EN.520.433 MEDICAL IMAGE ANALYSIS

Project 1 Report

Mammography: Abnormality Detection and Diagnosis

Tae Jin Kim

Team 11: Laura Chicos, Ganesh Arvapalli, Shayan Roychoudhury

Abstract

Breast cancer, statistically shown to be the most common cancer found among women, can be treated with improved survivability rates if diagnosed early in its development. As such, the production of effective CAD systems that are able to detect the existence of cancerous masses, to determine whether the mass exists on the left or right breast, to diagnose the mass as benign or malignant, and to produce an accurate generalization of the mass's location with masking will greatly advance the medical field. After implementing various algorithms that relied heavily on examining the intensity frequency distributions provided by different histograms, we eventually implemented a system that was able to obtain results with low computational time and high accuracy. As such, we were able to conclude that, at least with the training and testing data sets we specifically received for this project, implementing unestablished methods can at times compete with popular established methods.

1 Introduction

Deemed to be the most common cancer found among women, breast cancer has, according to the 2018 statistic provided by the American Cancer Society, estimated figures of 266,120 new cases and 40,920 deaths for females as of February 26, 2018 [1]. Fortunately, early detection can greatly improve survival rates as the cancer is most treatable before it begins spreading throughout the body.

Currently, the oldest and most commonly used method for finding early indications of breast cancer among patients is mammography, a specialized medical imaging that uses low-dose x-rays in order to examine the inside of the breast. For the past couple decades, recent advances have been made to this area of study, such as digital mammography, where the x-ray image is stored digitally in a computer for review by a radiologist, computer-aided detection (CAD) systems that uses computers to search the digital images for any abnormalities, and breast tomosynthesis, an advanced form of breast imaging that produces a 3-D digital model of the breast from images taken of the breast at various angles [2].

For this specific project, we were asked to come up with our own CAD system that would be able to implement four basic tasks: For each subject 1) diagnose whether or not the breasts are healthy (i.e. has no mass) or suspicious (i.e. has mass), 2) determine whether the left or right breast contains the cancerous mass (only one of the two provided breast for each suspicious subject was guaranteed to have a mass), 3) diagnose whether the cancer is benign or malignant, and 4) produce masks of the mass for both breasts.

Before committing to a specific method of analysis, we decided it would be best to observe other existing CAD systems, which commonly consists of four basic steps: Pre-processing, Segmentation, Feature extraction, and Classification [3]. Generally, the pre-

processing step would involve resizing the digital images for both faster computational time and to allow for a more balanced pixel distribution between the two breast images. This step can also involve image enhancement, which has been previously implemented via various methods, such as thresholding, gamma correction, histogram equalization, unsharp masking, subtracting Laplacian, adaptive contrast enhancement, diffusion filtering, and median filtering [4]. Segmentation will then use methods such as region growing, fuzzy sets, active contours, Hough transform, Gabor wavelets, and edge flow propagation in order to distinguish the various regions of interest presented on the digital mammogram, such as the breast boundary, the pectoral muscle, and the mass itself [5]. The mass's features, such as its shape or spiculation, can then be extrapolated in order to determine whether the mass is benign or malignant. Upon obtaining all of this information, the images and their masses can finally be classified.

Our implementation relied heavily on understanding how the cancerous masses would manifest themselves on digital mammograms, each of which would contain three dominant features: clustered micro calcifications, thin white specks scattered throughout the breast that represent calcium deposits, the pectoral muscle, and the mass, which will generally have a low contrast that is similar fibro glandular tissues [6].

Keeping in mind how the radiological representation of these features would tend towards brighter intensities, we decided to focus on the use of various histogram distributions as a means to detect any abnormalities in the breasts and make a final diagnosis. The entire algorithm was written in MATLAB and produced surprisingly effective results for both the training and testing set. This surprise is due to our initial belief that our algorithm relied too heavily on the values obtained from the training set and would therefore not be able to correctly accomplish the necessary tasks if provided a completely new dataset.

