SMU Data Science Review

Volume 2 | Number 1 Article 10

2019

The Simultaneous Detection and Classification of Mass and Calcification Leading to Breast Cancer in Mammograms

Scott Gozdzialski
Southern Methodist University, sgozdzialski@smu.edu

Alex Stern
Southern Methodist University, kastern@smu.edu

Ireti Fasere
Southern Methodist University, ifasere@smu.edu

Daniel W. Engels
Southern Methodist University, dwe@smu.edu

The Simultaneous Detection and Classification of Mass and Calcification Leading to Breast Cancer in Mammograms

Scott Gozdzialski, Ireti Fasere, Alex Stern, Daniel W. Engels
Master of Science in Data Science, Southern Methodist University,
Dallas, TX 75275 USA
{sgozkzialski, ifasere, kastern, dwe}@smu.edu

Abstract. In this paper, we present a novel method for detecting and classifying breast cancer calcification and masses in a single step. The detection and classification steps of calcifications and masses identifiable with a mammogram image are typically performed independently even though their simultaneous solution may lead to a more efficient approach. Our novel method utilizes a Convolutional Neural Network (CNN) to classify the calcifications and masses of different cropped images of a mammogram. We utilize a sliding window detector to break apart full mammogram images into sub-images, and identify and classify the observable objects in the sub-images. We receive multiple probabilities for each sub-image for the different possible classifications. We rank the sub-images, displaying the coordinates of the highest ranked sub-images for each classification. The results of this process are that we detect 46% of cancer within the mammograms and properly classify 64% of the calcifications and masses identified.

1 Introduction

Breast cancer is the second leading cause of death among women. The chance of a woman dying of breast cancer is 1 in 37 or 2.7% [1]. Forty thousand women are estimated to have died from breast cancer in 2017 alone [1]. It is important that we understand the effect breast cancer has upon those around us, that we can diagnose it easily, and that we have a treatment path for the disease. Awareness and early screening are crucial 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 or as a self-exam. The second way is with imaging. Imaging is sub-divided into a mammogram, ultrasound and MRI scan [1]. 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 present a method to detect and classify breast cancer within a mammogram in one step using a Convolution Neural Network (CNN). The biggest issue to overcome is the ambiguity of the data itself since cancer can be obscured by the natural variation of the tissue around it. This obscuring can be resolved with different views of the breast tissue, the most common are bilateral craniocaudal (CC), overhead, and mediolateral oblique (MLO), at a 45-degree angle. Also, the differences between malignant and benign cells cannot be easily distinguished on a mammogram. This is displayed by the fact that doctors have 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 must 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 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 the detector and the method above, we detected 32% of the cancer within the mammograms and properly 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, with good results, have focused on a part of this solution. But to bring it to a single step will be hard work.

2.0 Background

2.1 Mammogram

Mammograms is an early screening tool used in detecting early breast cancer in women with no indication of a condition or disease. It can also be used in detecting and diagnosing breast cancer in women with a known condition such as a lump, skin dimpling and nipple discharge. The mammography exam is referred to as the mammogram. Mammography can be divided into two types the screening mammography and the diagnostic mammography. The screening mammography shows early detection of breast cancers and show changes in the breast for up to two years before the patient or doctor can feel it. The U.S. Department of Health and Human Services (HHS) and the American College of Radiology (ACR) as recommended screening for women at age 40 annually [1]. "These annual mammograms save the lives of many women through the early detection of breast cancers when they are most curable. The National Cancer Institute (NCI) stated that women with cancer, and those with family history of breast cancer should seek medical advice on whether to start early screening before the age of 40 [1]. Women who are at risk of breast cancer may need to obtain a breast Magnetic Resonance Imaging (MRI) in addition to their annual mammogram. The Diagnostic Mammography is used to examine women with abnormal findings like breast lump or nipple discharge. It is done after an abnormal screening mammogram to review the area of concern on the mammogram image.

Mammography is a medical technique that is used in detecting and diagnosing early cancer by taking images of the breast tissue using a low-dose x-ray [1]. The low- dose x-ray is referred to as the radiograph. The radiograph is non- invasive, and it aids physician in detecting and treating early stage of breast cancer in women. Women ages 40 to 44 have the choice to begin a mammogram annually, ages 45 to 54 are required to take a mammogram every year, ages 50 and above are required to take the mammogram every two years [2]. Mammography has evolved over the years, and the three recent mammography are the digital mammography, computer-aided detection and the breast tomosynthesis [3]. The digital mammography also referred to as the Full-Field Digital Mammography (FFDM). The FFDM system can be compared to a digital camera which enables it to take better images of breast tissue using a lower radiation dose. The images of the breast tissue are loaded into the computer to be stored and reviewed by the radiologist. Computer-aided detection also referred to as the CAD, search the breast tissue for abnormal areas of density, mass, and calcification that may show the presence of cancer. These areas are further reviewed by the radiologist for better assessment. The breast tomosynthesis is also referred to as the three-dimensional (3-D) mammography and the digital breast tomosynthesis (DBT), it is a more advanced system of taking multiple images of the breast tissue from different angles in 3- dimensional image. Studies have shown that the DBT screening as a better and improved detection of breast cancer and a lesser rate of patient call-backs due to potential abnormal findings [3]. The DBT gives a better accuracy in showing the shape, size, and the location of breast abnormalities. It

shows a clearer image of the abnormalities within the dense breast tissue and detecting multiple breast tumors.

The American Cancer Society (ACS) recommends that patients' going in for their mammogram should not wear deodorant and lotion as these can show on the mammogram as calcium spots [3][1]. Mammograms should not be scheduled when pregnant because a possibility exists for the baby to develop cancer due to excessive exposure to radiation. The screening should not be scheduled a week before the menstrual period as the breasts is tender during this period. The best time to schedule a mammogram should be a week after the menstrual period.

2.2 Normal and Abnormal Mammogram

2.21 Normal Mammogram

What does a normal mammogram look like? A mammogram will mostly be gray with white areas showing healthy tissue. The small white specks image on the mammogram shows no health problem. The denser tissue, the connective tissue, and the glands show white on the mammogram [4]. A whiter speck on the mammogram usually does not indicate a health issue. However, not all normal mammograms are cancer free. About 20 percent of breast cancers are not detected on a mammogram [4]. This percentage is apparent for women with denser breast tissue. "Some cancer is overlooked on a mammogram, especially cancer that does not result in a mass. MRI is recommended to get a deeper detail of an area of concern on mammogram and for women that are at high risk with denser breast tissue. If a mammogram shows an area of concern, the MRI shows the problem. However, MRI are more expensive than a mammogram and are mostly used to diagnose abnormal tissue or high-risk developing breast cancer.

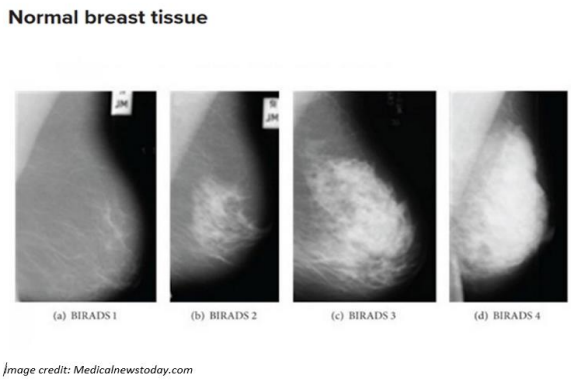


Fig. 1. Normal breast tissue showing different categories of BIRADS.

