**Abstract.** In this paper, we will be detecting and classifying breast cancer calcification and masses in a single step. We tackle this problem by training a Convolution Neural Network (CNN) to classify different cropped images of the types of breast cancer. We are then building a sliding window detector to break apart full mammogram images into sub-images, classifying the sub-images. We receive multiple probabilities for each sub-image for the different classification. We then rank the sub-images, displaying the coordinates of the highest ranked sub-images for each classification. The results of this process are we detect 32% of cancer within the mammograms, and properly classify 68%.

**1 Introduction**

Breast cancer is a dreadful disease, it is the second leading cause of death among women. The chance of a women dying of breast cancer is 1 in 37 or 2.7 percent [4]. 40 thousand women are estimated to die from breast cancer in 2017 alone [4]. The affect breast cancer has upon those around us it is important we understand it, can diagnose it easily, and have a treatment path for the disease. Awareness and early screening are very important in reducing the risk of breast cancer. Early detection of cancer can save the patient’s life. If the cancer is not detected on time and treated it can lead to death. There are three main types of tools used to diagnose breast cancer. The first is a physical breast exam. This is completed by a physician of as a self-exam. The second way is with imaging. Imaging is sub-divided into a mammogram, ultrasound and MRI scan [4]. Each one of the imaging tests has a different purpose. Mammograms are used to detect lumps within the tissue. Ultrasounds determine the difference between a solid mass and a cyst. The MRI is used to determine how invasive the cancer has become. The final test is the biopsy. In a biopsy they remove some tissue and test it in the lab, identifying if the cells are cancerous and the type of cancer. Each step taken to aid in the proper detection of breast cancer can make a vast improvement in the lives of women and those around them.

We are looking to detect and classify breast cancer within a mammogram in one step using a CNN. The biggest issues to overcome is the ambiguity of the data itself, cancer can be obscured by the nature of the tissue around it. Tis can be resolved with different views of the breast tissue, the most common are CC, overhead, and MLO, at a 45degree angle. Also, the differences between malignant and benign cells cannot be easily distinguished on a mammogram. This is displayed by doctor's having three pathologies for possible cancer found on a mammogram, benign without callback, benign, and malignant. Benign indicates that the doctor could not determine the pathology of the cancer by reading the mammogram alone and further tests. With 38% of our training data is marked benign we can see that even medical providers have a difficulty classifying breast cancer with a mammogram alone. The final issue we encounter is our ignorance on the subject itself. When our detector picks a section of the mammogram as important and it looks like the text book answer to the classification picked, we cannot explain why it is not cancerous and instead it is a false positive.

The first step we utilized to solve this problem is to build an image classifier to differentiate between benign and malignant, calcifications and masses found within mammograms, by classifying cropped images of the different pathologies of breast cancer. We utilize this method because the nature of our data and limit of our computing power. Our dataset contains full mammograms and cropped images of the different pathology. By using the cropped images, we have to use less compression on the images to load them in our memory. We have a finite amount of memory, 8Gbs of main memory and 3Gbs of video memory to use for training and testing our CNN.

Finally, we use the model built by the image classifier to build a sliding window image detector. We pull sections of the of the full mammogram into our model one piece at a time and run them through the classifier, predicting the possibility of the sub-image belonging to a different class of cancer. We record the different sub-images probabilities and coordinates. Finally, we display the highest ranked of each sub-category of cancer.

If the known area of cancerous tissue is detected and classified properly, and also, displayed in the highest ranked of its subcategory, our system works and there are no false positives. If the known are of cancerous tissue is detected and classifies properly, but not displayed in the ranked area we have a false positive elsewhere within the mammogram. If we detect the tissue but misclassify the tissue, we know the detector is working by the pathology is obscure and more analysis is needed.

Utilizing them detector and the method above we detected 32% of the cancer within the mammograms and proper classified 68% of the cancer we discovered. Only 50% of the cancer detected, or 15% overall, was detected as the most important anomaly on the mammogram. This leaves a large group of false positives, with most of the false positives being outside of the breast tissue completely or within the black outside of the mammogram.

After completing this process of building the classifier to feed into the sliding window detector for the purpose of detecting and classifying breast cancer within a mammogram we can conclude this is a solvable problem but will need a serious effort to bring to a solid solution. Many studies [studies], with good results, have focused on a part of this solution. But to bring it to a single step will be hard work. When we divide our model up to two different focuses, calcification and masses, we raise our result into the 80% range without tweaking our model. To have everything done in one step might be ready for real work use, but it is a step needed to be taken.