2 Theory and Background

As stated before, the implementation we used for this project relied heavily on histograms as they provided the various ranges of intensity frequencies for each breast. The reason why we decided to implement intensity frequencies as the basis of our implementation was because we understood that each distinguishable section in the mammogram would belong in its own unique intensity range. As such, all we would have to do would be to find those specific intensity ranges and separate each section of the mammogram accordingly. We also had a weaker understanding of the more complicated algorithms proposed by other scientific journals and therefore decided to focus on the implementation of a faster but computationally weaker algorithm.

2.1 Intensity Frequency Representation

Although there exist various ways of presenting the intensity frequencies of an image, the two main methods we focused on were the use of image histograms and polar histograms, otherwise known as rose diagrams.

In MATALB, `imhist` automatically creates an image histogram of an image; however, this information can be made even more useful by using another MATLAB function called `findpeaks`, which presents the intensity frequencies from a range from 0 to 255 and stores the location and values of the histogram's peaks.

Meanwhile, polar histograms are useful for obtaining a more comprehensive understanding of how the intensity frequencies are distributed in the image. In other words, polar histograms can provide us with a better idea of how to make distinctions among various images. Shown below are examples of such representations:

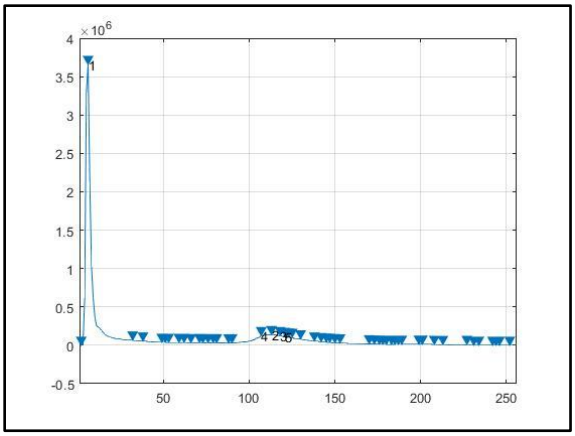

Figure 1a: Original Peak Histogram

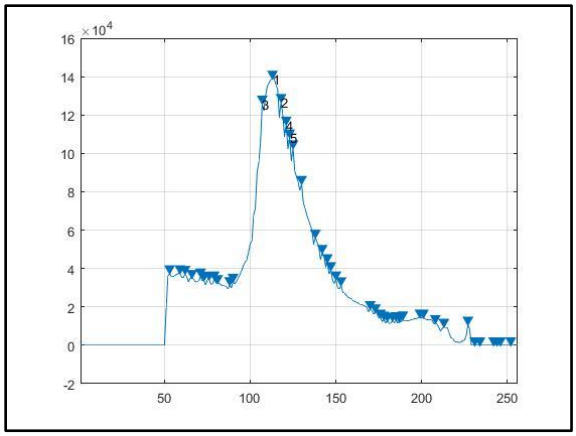

Figure 1b: Updated Peak Histogram

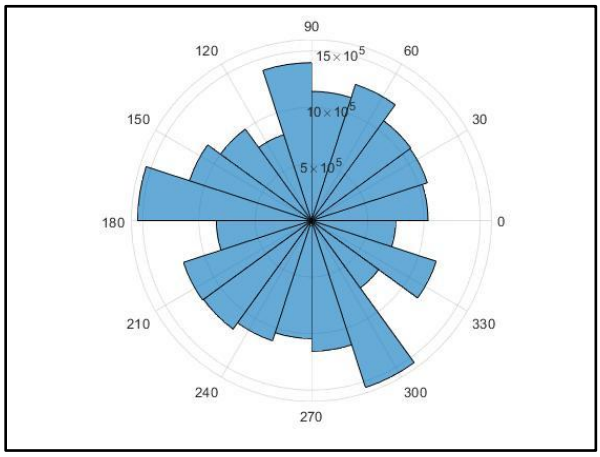

Figure 1c: Polar Histogram

