Breast Cancer Segmentation and Classification

By Tanmay Thaker

About the Project:

- Breast cancer is one of the most common causes of death among women worldwide. Early detection helps in reducing the number of early deaths.

- The data review the medical images of breast cancer using an ultrasound scan.

- Breast Ultrasound Dataset is categorized into three classes: normal, benign, and malignant images.

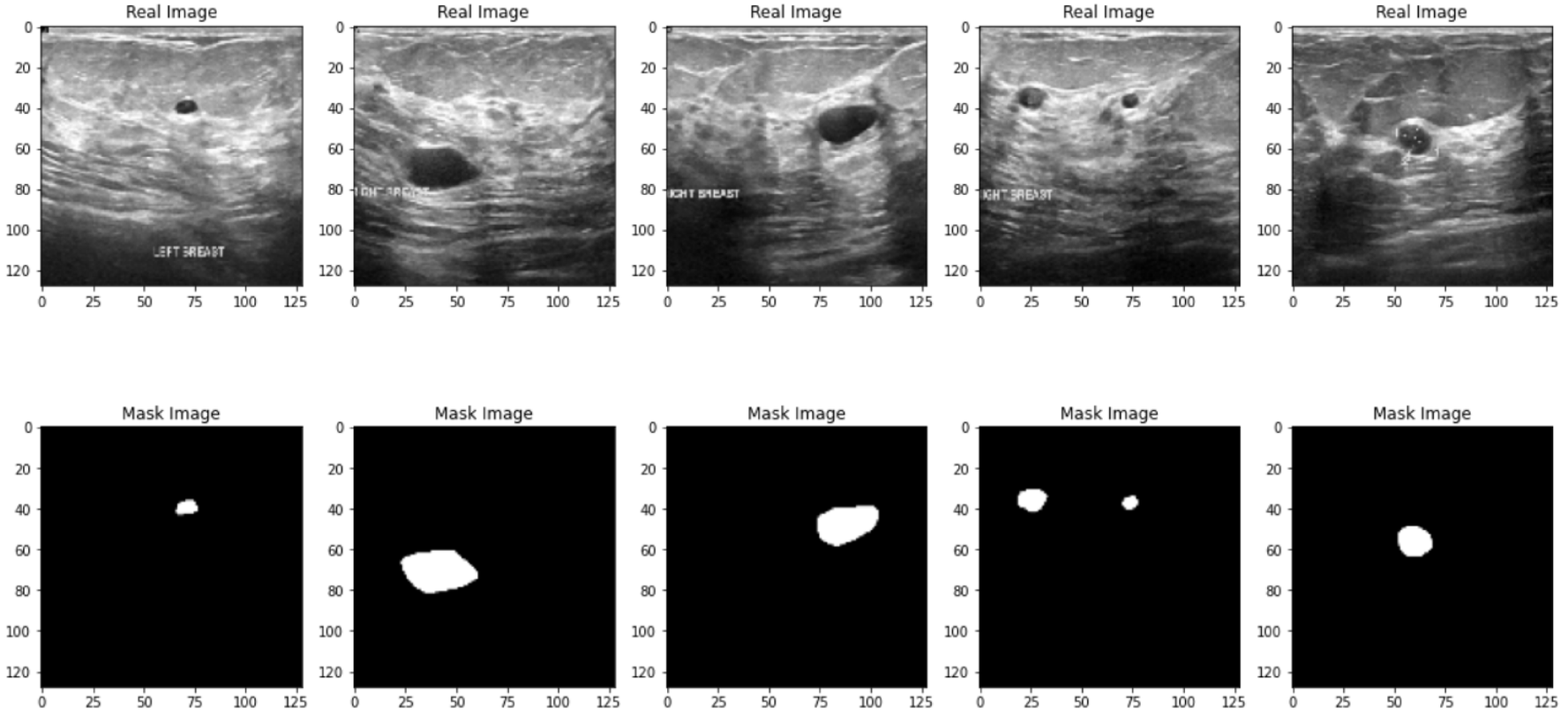
About the dataset

- The data collected at baseline include breast ultrasound images among women in ages between 25 and 75 years old. This data was collected in 2018.

- The dataset contains ultrasound of 600 female patients. The dataset consists of 780 images with an average image size of 500*500 pixels and the images are in PNG format.

