Currently there are no major surprises from the domain of the dataset. The dataset contains many images of normal breast tissue and breast tissue that show signs of IDC. There are so many images that I did decide to take a sample size as it would take such a long time to process them all. If I can get everything working, I may add some more in the final version as to (hopefully) bump up the accuracy scores. Kaggle is a great place to go for datasets as it has many options and varieties.

The dataset is what I thought it would be, as it contains so many images of breast tissue. What was somewhat surprising to me is that there are some abnormal breast tissue photos that do not show a lot of IDC. It will be interesting to see how much this is picked up in the final results. I am hoping that the accuracy is at least 80%. As mentioned earlier, there are so many photos that I am taking a sample of them to begin with. As I gain more confidence and know more of my approach by flushing it more out, I am hoping to add more.

So far, I have not had to adjust my initial questions as I still believe they are relevant no matter what. I am hoping to still achieve a high accuracy and would like to see if the model will be able to accurately predict if the image contains cancerous cells. From my initial testing, I did have too many “normal” images and not enough that contain IDC when training. I had to add more to give more of the cancerous images so that it is able to use the prediction better.

I have not yet gotten to the real method yet, but so far, the training of the model has been going well. I am hoping that when I get more into the testing that it will go according to plan and will run smoothly without many challenges. I am also going to attempt to do some data augmentation to see if the predictions will work if they are slightly altered. I do plan looking more into the training and validation accuracy to see if the model worked out how it was intended.

Current challenges that I am facing is that there are so many images. If I were to use them all, it would really bog down my computer and it would take more time than I would like for it to run. This is why I chose to work with not all of the images, but if I can add more than I will. Other challenges that I am facing is trying to find time to really focus. At work, we are doing an upgrade and need to make sure all of that is completed by the end of the week. In my personal life, I am in the process of moving. Once the move is more settled and this project is rolled out, I am sure that more focus time can come.