Once should notice immediately that the initial peak histogram (Figure 1a) tells us very little of how the intensity frequencies are distributed in the image, besides the fact that much of the image is composed of a black background. It is therefore necessary for us to ignore all of the lower intensities that do not have any contribution to finding the abnormalities in the breast, as shown in Figure 1b, where the more relevant top 5 peak intensity frequencies are clearly distinguished. Meanwhile, without having to ignore any of the lower intensities at all, the polar histogram is able to present a much more easily digestible representation of how the intensity frequencies are distributed in the image. However, it is important to remember that because the images of each breast have different dimensions, there will be an asymmetric appearance

between the two breasts. Therefore, in order to ensure a balanced representation of each image in the polar histograms, we must set a constant limit on the bin sizes; specifically, every bin size was set to pi/10 (i.e. every image was split among 20 equally-sized intensity ranges).

3 Methods and Implementation

Presented below is the block diagram that provides a very basic overview of the implementation we used for breast abnormality detection and diagnosis. The four main tasks described in the introduction are shown as follows: 1) The first line diagnoses the breasts as either healthy or suspicious. If the breasts are diagnosed to be healthy, the algorithm is terminated in order to save computational time. 2) The next task is to determine which breast contains the mass. 3) Upon making this determination, the tumor in the affected breast is categorized as either benign or malignant with thresholding. 4) Finally, the polar histogram obtained before is used to update the mask.

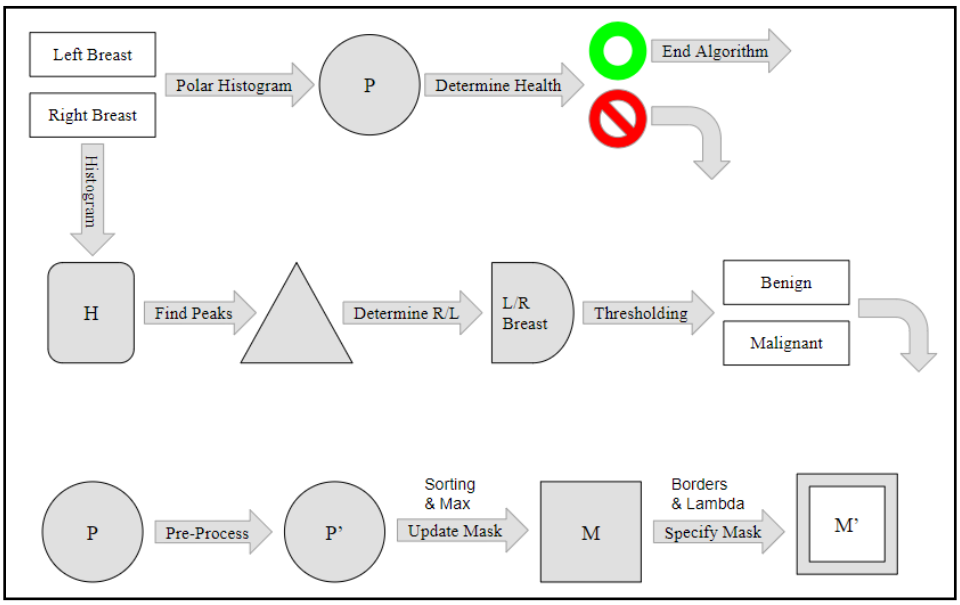

Figure 2: Block Diagram

3.1 Healthy or Suspicious

In order to better understand the reasoning behind the algorithm we used for determining whether a pair of breasts is healthy or suspicious, please take some time to notice a trend among the 3 pairs of rose diagrams for the healthy, benign, and malignant breasts.