- . The images are categorized into three classes, which are normal, benign, and malignant.

About the dataset -continue



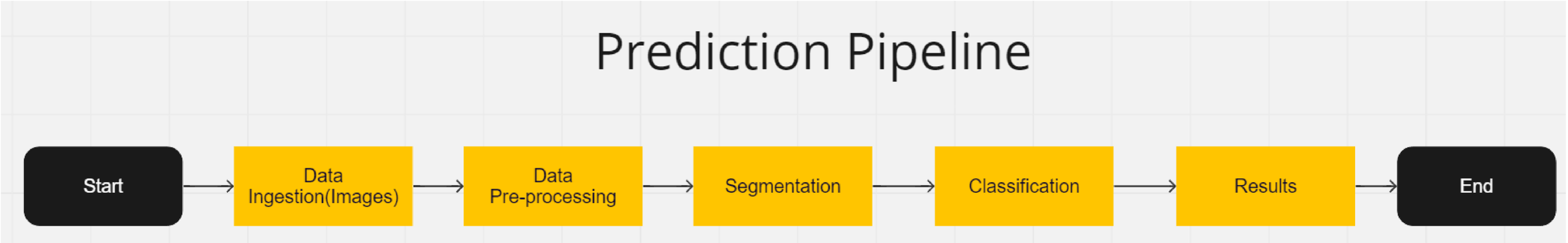
Proposed Approach

- The first step is to ingest the data and perform pre-processing on it like resizing the image or performing some data augmentation techniques like horizontal or vertical flip.

- In the next step, we will perform image segmentation using U-Net which will give us the segmented image which is our region of interest.

- Now, we will take these segmented images and make our classification model which will take these images as input and will classify if the given ultrasound image is normal, benign, or malignant.

How the pipeline will work?