2.22 Abnormal Mammogram

Women with a dense breast are at higher risk of breast cancer. The denser the breast the more difficult it is to find abnormalities on a mammogram. If a mammogram results shows abnormalities, a follow up is necessary to check if it is breast cancer. An abnormal mammogram most often does not indicate breast cancer. Most abnormal mammogram is a benign breast condition or just a normal breast tissue. Some abnormalities found in the breast include cysts, calcifications, fibroadenomas, and scar tissue [4]. A cyst is a small filled sac, most often is not cancerous but needs further check-up to determine if it is cancerous or not. Calcifications is a deposit of calcium, often due to old age. Depending on the appearance further checkup is needed to test for cancer. Fibroadenomas is a round-like benign tumors in the breast and can occur at any age. Scar tissue appears white on the mammogram, your doctor should be aware of any scar on your breast at the earliest.

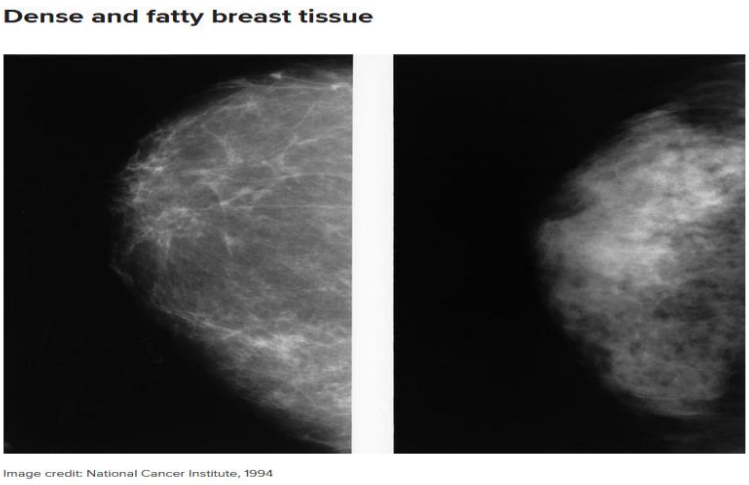


Fig. 2. Dense and Fatty breast tissue becomes less dense as the age progresses.

Cancerous tumor

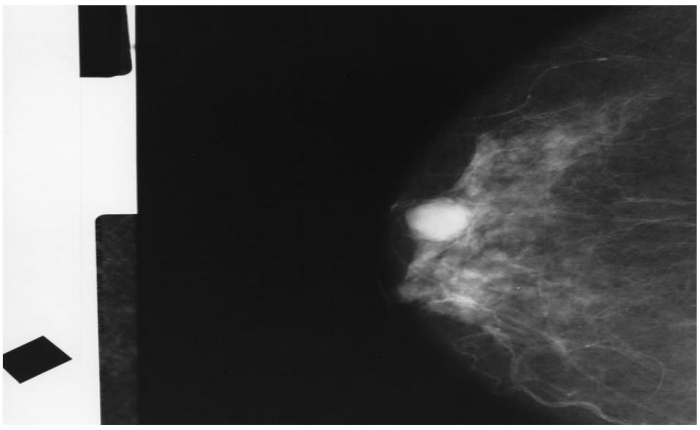
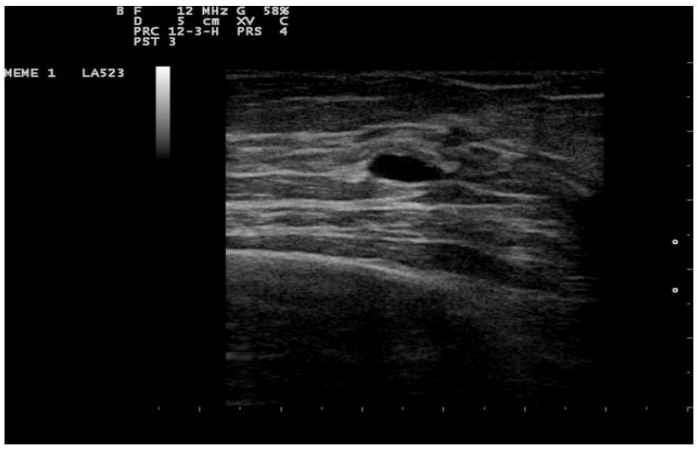


Image credit: National Cancer Institute, 1990

Fig. 3. Breast tissue showing a cancerous tumor.

Breast cyst



In a mammogram, cysts show up as dense white masses. A doctor may recommend a breast ultrasound, like the one above, to help determine if the cyst is benign and fluid-filled. They may also take a sample of the fluid for testing.
Image credit: Nevit Dilmen, 2011

Fig. 4. A breast cyst appears as dense white masses, further check-up is needed to determine if it is benign and fluid-filled.

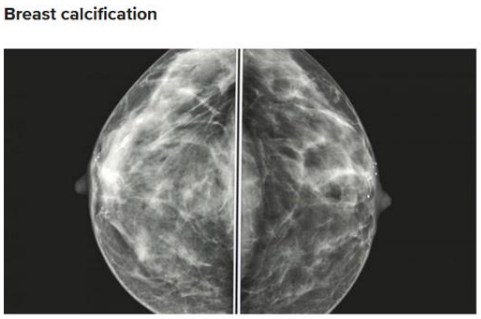


Image credit: Medicalnewstoday.com

Fig. 5. Breast calcification often due to old age showing deposit of calcium.

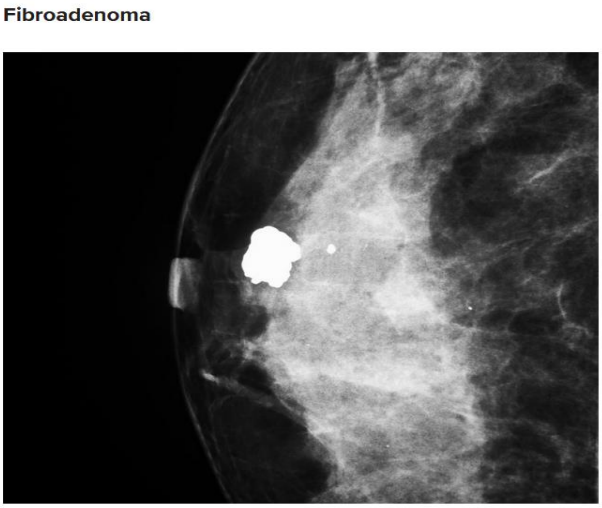


Image credit: SCiardullo, 2014

Fig. 6. Fibroadenoma appears as a a round-like benign tumors in the breast and can occur at any age.

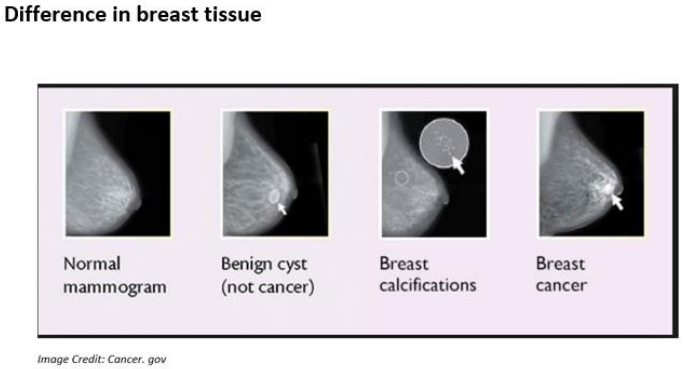


Fig. 7. The difference in breast tissue showing a normal breast tissue, benign cyst, calcifications, and breast cancer image.

