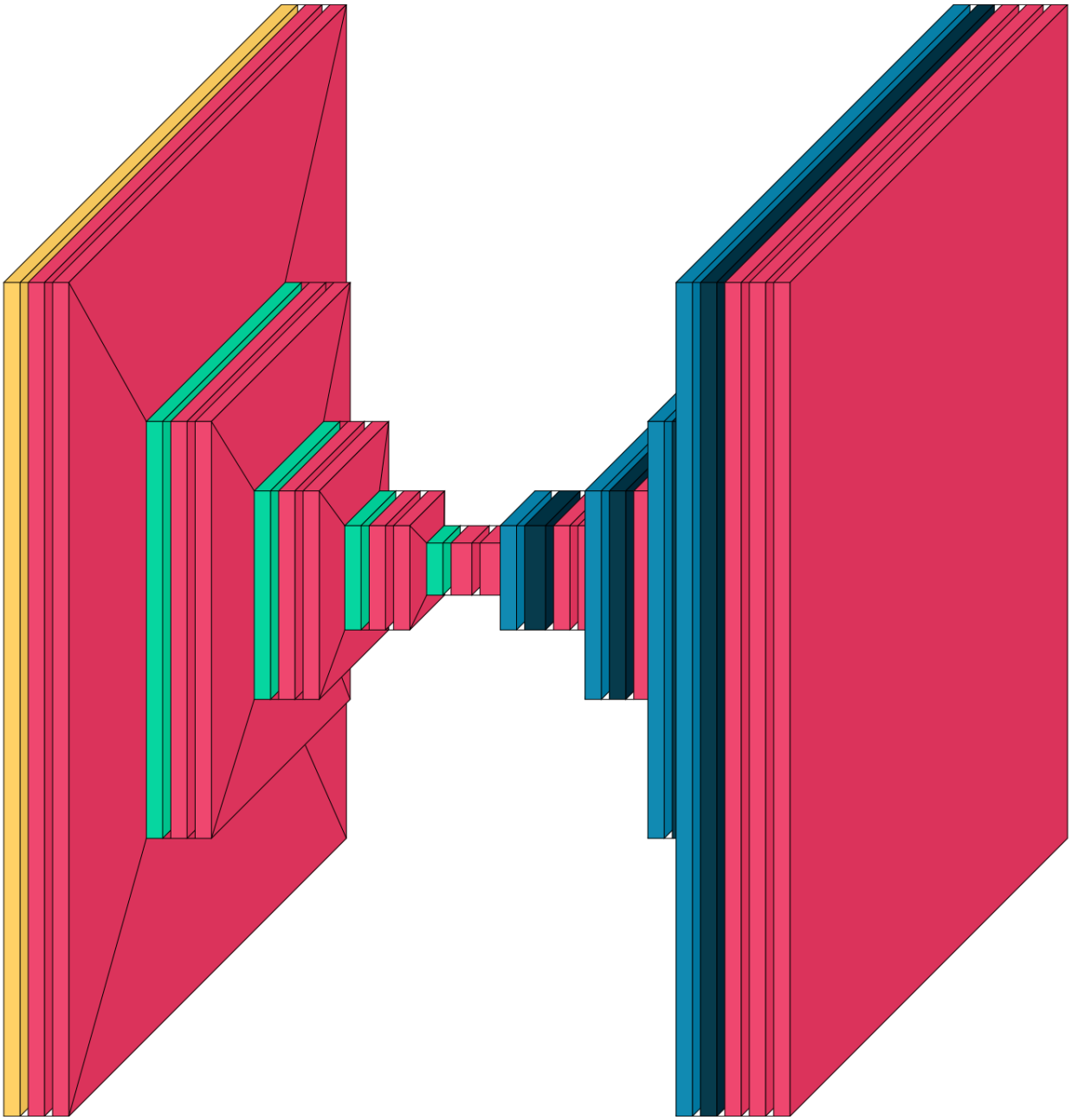
Images of breast cancer in the healthy group with the corresponding mask:



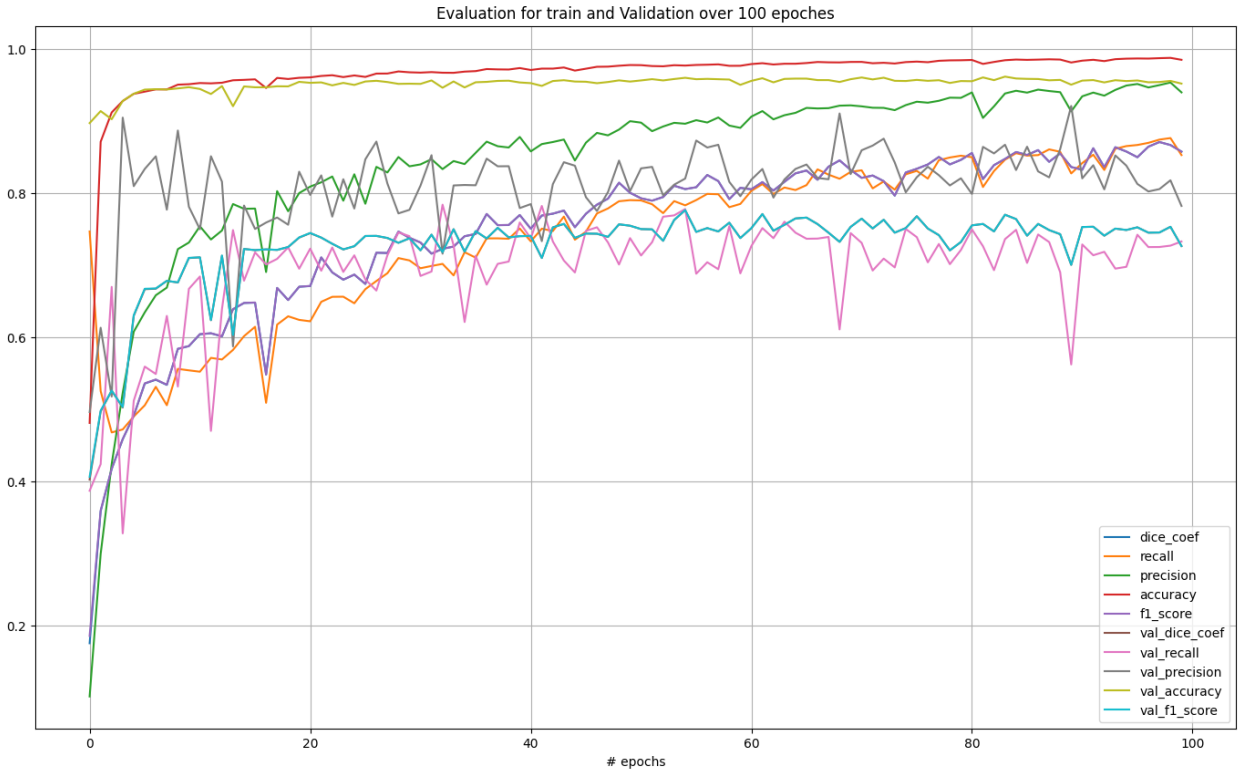
2) Model:

In recent years, deep learning has made great progress in medical image segmentation. In this regard, U-Net has been the most popular architecture in medical image segmentation. The U-Net network has shown outstanding overall performance in medical image segmentation through extensive testing on challenging datasets. Improved U-Net networks in special applications, in addition