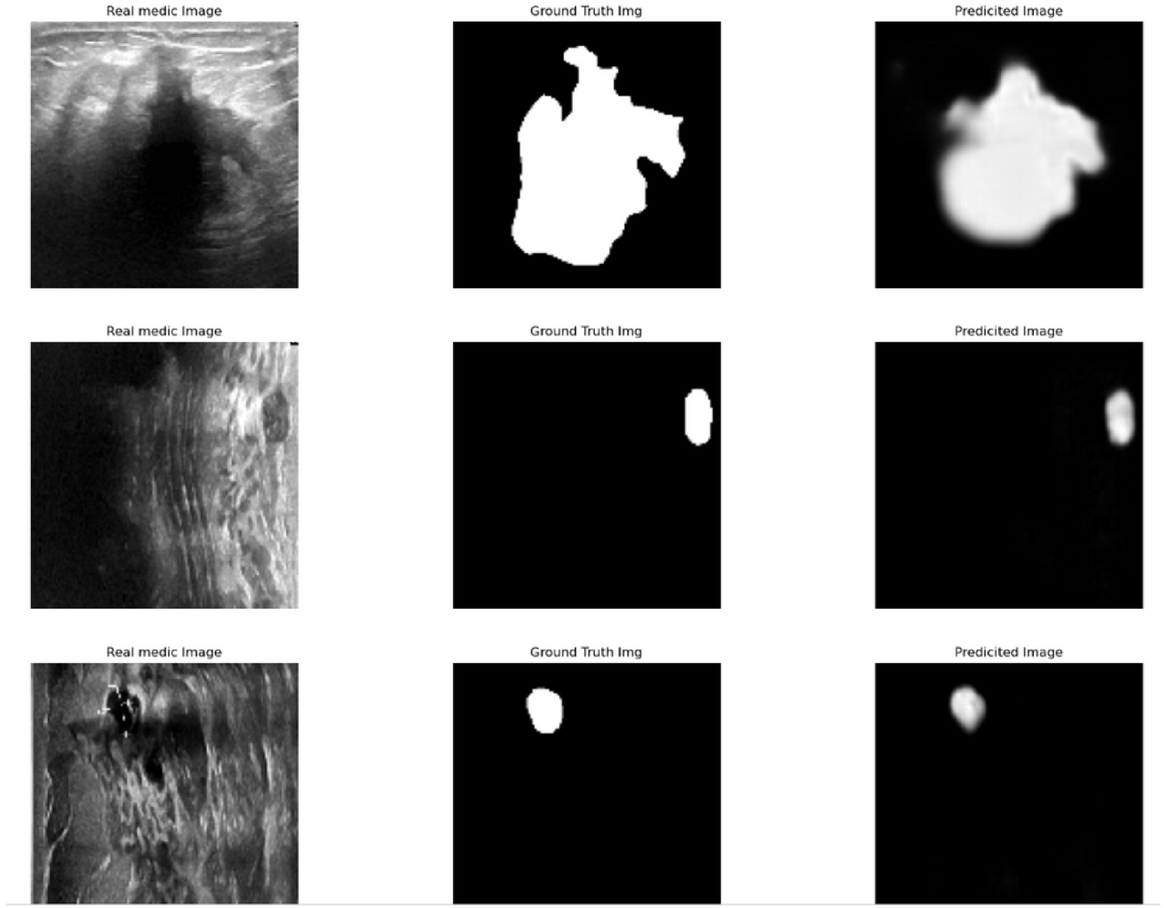
Segmentation model

- I have used the U-net architecture to segment breast cancer images. The model summary is shown below:

```
Model: "U-net"
_________________________________________________________________________________________________
Layer (type)                    Output Shape         Param #     Connected to
=================================================================================================
input_7 (InputLayer)            [(None, 128, 128, 1) 0
_________________________________________________________________________________________________
conv2d_75 (Conv2D)              (None, 128, 128, 64) 640         input_7[0][0]
_________________________________________________________________________________________________
conv2d_76 (Conv2D)              (None, 128, 128, 64) 36928       conv2d_75[0][0]
_________________________________________________________________________________________________
max_pooling2d_24 (MaxPooling2D) (None, 64, 64, 64)   0           conv2d_76[0][0]
_________________________________________________________________________________________________
dropout_21 (Dropout)            (None, 64, 64, 64)   0           max_pooling2d_24[0][0]
_________________________________________________________________________________________________
conv2d_77 (Conv2D)              (None, 64, 64, 128)  73856       dropout_21[0][0]
_________________________________________________________________________________________________
conv2d_78 (Conv2D)              (None, 64, 64, 128)  147584      conv2d_77[0][0]
_________________________________________________________________________________________________
max_pooling2d_25 (MaxPooling2D) (None, 32, 32, 128)  0           conv2d_78[0][0]
_________________________________________________________________________________________________
dropout_22 (Dropout)            (None, 32, 32, 128)  0           max_pooling2d_25[0][0]
_________________________________________________________________________________________________
conv2d_79 (Conv2D)              (None, 32, 32, 256)  295168      dropout_22[0][0]
_________________________________________________________________________________________________
conv2d_80 (Conv2D)              (None, 32, 32, 256)  590080      conv2d_79[0][0]
...
Trainable params: 31,030,593
Non-trainable params: 0
```