2.3 Interpreting the Mammogram

The standard system for reporting a mammogram result is called the Breast Imaging-Reporting and Data System (BI-RADS). The BI-RADS category ranges from 0 to 6. The doctors across the United States adhere strictly to the BI-RADS system. Table 1 shows how to interpret the BI-RADS result.

Table 1. Interpreting a Mammogram

Category	Interpretation
0	Result unclear needs further test, and previous mammogram comparison
1	Abnormalities absent
2	No cancer but presence of abnormalities like benign calcifications
3	Abnormalities like benign present needs follow-up
4	Abnormalities present could be cancerous, may need biopsy
5	Abnormalities present likely to be cancerous, requires biopsy
6	Presence of cancer requires a mammogram for progress.

The result of a mammogram is clearly explained by a medical practitioner to a patient after the result comes in. Further check-up is then required if there are abnormalities on the mammogram.

Mammograms are the most important method out there for detecting breast cancer and checking the response to treatment of breast cancer. However, mammograms are not perfect for checking abnormalities, especially women with dense breasts. There is no standard for abnormal and normal mammogram out there. Mammograms appear

different for every patient. White areas on the mammograms may need follow-up, but most often do not result to cancer.

2.4 Benefits and Risks of a Mammogram

Table 2 summarizes the primary benefits and risks of mammograms.

Table 2. Benefits and Risks of a Mammogram

Benefits	Risks
Early detection of breast cancer like small tumors gives the patient more treatment options.	There is possibility of small exposure to radiation after mammogram, but the benefits outweigh the risk.
Mammogram is used to detect all types of breast cancer such as ductal and invasive lobular cancer.	About 5 to 15 percent of screening mammograms needs retesting. Most cases turn out to be false-positive mammogram. This means the result is a normal mammogram.
Special care is taken during x-ray exam to use the lowest dose of radiation. No radiation is left behind in the patient's body after the x-ray exam.	It is not advisable to take a screening mammogram if you are pregnant. The baby stands the risk of possible exposure to radiation.

2.5 Neural Networks

Neural networks have a long history in the machine learning realm. In 1958, Frank Rosenblatt theorized the perceptron to mathematically model a neuron in the human brain [1], Figure 2.2.1. In his theory, the neuron takes multiple inputs and multiplies the input by a weighted value. The neuron then sums all the inputs and weights, and if the sum is above a threshold, the neuron is activated. If the sum is below the threshold, the neuron stays dormant.

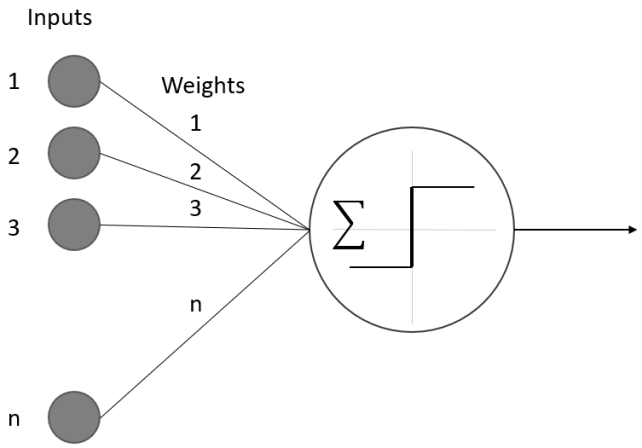


Fig. 8. Perceptron

Over time, the perceptron was shown to be able to handle some simple functions, but more complex problems need more than one perceptron to calculate the solution. The next step was to build networks of perceptrons to work in conjunction, called neural networks. The first neural networks were made up of only a single input layer and a single output layer, this was caused by the need for multiple outputs for a simple problem [2]. An example of the would be if you wanted to classify a picture of a vehicle as a truck, a car, or motorcycle; then, you could have multiple inputs of the images, but a single perceptron could only output one response. For this example, we would need three different outputs, one for truck, one for car, and one for motorcycle.

Neural networks continued to become more and more complex as the problems increased in difficulty. We learned that having a single layer of neurons is needed to have hidden layers, Figure 2.2.2. The hidden layers are fully connected to the neurons in the previous layer and the next layer, but they do not have an input or output connection outside of the neural network [2].

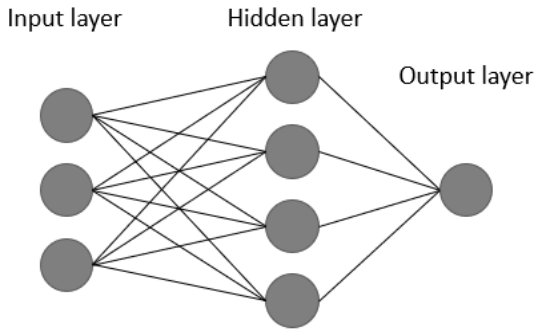


Fig.9. Hidden Layer

As the layers of neural networks grew to include hidden layers, the way we trained the models had to change. Since the output did not have a direct connection to the output layer, we had to find a way to push the corrections to the weights of the connections of layers further back. The way we achieve this is with backpropagation, Figure 2.2.3. As the network is trained, it outputs the results, then the results are calculated for an error. From the error, we build a correction to the weights of the neurons. We backpropagate those corrections through the network adjusting the weights until the loss is minimized within the loss function, giving us the best results [2].

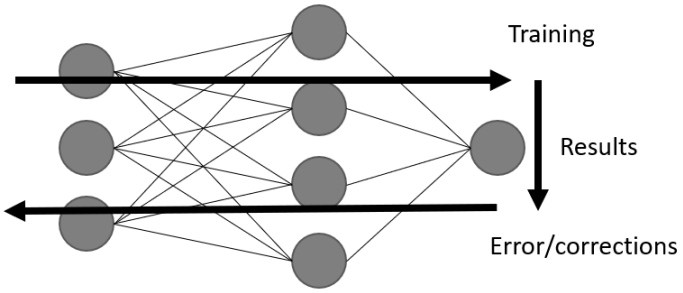


Fig. 10. Backpropagation

2.5.1 Convolutional Neural Networks

The convolutional neural network (CNN) is a neural network used to identify images. It does this on a high leave by examining the pixel of the image with filters and identifying different features of the image structure. It then compiles the different features extracted and builds a model of the object with in the image. After the model has been properly built an unknown image can be feed into the model and a probability of the object in the image belonging to the different classifications trained in model can be output [3].

A CNN is made up of different layers performing different functions, Figure 2.2.1.1. The first type of layer is the convolution layer itself, which applies a filter to the image extracting the features from the image. The next type of layer is the pooling layer, which reduces the size of the previous layers cutting the computational complexity of the problem. The final type of layer is the fully connected layer which is responsible to transform the outputs from the other layers into the final output [3].