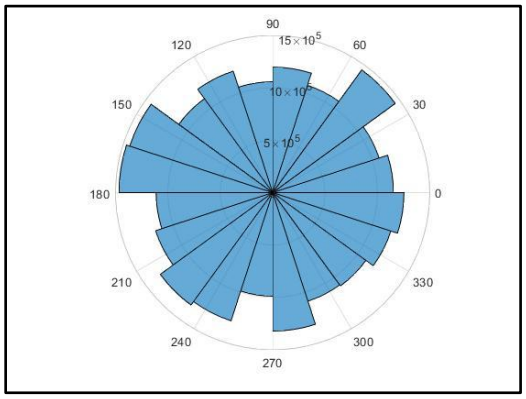

Figure 3a: Polar Histogram of Healthy Sample 1008_LEFT

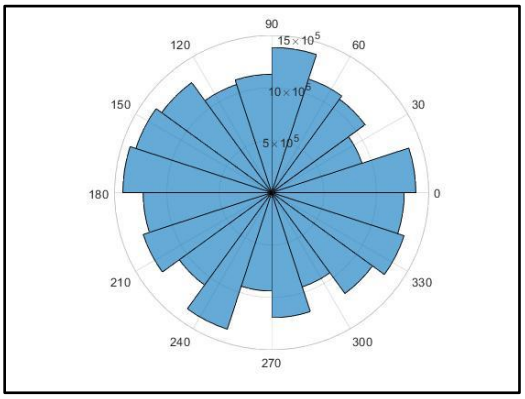

Figure 3b: Polar Histogram of Healthy Sample 1008 RIGHT

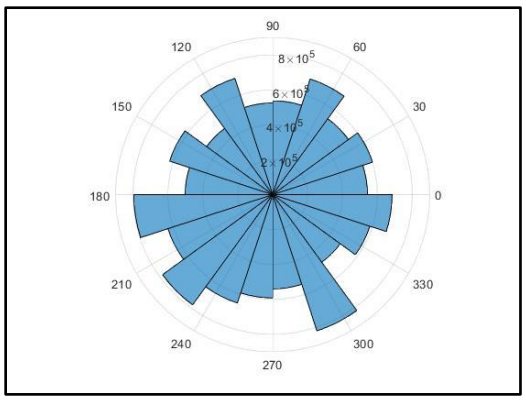

Figure 3a: Polar Histogram of Benign Sample 2008 LEFT

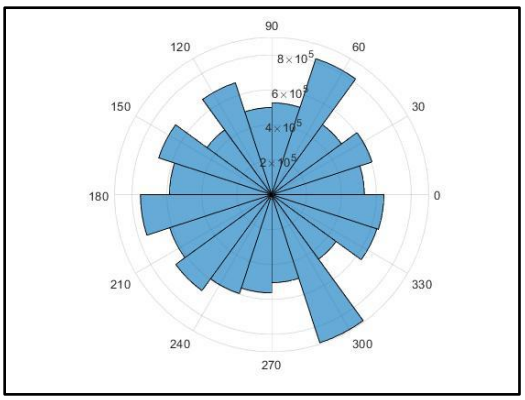

Figure 3b: Polar Histogram of Benign Sample 2008 RIGHT

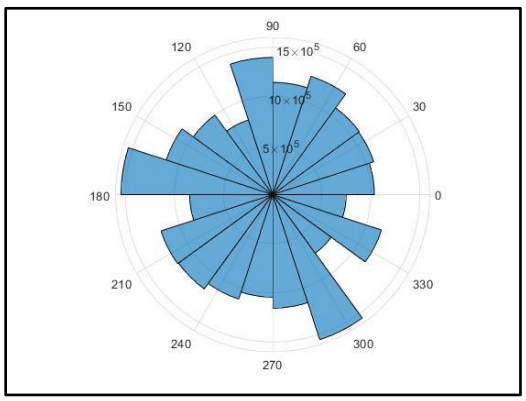

Figure 3a: Polar Histogram of Malignant Sample 3008 LEFT

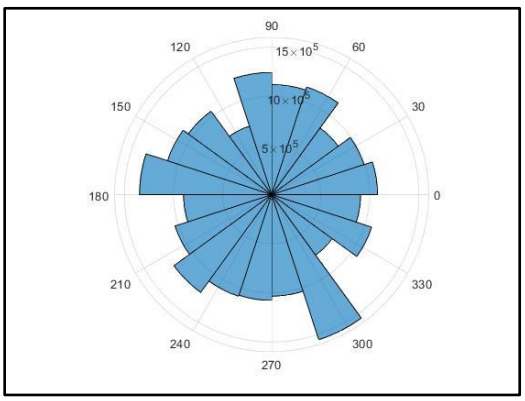

Figure 3b: Polar Histogram of Malignant Sample 3008 RIGHT

After examining the images above, we noticed that all of the suspicious samples (i.e. the benign and malignant samples) have a protruding slice at the 17^{th} bin (the bin located at angle 300) while the healthy samples have a more balanced distribution of intensity frequencies. The reason for this protrusion can be attributed to the fact that all of the suspicious samples would have a section of the image dedicated to some higher intensity, allowing us to make the generalization that all suspicious samples would have a protruding slice at the 17^{th} bin. Keeping this trend in mind, we then implemented thresholding where if the value obtained after subtracting the sum of the left and right counts at slice 18 from the sum of the left and right counts at slice 17 is negative, then the sample was diagnosed as healthy; otherwise, the sample was diagnosed as suspicious.

3.2 Left or Right Breast

We then had to come up with an algorithm that would be able to determine whether the left or right breast contained the mass. After taking some time to examine all of the digital mammograms, we realized that, for the most part, the breast infected with the cancer had a more distributive range of intensities as it will have a mass that tends to be surrounded by an intricate set of microfibers. However, because we knew that the rose diagram would not be effective at displaying this minute distribution, considering that we had preset all of the bin sizes to be pi/10, we decided that it would be better to use a regular histogram instead.

After obtaining the histogram, we made sure to implement some pre-processing by ignoring all of the lower intensities as the only intensities we wished to truly focus on were the brighter ones. Upon obtaining this updated histogram, we then use MATLAB's function `findpeaks` in order to obtain the peak frequencies of the intensities on the histogram. We then took the 5 highest peaks and saved the mean values of the 5 highest peaks for the left and right