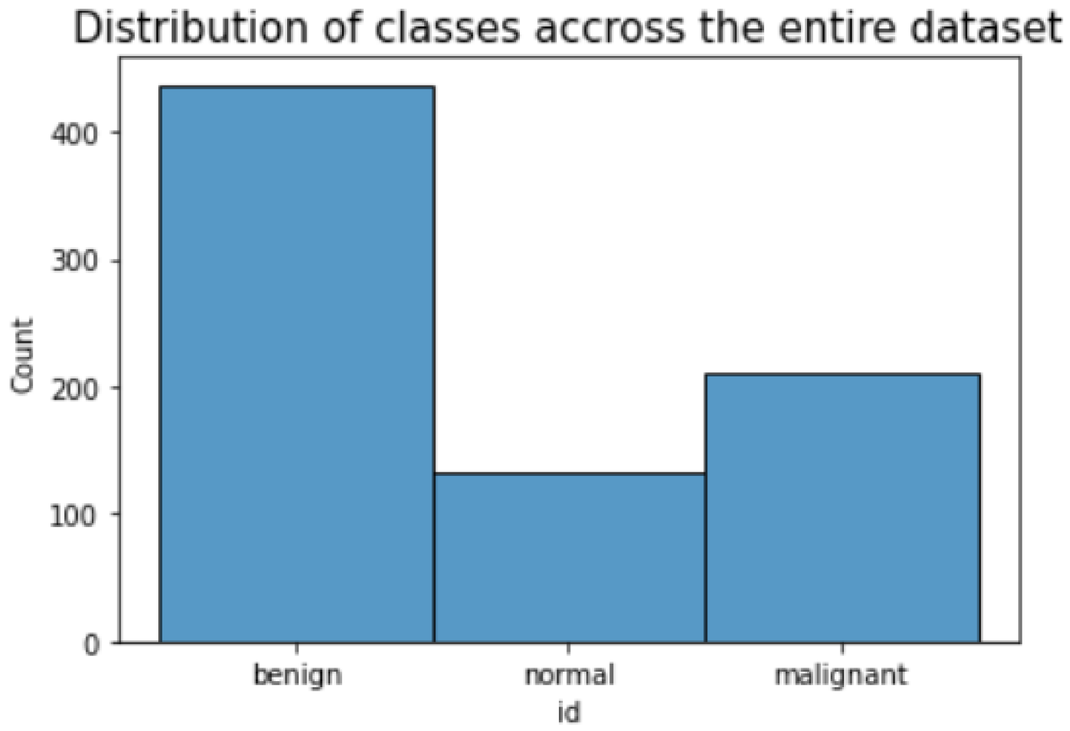
Segmentation Model Results

- The accuracy score for my segmentation model on test data is around 97.69%. The results are shown below:



Frequency distribution

- Before we start training our model, let's see the frequency distribution for each of the classes.
- Here, the number of instances for benign is significantly more as compared to that of normal and malignant.
- Hence we need to remove these extra instances in order to balance out the dataset.



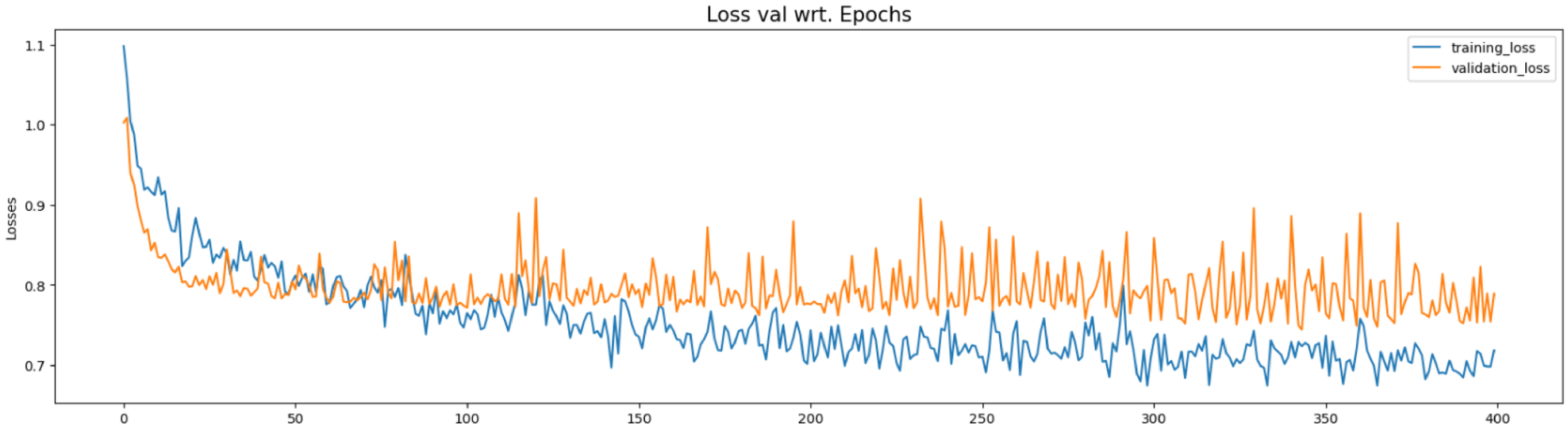
Classification model

- I trained various models like ResNet50, ResNet101, and DenseNet, out of which ResNet101 gave the highest accuracy. The model summary for ResNet101 is shown below:

