Segmentation and Classification of Breast Cancer In Breast Ultrasound Images Using Deep Learning

Project's Domain Background

The domain of this project is the application of AI an ML in healthcare specifically in medical imaging. Artificial intelligence (AI) has powerful potential within healthcare, promising the ability to analyze vast amounts of data quickly and in detail. AI technologies are making great strides in medical imaging. Studies have shown that the use of AI may be able to enable earlier disease detection.

AI and machine learning (ML) can be used by radiologists, pathologists, and oncologists for diagnosis, segmentation, classification, and other medical applications Minute changes that are often missed by the human eye can be detected with ML algorithms.

Problem Statement

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

In 2020, there were 2.3 million women diagnosed with breast cancer and 685 000 deaths globally. As of the end of 2020, there were 7.8 million women alive who were diagnosed with breast cancer in the past 5 years, making it the world's most prevalent cancer. There are more lost disability-adjusted life years (DALYs) by women to breast cancer globally than any other type of cancer. Breast cancer occurs in every country of the world in women at any age after puberty but with increasing rates in later life. WHO

Breast cancer treatment can be highly effective, especially when the disease is identified early. to achieve that goal I will be utilizing the power of ML algorithm to segment and classify the types of lesion present in Breast Ultrasound Images Dataset.

Dataset and Inputs

we will be using the data set of Breast Ultrasound Images which was collected in 2018. This Dataset can be found at Kaggle

which consists of 780 images with an average image size of 500*500 pixels. The images are in PNG format. The ground truth images are presented with original images. The images are divided into 3 categories (normal, benign, and malignant). This images were taken from 600 female patients ages between 25 and 75 years old.

Solution Statement

The proposed solution is divided into two stages:

1. image segmentation (to locate the location of the tumour or lesion)
2. image classification (to classify the image into one of the three classers mentioned above)
   These two stages will be implemented using AWS, starting from uploading data to s3 followed by

EDA and preprocessing jobs and going through training and hyper-parameters tuning jobs and finally debugging and deployment of the model.

Benchmark Model

1. for image segmentation task:
   I will compare my work with the results of this paper STAN: SMALL TUMOR-AWARE NETWORK FOR BREAST ULTRASOUND IMAGE SEGMENTATION
   specifically, I will be comparing the following metrics:
   1. True Positive Rate (TPR)
   2. False Positive Rate (FPR)
   3. Jaccard index (JI)

2. for image classification task:
   I will be using the result of this paper to compare my work to it. Breast Cancer Classification in Ultrasound Images using Transfer Learning
   specifically, I will be comparing the following metrics:
   1. Precision
   2. Recall
   3. F1 score

Evaluation Metrics

the solution will be evaluated using the following metrics:

1. image segmentation metrics:

Metric	Brief explanation
True Positive Rate (TPR)	measure the percentage of actual positives which are correctly identified.
False Positive Rate (FPR)	measures the percentage of false positives against all positive predictions
Jaccard index (JI)	used for gauging the similarity and diversity of sample sets.
Dice similarity coefficient (DSC)	statistical tool which measures the similarity between two sets of data.
absolute mean error (AME)	mean absolute distance between the pixel in one image and its corresponding pixel in another image

2. image classification metrics:

Metric	Brief explanation
Precision	how precise the model is, i.e. it is a measure of the quality of the predictions we make
Recall	proportion of actual positives was identified correctly
F1 score	harmonic mean of precision and recall

Project Design

The following flow chart depicts the project design.