Sanjana Kuchipudi
DS 200
Course Project

Breast Cancer Segmentation Project

Breast cancer is one of the most common causes of death in women worldwide and it is the number one diagnosed cancer in women in the US and many countries around the world. As mentioned in the Carol Milgard Breast Center if breast cancer was detected at an early stage, they would have a 93% survival rate from the cancer. Therefore, early and accurate detection of breast cancer is critical in saving many lives in the US and all around the world. Almost everyones knows someone who has been affected by breast cancer, including me. This is why this topic means so much to me.

A study was conducted in Cairo and they collected breast ultrasounds from women from age 25 to 75 years old. It is the biggest dataset of ultrasound images which can then be used for machine learning. The goal of this project is to take the ultrasound images and create a model to classify the ultrasound images to see if it is normal, benign or malignant. A normal ultrasound is a healthy breast with nothing abnormal, a benign tumor is an unusual but not cancerous growth and a malignant tumor is a cancerous tumor. The main reason for this work is to save many, many people from breast cancer through early detection and classification. It will also be comprehensive enough to contain different breast cancer stages. The result should be a machine learning model that is able to take an image of a breast ultrasound and then classify and detect breast cancer. The medical images are used for deep training the model which will make the results accurate. A note to add is that the researchers took into consideration the ethical concerns of privacy so all the images of patient ultrasounds are anonymous to protect their identity.

In the preprocessing step, the duplicate images are removed and the errors with the classification were fixed. The type of image was changed from DICOM to PNG to make the process of exchanging images and using it for machine learning easier. The unnecessary details outside the ultrasound are removed and the images are then verified and checked by specialists to make sure the classification is all correct. The classification of the ultrasound is vital when running the images through the machine learning algorithm.

U-net is a neutral network which is used for image segmentation for biomedical sciences. A part of that Conv2d is used for high accuracy in image recognition and kernels are used in image processing. In this case it is used for edge detection. U-net is a neutral network used for image segmentation for biomedical sciences. The goal is to be given an image and then predict a segmentation mask that highlights the breast growth (cancerous or not).