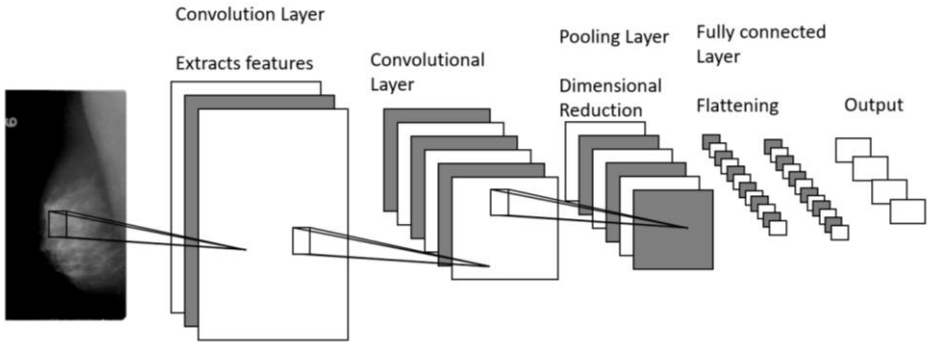


Fig. 11. CNN Example

The first type of layer of a CNN is the Convolutional layer. This layer applies a convolutional filter to the image. The filter is used to learn different features and relationships of the image. It will look for structural or gradient on each different channel of the image. The different channels can be three for each color in the red, green, blue image or one on a greyscale image. These filters take the pixels within the filter size and apply a different linear algebra function to them to extract the feature from that section of the image. They then travel around the entire image, mapping the entire image. Each of the filters produces different maps of the image [3].

With the different filters creating the unique maps on each channel of the image we need to reduce the amount of data we are using. The pooling layer of the CCN is used to reduce the size of the data needed to process. The pooling layer completes this by taking a piece of the filtered image and using a function to combine the properties of the piece, while maintaining the integrity of the information contained within [3].

After multiple convolutional and pooling layers have been applied to the image, we need to send the information to the output layer. Until now, we have not talked about the dimensions of the CNN, but each of the convolutional and pooling steps change the shape of the information. The original image may be 100 x 100 x 3 pixels for the RGB channels, and after the first convolutional layer with 10 filters, it can be 100 x 100 x 3 x 10. Even with pooling it will not be reduced to a perfect shape to feed the output. The flattening layer is responsible for reducing the response from the previous layers into a single stream of bits. This stream of bits is fed into the output layer to get the probability results [3].

Each CNN uses a different number of each layer in different orders to achieve their results.

2.5.2 Object Detection

Utilizing the CNN built above, we can develop a way to detect objects within an image and highlights it location. There are three main ways to do this sliding window, "You Only Look Once, YOLO" and the "Single Shot Multiple Box Detector." The different ways to detect an object have some similarities and a few key differences, but fundamentality they all work the same. They all work by first dividing the image into smaller sub-images. Then, they run the image classification model we described above on the sub-images. A probability is retrieved from the image classifier and recorded with the location within the image. If the probability is above a certain threshold, then that section of the image is highlighted in some fashion [1].

3 Data

The data set we used for our paper was put together by the Cancer Imaging Archive [1] for public use. It was made from the Digital Database for Screening Mammography. The dataset consists of 10,239 images each containing cancer. The image set is made up of full MLO (mediolateral oblique, angled) images, CC (Cranial-Caudal, top-down) images, Region of Interest masks, and cropped images. The images are split into 704 test images and 5324 training images. We pull 100 full mammogram images from the test and training sets before the model is made for validation of the object detector. The images chosen for the object detector are chosen because there is only one image from that patient, removing a chance they could have been looked at before the final validation, skewing the results.

The data is stored in separate folders for each patient and each image type (MLO, CC) and each breast (left, right). These folders have subfolders by the day the image was taken. The next subfolder is used to explain the type of image contained (full mammogram, cropped, ROI). Finally, the images are located within these subfolders named either 000000.dcm or 000001.dcm based on how many images are located within the folder (Figure 12).

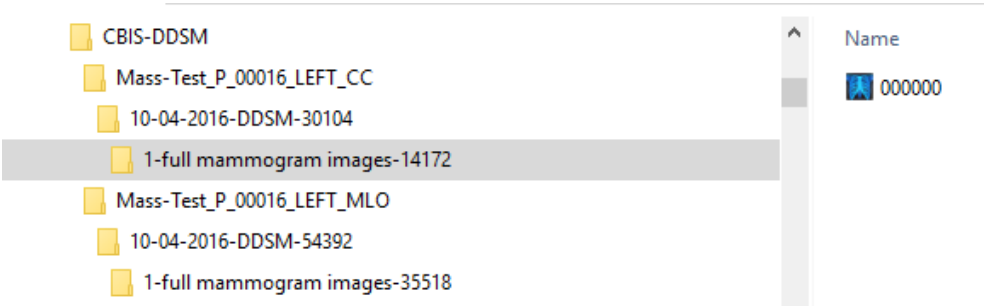


Fig. 12. Folder storing the full mammogram images.

The first step in processing the images we brought them into individual folders used for training and testing of the convolutional neural network. Next, we identify

the individual files used for this study. We converted all the DICOM files into lossless JPEG files to be used by the CNN for detection and classification.

4 Methods

In order to detect and classify breast cancer within mammograms, we utilized CNNs to do the heavy lifting. Using a Keras front end with a Tensorflow working in the back ground, we build a 4-layer deep learning image classifier to differentiate between the four states of the cancer we are looking to detect, calcification benign, calcification malignant, mass benign, and mass malignant. We used the checkpoint from the image classifier to load into the sliding window detector and retrieve the final detection and classification of anomalies in the mammograms.

The building of the image classifier is broken down into four fully connected layers: one flattening layer and three dropout layers using relu activation on the first two and softmax on the final flattening layer. It then uses an adam optimizer looking at categorical response. The four fully connected layers start with 10 filters working up to 80 layer per image, with a shrinking kernel size at the filters increase. After flattening the results, the four drop layers are utilized, bringing the results down to the final four categories. We run 75 epochs validating after each epoch saving the best weights of the epochs to be used by the image classifier. We train our model on only the images with pathologies of "benign-with no call back" and "malignant" to remove the middle ground with an unclear pathology even for the doctors reading the mammograms. We still validate the classifier and test detector with all three pathology types. Removing the "benign" pathology gave an immediate validation accuracy of 11% in classification with no other changes. We achieve a peak validation accuracy of 68% as shown in figure, we load these weights into our object detector.