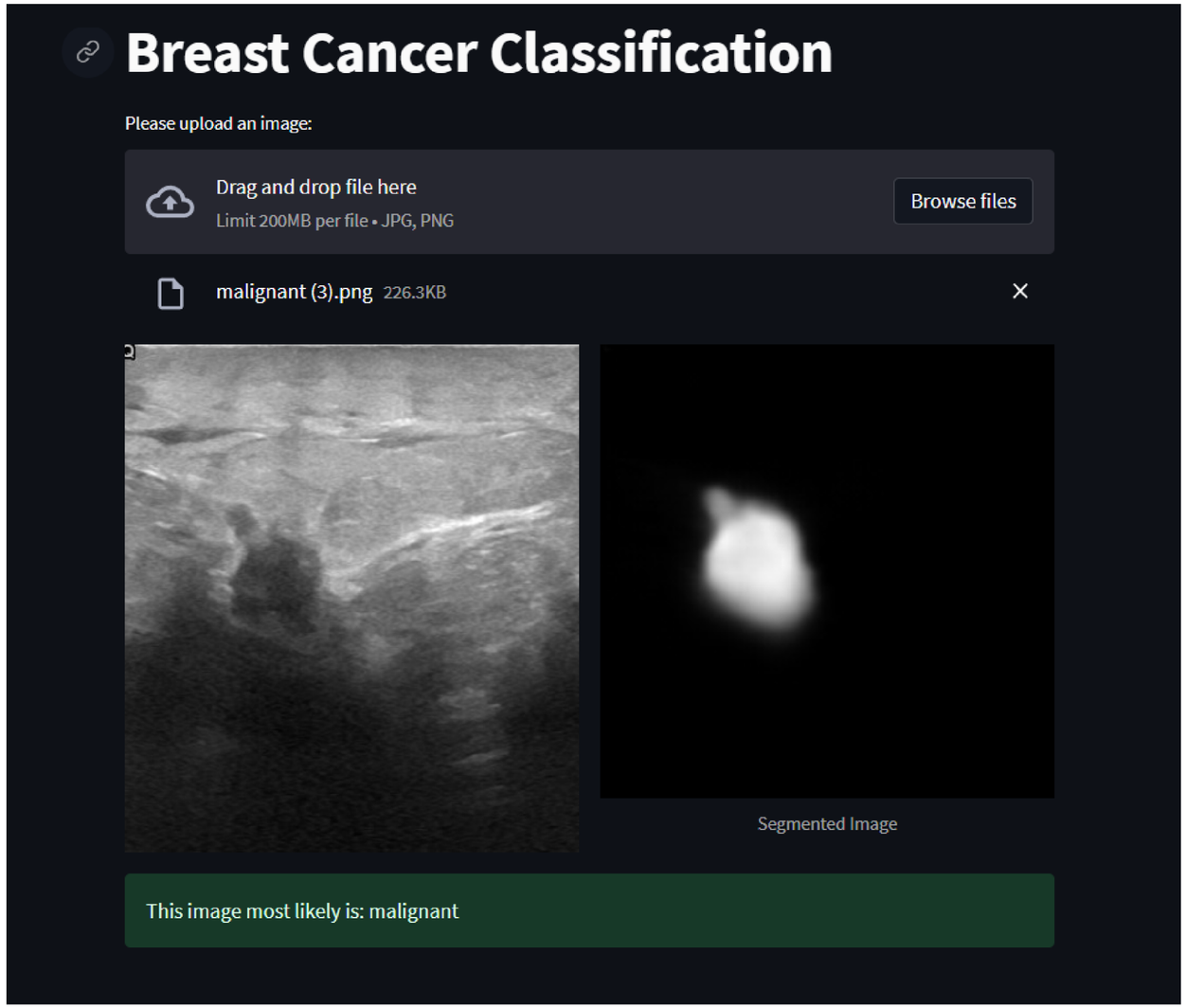
```
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 resnet101 (Functional)      (None, 2048)              42658176

 flatten_2 (Flatten)         (None, 2048)              0

 dense_4 (Dense)             (None, 512)               1049088

 dense_5 (Dense)             (None, 3)                 1539

=================================================================
Total params: 43,708,803
Trainable params: 1,050,627
Non-trainable params: 42,658,176
_________________________________________________________________
```

ResNet101 model results

I trained ResNet101 on images of breast cancer using transfer learning for 400 epochs and a batch size of 64. The accuracy I got was around 68% on the test data. The graph of loss v/s epochs is shown below:



Deployment using streamlit



Thank You!