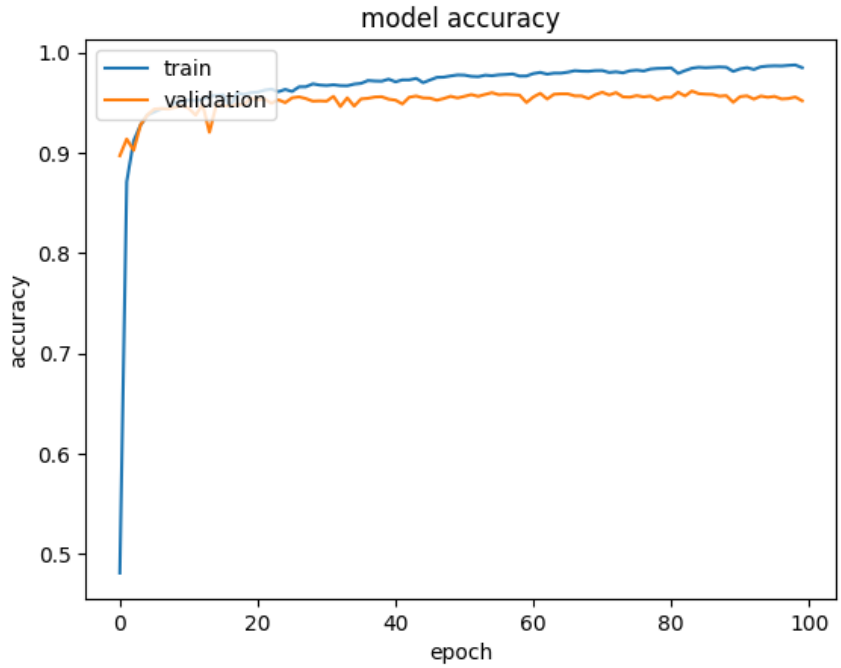
to the advantages of the U-Net network, benefit from the improved advantages created in the network according to the specific application. Therefore, some modifications have been proposed to improve the U-Net model in various articles.

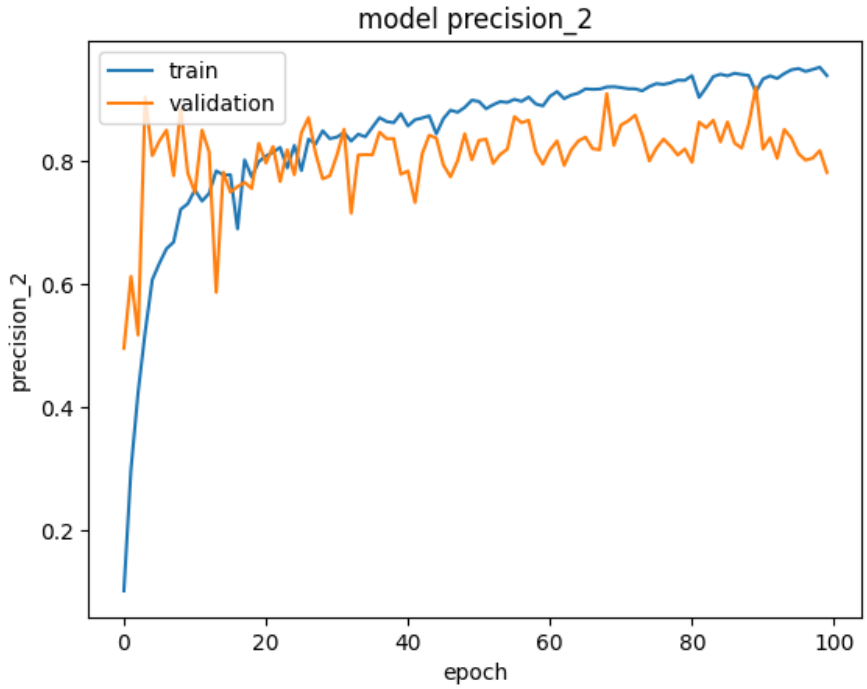


Evaluation of model performance after 100 Epoch:

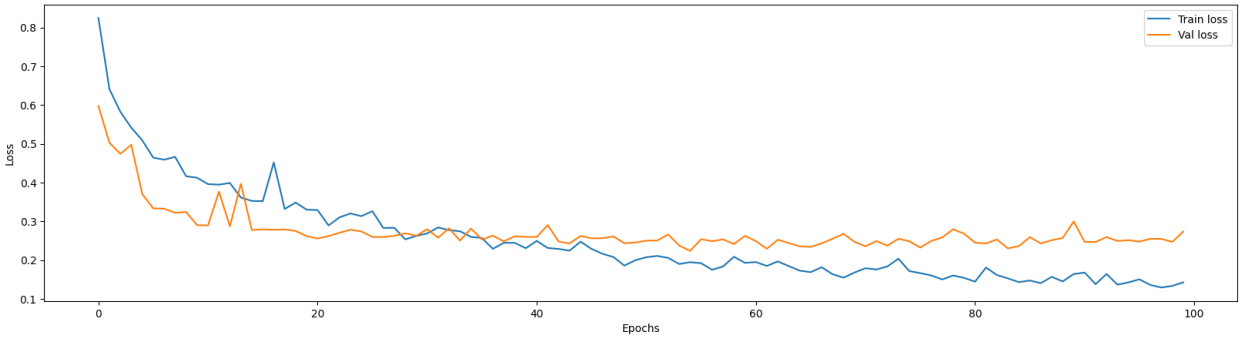


Accuracy diagram of the model:


model accuracy
train
validation
accuracy
epoch
1.0
0.9
0.8
0.7
0.6
0.5
0
20
40
60
80
100


model precision_2
train
validation
precision_2
epoch
0.8
0.6
0.4
0.2
0
20
40
60
80
100

Loss function diagram:



Model evaluation on test data:

The evaluation of the accuracy of the test data is as follows:

Loss: 0.2826

Dice_coef: 0.7174

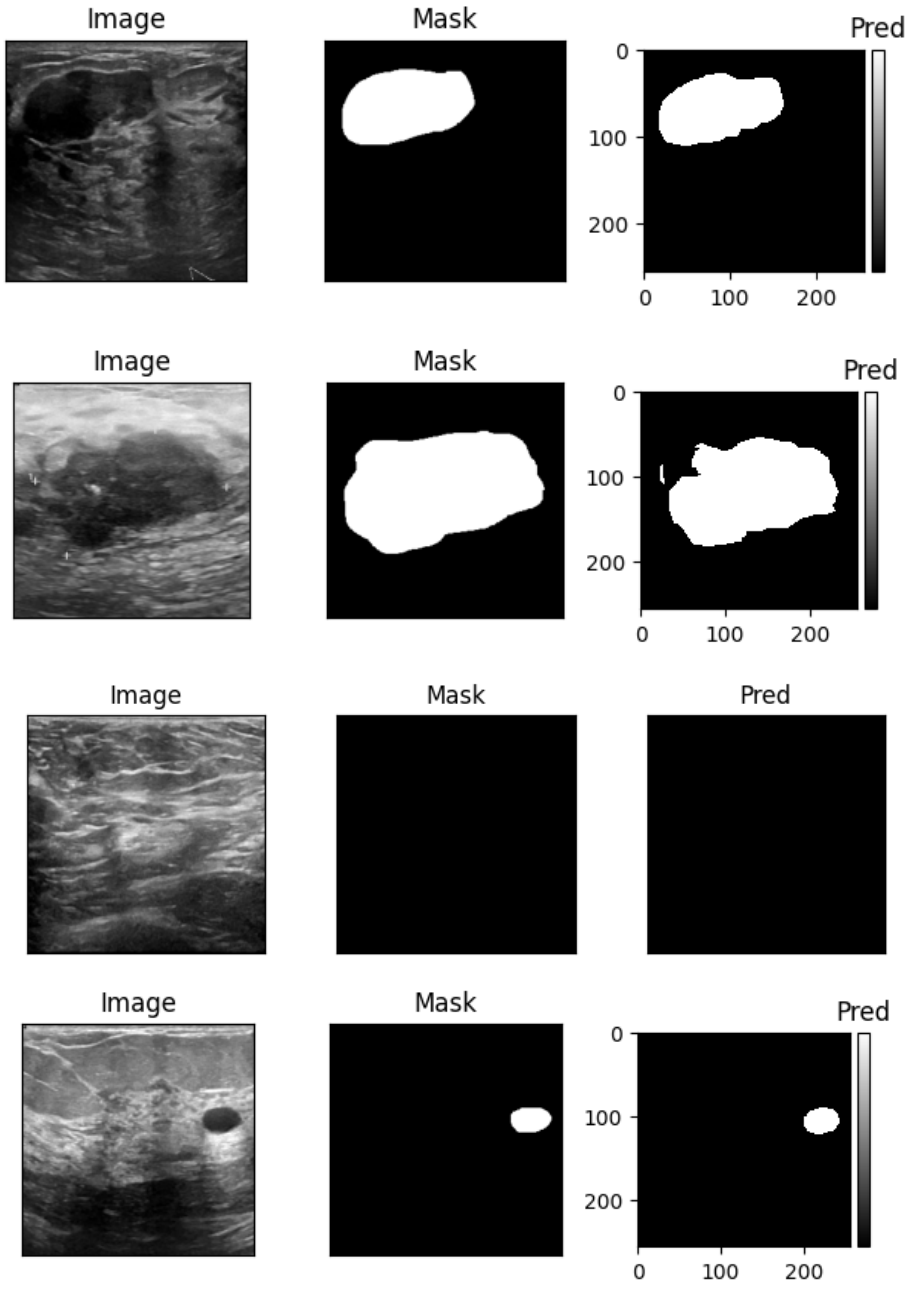
Recall: 0.7536

Precision: 0.6847

Accuracy: 0.9662

F1_score: 0.7175

Also, some examples of test data along with the original mask and the predicted mask are shown in the figure below:



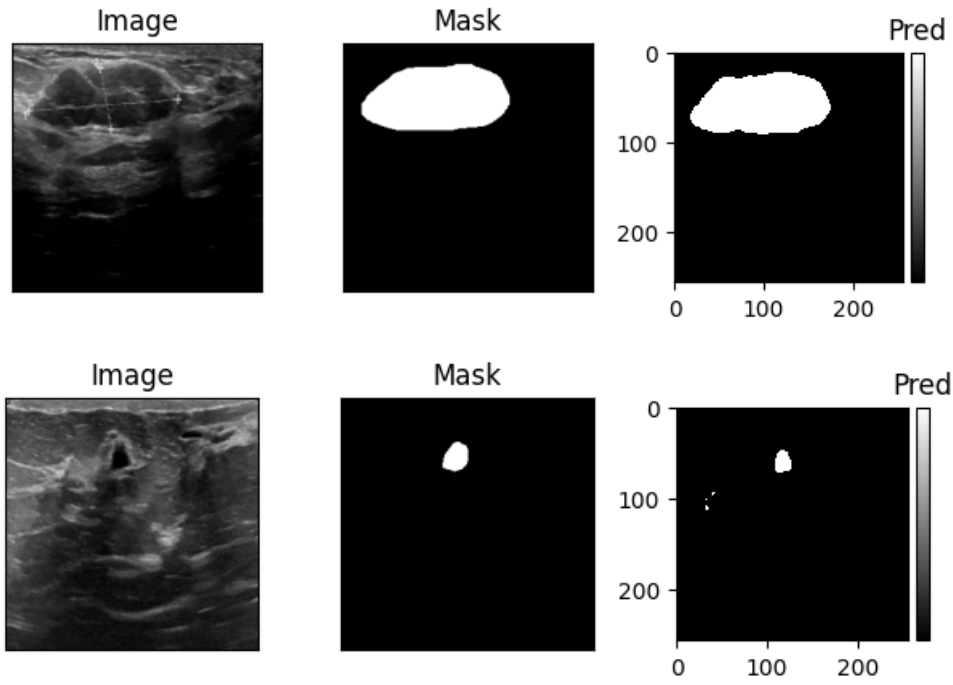


Image
Mask
Pred
0
100
200
0
100
200
Image
Mask
Pred
0
100
200
0
100
200