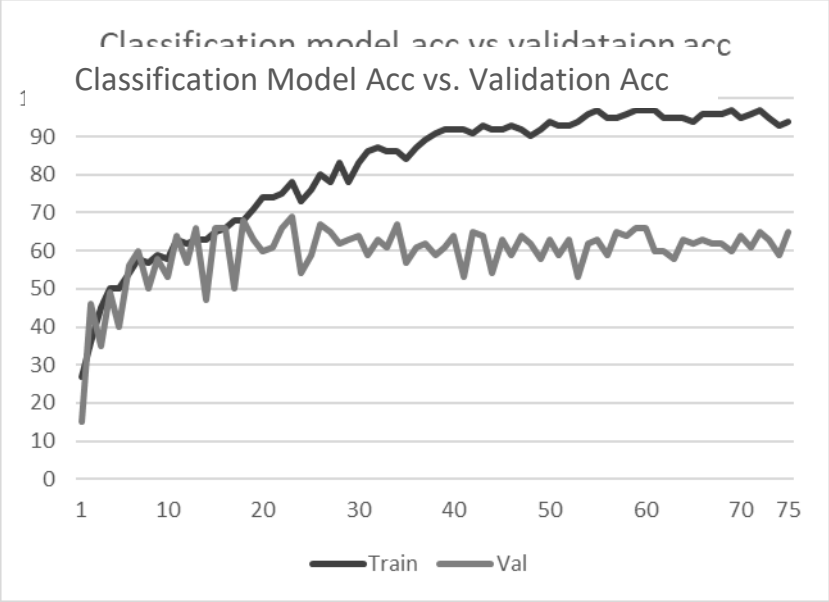


Fig. 13. Classification Model Accuracy vs Validation Accuracy

We then build a sliding window object detector to pull pieces of a full mammogram ranging from 215x512 pixels to 1024x1024 pixels, the images are shrunk to 512x512 sub-images. The different sub-images are run through the image classifier predictor one at time saving the results and x, y coordinates. We then look through the results observing any spikes in the predictions indicating the presence of cancerous tissue. Then the different classifications for the area of interest are compared to see which classification of caner type and pathology are determined by the CNN.

5 Results

We achieved a detection accuracy of 46% and classification percentage of 64%. We utilized 25 full mammograms of each type of cancer and pathology for our final validation of our cancer detector and classifier. Of the different detection rate of cancer type, Mass or Calcification, Mass consisted of 65% of those detected while Calcification makes up the remaining 35%.

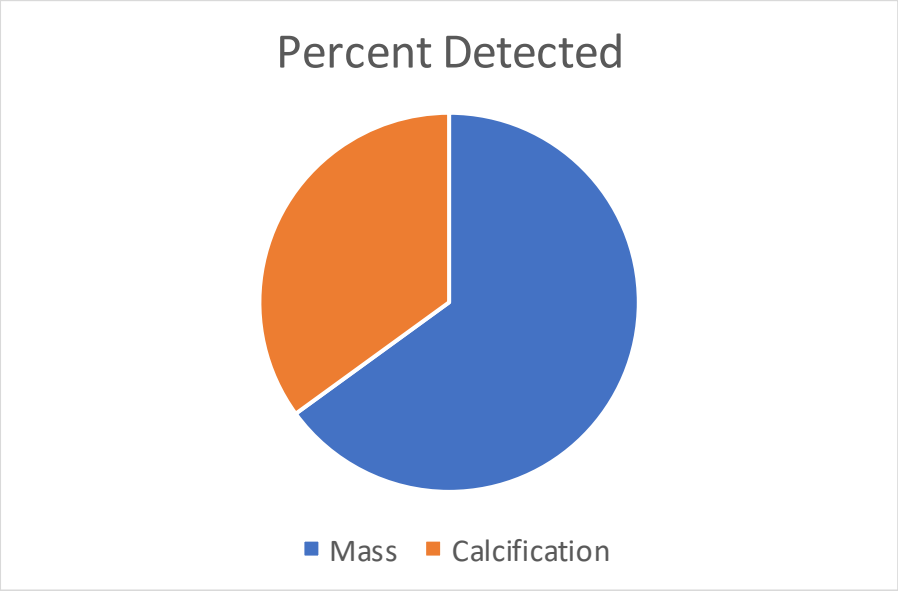


Fig. 14. Mass vs Calcification

Of the pathology, Malignant of Benign, Malignant was 58% of the detected cancer and benign was 42%.

Breaking the results down further into individual combinations of type and pathology we can see our best category for detection is malignant masses, we detected 16 out of 25 or 64% of this type of cancer.

When it comes to classification of the detected cancer, we classified 64% of the detected cancer properly. We had two that we improperly classified as the wrong type of cancer. When it comes to improper classification of the pathology there are some results.

Malignant calcification was classified very well, with only one miss classification. While with masses, both malignant and benign, we properly classified 50%. Finally, benign calcification was miss classified over half the time, but only 7 cases we detected.