breast. If the left mean value was found to be greater than the right mean value, then we concluded that the right breast must contain the mass; otherwise, we concluded that the left breast must contain the mass. As explained before, the histogram for an infected breast will have lower peaks spread around the biggest peak as the intensity frequencies would be more distributed while the healthy breast will have many high peaks around biggest peak because the intensity frequencies would not be as distributed. In other words, the mean peak value will usually be higher for the healthy breast.

3.3 Benign or Malignant

We then had to come up with an algorithm that would be able to determine whether the infected breast contained a benign or malignant mass. After taking some time to compare all of the mammograms and rose diagrams of the benign and malignant samples, we realized that, generally, the malignant mammograms had a more distinct contrast between the mass and the rest of the breast. We then concluded that this would result in higher standard deviation values for the counts of the bins of the malignant polar histograms than for the counts of the bins of the benign polar histograms. After calculating all of the standard deviations for the infected breasts, we found out that a threshold of 1.65e5 brought about the most accurate results.

3.4 Mask for the Mass

The final task for us to implement was to create a mask of the mass in the infected breast. We decided to use the rose diagram because, as stated before, the protrusion created in slice 17 best helped reveal what range of intensities that contained the tumor. In order to ensure that only the affected areas would be taken into account, I applied a limit upon the rose diagram so that all the darker intensities were ignored. The intensities were then ordered based on their frequency,

with the most frequent intensity being used as the basis for where the mask should set its values to 1. Upon updating the mask, the highlighted regions were edited so that any regions outside the most concentrated region in the mask had their values set to 0. The algorithm then took the mean position of this concentrated region and treated it as its center. A constant lambda was used as the radius of the area around the calculated center in order to account for possible error.