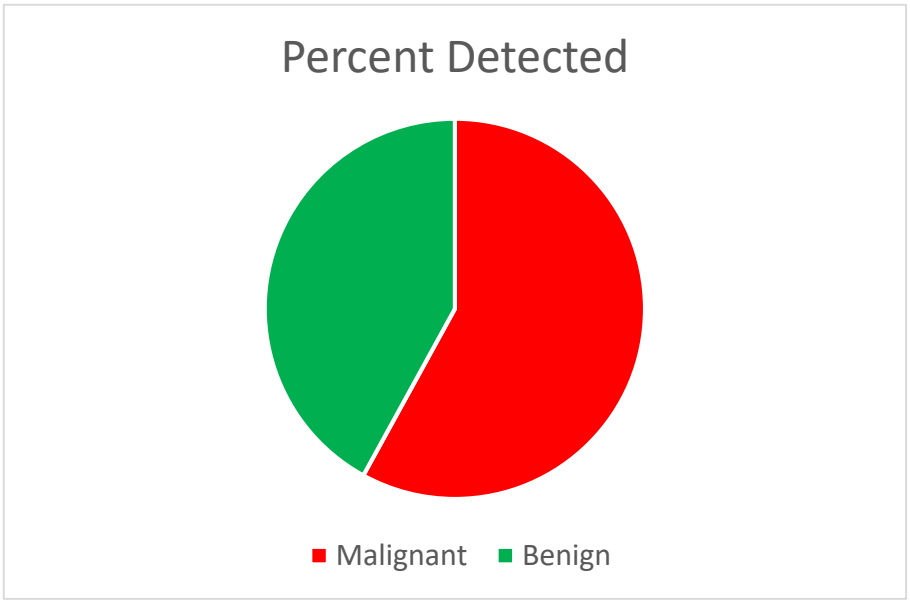


Fig. 15. Malignant vs Benign

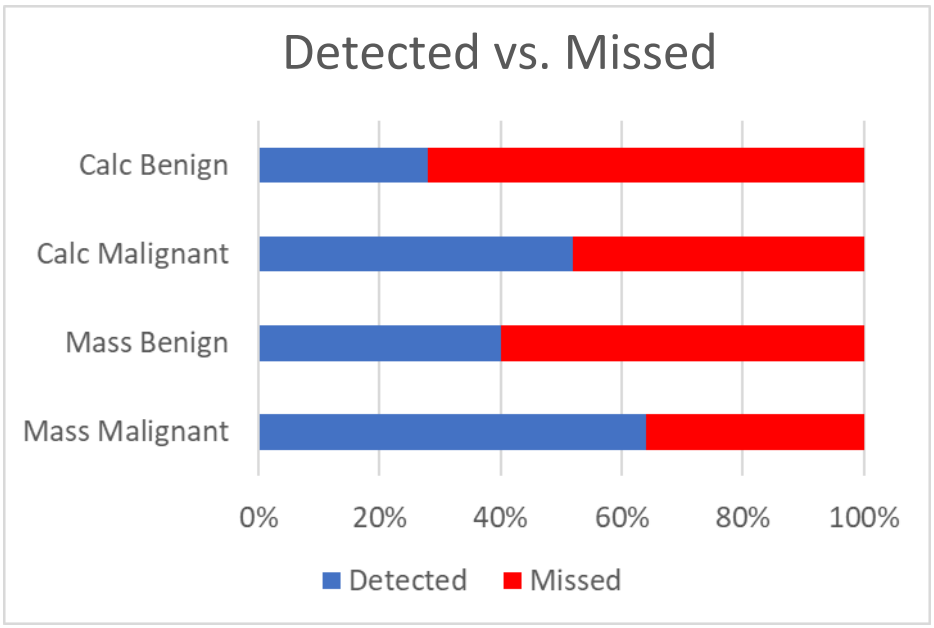


Fig.16. Detected vs Missed

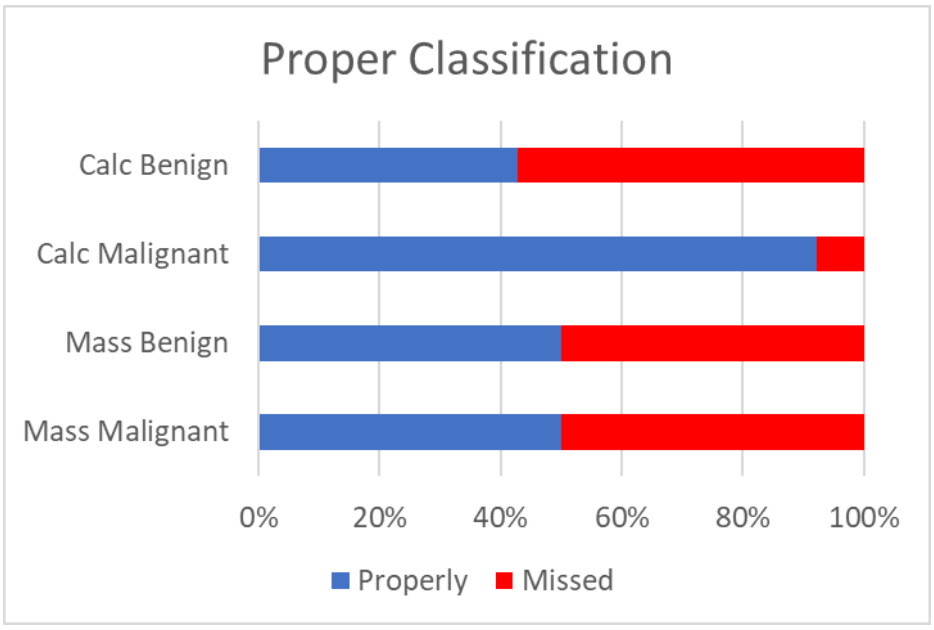


Fig. 17. Proper Classification

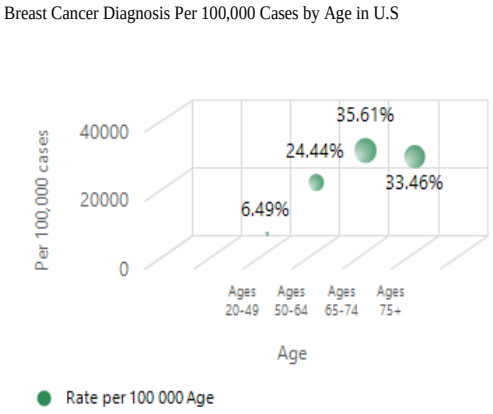


Fig. 18. Scatter plot per 100,000 cases in breast cancer diagnosis (y-axis) at each changes to age at time of diagnosis with 20 -75+ years of age (x-axis)

The scatter plot in Figure 18 of 100,000 cases in breast cancer diagnosis by age shows percentages per age of diagnosis with age range 65-74 years recording the highest at 35.6% while the age range 20-49 years recorded the lowest at 6.49%. Ages 75 and above and 50-64 years recorded 33.46% and 24.44% respectively.

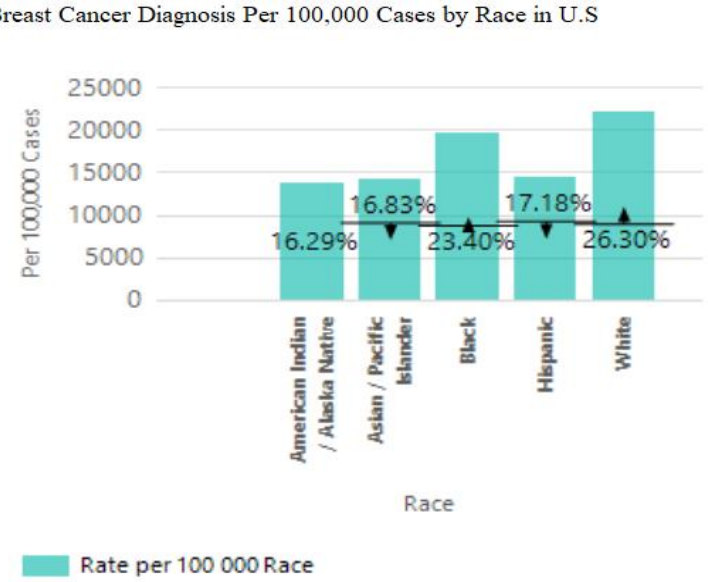


Fig. 19. Bar chart per 100,000 cases in breast cancer diagnosis to the percentage by race in breast cancer diagnosis

The bar chart in Figure 19 of 100,000 cases in breast cancer diagnosis by race shows percentages per race at time of diagnosis with White recording the highest at 26.30%, followed by Black with 23.40%, while American Indian/Alaska recorded the lowest at 16.29%. Hispanic and Asian/Pacific Islander recorded 17.18% and 16.83% respectively.

The line plot in Figure 20 of 100,000 cases in breast cancer diagnosis per year at time of diagnosis with year 2001 recording the highest with a significant decrease yearly and slight increase every other year. Year 2015 shows a significant decrease from 2001.

The Line plot in Figure 21 of 100,000 cases in breast cancer diagnosis per year at time of death by age with ages 75 and above recording the highest mortality rate, followed by ages 65 -74 years and 50 - 64 respectively. Ages 20 - 49 recorded the lowest mortality rate.

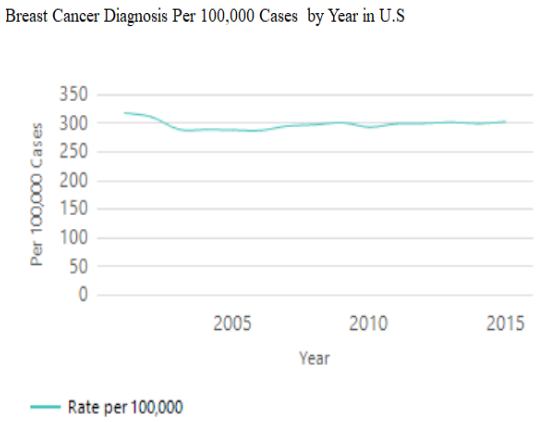


Fig. 20. Line plot per 100,000 cases in breast cancer diagnosis by year (2001 - 2015) in breast cancer diagnosis

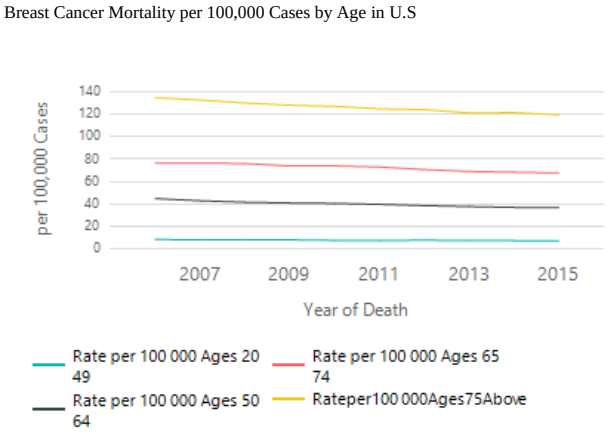


Fig. 21. Line plot per 100,000 cases by age in breast cancer diagnosis (y-axis) at each change to year at time of death (x-axis) with 20 -75+ years of age

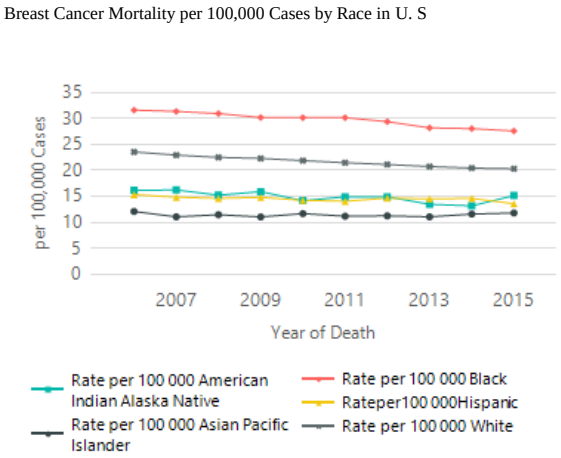


Fig. 22. Line plot per 100,000 cases by race in breast cancer diagnosis (y-axis) at each changes to year at time of death (x-axis) per race

The Line plot in Figure 22 of 100,000 cases in breast cancer diagnosis per year at time of death by race with Black recording the highest mortality rate, followed by White and American Indian/Alaska Native respectively. Hispanic and Asian Pacific Islander recorded the lowest mortality rate respectively.

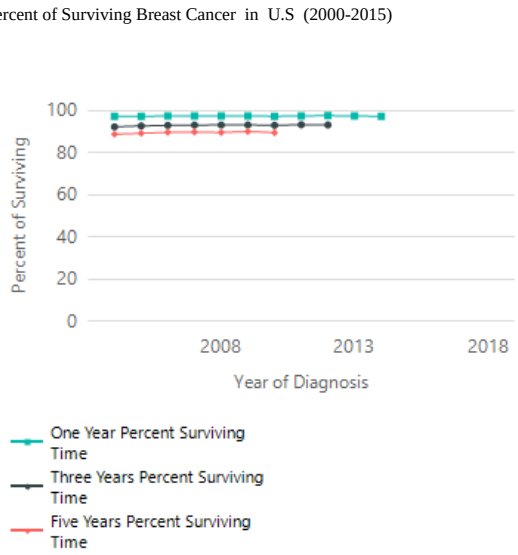


Fig. 23. Line plot per 100,000 cases by percentage of surviving breast cancer from time of diagnosis (y-axis) at each change to year at the time of diagnosing breast cancer (x-axis) per one-year percent surviving, per three- years percent surviving, and per five-years percent surviving

The Line plot of 100,000 cases in breast cancer diagnosis per year at time of diagnosis by surviving time with percentage of one year of surviving breast cancer recording the highest surviving rate at 97.61%, followed by three years of surviving breast cancer recording 93.21% surviving rate. Five years of surviving breast cancer recorded 90.07% surviving rate, the lowest surviving rate.

Fig. 24. Breast Cancer dashboard by race, age, year, mortality rate by race, mortality rate by age, and surviving time per 100,000 cases in U.S

6 Analysis

Looking further into our results we can see the strengths and weaknesses in our process. In comparing our results to other studies and doctors results we can see how it stands up to the current systems. Finally, looking into our data, its characteristics, and how it effects our results we can see what limitations it has imposed upon our study.

Our method of using a CNN has a few advantages and weaknesses. The biggest advantage of using a CNN is its strength in image classification. The ImageNet competition was the gold standard for computer vision. It started in 2010 and featured over a million images in over 1 thousand different classes. For many years many models won the competition with an error rate of 26%, until 2012. Then a CNN call AlexNet won with an error rate of 16.4% [1]. Since that year CNNs continued to win with decreasing error rates [2]. Neural networks are also, very good at dealing with large datasets, at a certain point older algorithms stop improving with an increase of

data while neural networks continue to grow. Another strength is they build their own algorithm on the data. Some of the weakness for us to use CNNs is they need a large amount of data to work properly, even with over 3 thousand images we are on the lower end of what is needed to properly build a CNN. On top of this they are a black box, we do not know exactly what is happening on the inside of our model, we cannot make changes and if we add data, we have to start the whole training process over again [3].

To truly understand the accomplishment of our model we have to compare the results to other studies and doctors. When it comes to detection, we have a rate of 46%. When compared to doctors, who detect 87% of breast cancer [4], it does not look that good. Other studies have also looked into the detection of breast cancer in mammograms with detection rates from 80% [5] to 94% [6]. If you compare it to random chance if is a great improvement. Each mammogram is 3000x16000 pixels on average. We break this up into 512x512 pieces and feed them to our classifier. The probability of picking the right location at random is 1 in 192 or 0.005% chance. These studies have focused on only the detection of masses within mammograms, without looking into the classification side of the equation.

Fig. 25. Comparing our model to Studies and Doctors

While we add the additional step of classification to our model, increasing its complexity. Less research has been completed on the classification of breast cancer in mammography. The other research we found [7] achieved a proper classification rate of 86% on the proper classification of masses only. Our research has a proper classification rate on masses only of 52%.

Finally looking into our dataset is very biased with all our images contain one form or another of cancer. Using this for both our training and validation sets skews our results. It greatly enhances our chances of properly classifying any detected cancer because the CNN does not know what healthy breast tissue looks like. While, it greatly reduces our chances to detect breast cancer within the full mammogram, for the same reasons.

7 Ethics

All innovators must bear the consequences of their creations. Not all discoveries and new developments are used for the better, even when it was not what their creators intended. However, good intentions do not absolve anyone of their ethical responsibility. For example, those who worked on the Manhattan project were ethically responsible to consider the consequences of creating the atomic bomb, and the lives it could destroy. This does not mean they murdered everyone who died at Hiroshima and Nagasaki, but the possibility of these deaths was something they were obligated to consider when making an ethical decision whether or not to create the bomb. It is not different for us as data scientists. While we may not believe our creations will have nearly as large of consequences, we must still consider them.

As data scientists, we must understand how the methods we create to analyze data will be used, and factor that into our decision to release our findings. Once we share our methods of analysis, or even just the results we have found, we cannot control how others will use them. Therefore, we must think about all possible ways our results can be used, and which are most likely. While we can never anticipate every possibility, we can do our best to understand the nature of our creation.

In this paper, we create a way to detect masses in breast tissue based on mammogram images, and then classify the type of mass. Our intention is to detect breast cancer early, which can in turn help save the lives of breast cancer victims that may have otherwise died due to failure to treat the cancer early enough. However, we also must consider the possibility that our method could end up costing lives rather than saving them.

Our method ran a success rate of 64% detection, and only 2% were false negatives. Yet that means that 2% could die to the lack of detection. Our intention is that doctors will still catch the cancer in the same patients they would have prior to our method. However, it is possible that doctors could become too reliant on the automatic detection, and then would fail to identify the cancer in other patients they would have normally found. If this were to be the case, we could potentially be doing more harm than good with our methods and would be better off keeping it to ourselves until we could ensure a higher success rate than current doctors.

Even if we were to get our success rate high enough that we are clearly saving more lives than we could possibly be losing, there are still ethical issues to consider. Utilitarian ethics would dictate that the use of our detection would be a positive force since we save more lives than we would lose. However, the fact that we are now changing who lives and who dies puts our method under much more scrutiny. We may find out that our method has a lower success rate in people of a certain age, race, or sex, which would mean that it could be considered discriminatory. Is it fair to give certain groups of people a lower chance at survival if it means more lives are saved overall? This is an ethical question we would have to consider in this situation.

These ethical questions arise if we were to produce false negatives in our method, failing to detect some breast cancer that could result in a patient's death. However, we also need to consider the implications of producing false positives, where our method could create panic in a patient that is completely healthy. We could be held responsible for a patient overreacting to bad news, or even simply the grief they feel before they learn that they are healthy. Is it unethical to create false panic in more

patients to avoid more false negatives? The answer to this again stems from the way our model is used. If the use is made clear that the results simply indicate an increased risk, and that the patient should be examined further by a doctor, then our model should be justified. However, as said before, we can never guarantee how our model will be used if we were to make it used widespread, so we would need to consider the implications. There could even be a scenario where the results of our model are too accepted by doctors, and a patient receives treatment they do not need, which could prove harmful if not just expensive and uncomfortable.

Another ethical concern is the potential for invasion of privacy through the use of our dataset. Medical records are supposed to be private in our country, but we are using them to train our model. While the images are anonymous, there is still a question if we have the right to use these images without the consent of the patients. Are the medical images the property of the patient, or does the hospital own the images, and is simply ethically obligated to withhold the identity of the patient? If we determined the latter was true, we could assume that the hospital provided these images, and we would have not ethical concerns in using the dataset. The former would be more complex. First of all, we do not know if the patients gave consent for their images to be released by the database where we found our dataset. If they did not, then there is clearly a violation of privacy by the entity that released the information. We could argue that the data is already be publicly available, so our use of it would not result in any further invasion of privacy. However, our participation in the use of the potentially unlawfully procured images would most likely be considered unethical.

8 Future Work

Our work is a small step towards the detection and classification of breast cancer types and pathologies. There is still work to be done in this field. First, a larger dataset with normal mammograms and cropped sub-sections. An evaluation of different object detectors should be done to determine if you only look once or single shot multi-box detector have better results. Next, an evaluation of different forms of images pre-processing and their improvement. Also, being able to directly feed the images as DICOM format could yield strong improvements. Finally, our lack of knowledge about mammograms was our biggest weakness, a multi-disciple team could not only find ways to improve the accuracy of the CNN in detection and classification of mammograms but make a model ready for real word use.

9 Conclusions

An initial look at our results indicate there is still plenty of work to get done before a cancer detector and classifier will be viable for real world applications. Other studies have achieved success rates varying from 81%-98% by limiting themselves to focusing on the detection of a single type of cancer, masses or calcifications, without classifying the pathology. Since our method does not approach that figure, it is best to

attempt detection and classification separately for now. However, our method shows the potential of simultaneous detection and classification, and with some refinement, can be a powerful tool for diagnosing breast cancer in the future.

References

1. Cancer.org. (2019). *Types of Breast Cancer | Different Breast Cancer Types*. [online] https://www.cancer.org/cancer/breast-cancer/understanding-a-breast-cancer-diagnosis/types-of-breast-cancer.html [Accessed 2 Nov. 2018].

2. Mammography 2017. [Online.] https://www.radiologyinfo.org/en/info.cfm?pg=mammo [Accessed 2 Nov. 2018]

3. American Cancer Society Recommendations for the Early Detection of Breast Cancer 2017. [Online.] https://www.cancer.org/cancer/breast-cancer/screening-tests-and-early-detection/american-cancer-society-recommendations-for-the-early-detection-of-breast-cancer.html [Accessed 5 Nov. 2018]

4. Mammography 2017. [Online.] https://www.radiologyinfo.org/en/info.cfm?pg=mammo [Accessed 2 Nov. 2018]

5. Claire, Sissons. Valinda, Riggins Nwadike. What Does Breast Cancer Look Like on a Mammogram? [Online.] https://www.medicalnewstoday.com/articles/322068.php [Accessed 6 Nov. 2018]

6. The Surveillance, Epidemiology, and End Results (SEER): Breast Cancer Recent Trends in SEER Incidence Rates, U.S. Mortality, and Survival. (2000 – 2015) 2018. [Online.] https://seer.cancer.gov/explorer/application.php?site=55&data_type=1&graph_type=2&compareBy=race&chk_sex_3=3&chk_race_5=5&chk_race_4=4&chk_race_3=3&chk_race_6=6&chk_race_8=8&chk_age_range_1=1&chk_data_type_1=1&advopt_precision=2&advopt_display=1&showDataFor=sex_3_and_age_range_1_and_data_type_1 [Accessed 21 Jan. 2019]

7. F. Rosenblatt.: The perceptron, a perceiving and recognizing automaton Project Para. Cornell Aeronautical Laboratory, 1957.

8. Andrey Kurenkov: A 'Brief' History of Neural nets and Depp Learning (2015) http://www.andreykurenkov.com/writing/ai/a-brief-history-of-neural-nets-and-deep-learning/

9. Ujjwlkarn.: An Intuitive Explanation of Convolutional Neural Networks. (2016) https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/

10. Prince Grover.: Evolution of Object Detection and Localization Algotithms (2018) https://towardsdatascience.com/evolution-of-object-detection-and-localization-algotihms-e2410212d8bad

11. Cancer Image Archive.: Curated Breast Imaging Subset of DDSM. (2018) https://wiki.cancerimagingarchive.net/display/Public/CBIS-DDSM#83dd9b518d554b2983368fd0d778f31b

12. Demush, R. (2019, February 27). A Brief History of Computer Vision (and Convolutional Neural Networks). Retrieved from https://hackernoon.com/a-brief-history-of-computer-vision-and-convolutional-neural-networks-8fe8aacc79f3

13. Bandura, O. (2017, June 01). The Great Expectations of the ImageNet Challenge 2017. Retrieved from https://www.scnsoft.com/blog/imagenet-challenge-2017-expectations

14. Donges, N. (2018, April 17). Pros and Cons of Neural Networks. Retrieved from https://towardsdatascience.com/hype-disadvantages-of-neural-networks-6af04904ba5b

15. Mammogram Accuracy - Accuracy of Mammograms. (2019, January 07). Retrieved from https://ww5.komen.org/BreastCancer/AccuracyofMammograms.html

16. Sompaio, W. B., & Deniz, E. M. Detection of masses in mammogram images using CNN, geostatistic functions and SVM. Computers in Biology and Medicine. Vol. 41, Issue 8, August 2011, pp 653-664.

17. Dunghel, N., Carneiro, G., & Bradley, A. P. (n.d.). The Automated Learning of Deep Features for Breast Mass Classification from Mammograms. Retrieved from https://cs.adelaide.edu.au/~carneiro/publications/mass_classification_miccai.pdf