4 Alternate Considerations

For each of the four basic tasks we used to implement our design, we had to come up with various possible algorithms that all ultimately did not match up with our final design.

4.1 Healthy or Suspicious & Left or Right

When we first tried to determine whether the set of breasts were healthy or suspicious, we initially used regular histograms rather than the polar histogram. As such, the distribution of the intensities for each image was not as easily distinguishable as when using a polar histogram. This lack of understanding led us to assume that, we could determine which set of breasts was healthy or suspicious and whether the left or right breast contained the tumor at the same time. We believed that, if given a suspicious sample where the left breasts contained the tumor, then the left breast should have a higher frequency of higher intensities. However, what we failed to consider at the time was that the histograms we were using were not split into consistent bins that would allow for a balanced representation of the intensity ranges of each respective image. In other words, instead of looking for a trend found between healthy and suspicious samples, we

were just blindly comparing the most frequently found intensities of each image without understanding what those intensity frequencies actually meant.

4.2 Benign or Malignant

We ended up having to explore various methods for determining whether a mass was benign or malignant. We initially tried subtracting reflected breast images from each other in hopes of having the cancerous mass be all that was left. However, this assumption failed to take into considering the asymmetry of the two breasts.

We then took the top 10 peaks from each breast and took the mean of their locations, believing that the higher location mean should indicate that a malignant mass. The reason for this assumption was that, for the most part, the malignant masses had much higher contrast. However, this assumption failed to take into consideration that the higher contrast also let to high peaks at darker intensities.

4.3 Mask for Mass

Initially, when implementing masking, we found that, for some reason, the most isolated peak in the peak histogram (i.e. the peak farthest from all the other peaks) was a good indicator for finding the cancerous mass in a malignant sample. However, it performed very badly for benign samples. We soon realized that all we were doing was finding the area in the image with the highest contrast to its background, which made it so we were able to locate the masses in malignant sample so well, but not so much in the benign samples.

5 Experiments and Their Results

We will now examine the results obtained from implementing the methods described above. This will help clarify why we decided to stick with the implementations we had constructed rather than try to follow the proposed complicated, time-consuming algorithms.

5.1 Healthy or Suspicious

As explained in the methods section of this report, diagnosing the health of the patient was determined by obtaining the difference of the sum of the breasts' 18^{th} slice counts from the sum of the breasts' 17^{th} slice counts, resulting in the following table.

Healthy Sample Difference	Benign Sample Difference	Malignant Sample Difference
1002: -5.2e5	2002: 2.7e5	3001: 1.8e5
1003: -6.0e5	2003: 4.5e5	3002: 7.7e5
1004: -6.5e5	2004: 1.1e6	3004: 1.2e6
1005: -3.5e5	2005: 8.6e5	3006: 2.4e5
1006: -6.2e5	2006: 7.5e5	3008: 1.6e6
1007: -3.4e5	2007: 8.2e5	3009: 1.3e6
1008: -2.5e5	2008: 7.8e5	3011: 1.3e6

Table 1: Differences of the sum of 18^{th} slices from the left and right breast from the sum of 17^{th} slices from the left and right breasts

All positive differences were diagnosed as suspicious samples while all negative differences were diagnosed as healthy samples, and as seen from the table above, this resulted in us obtaining 100% accuracy for distinguishing between healthy and suspicious samples.

5.2 Left or Right

For this section, we examined the means of the top 5 peak intensities found in the image histograms of the left and right breasts, resulting in the following table:

File	L_mean	R_mean	prediction	actual
Benign				
2002	68844.6	62275.4	R	L
2003	96700.2	96879.2	L	L
2004	99371.8	131987.2	L	L
2005	138313.6	141141.4	L	L
2006	95855.4	91150.6	R	R
2007	186532.8	198775.2	L	L
2008	29129.0	41247.8	L	L
Malignant				
3001	57291.6	67283.0	L	L
3002	105963.2	112182.6	L	L
3004	155193.6	216382.6	L	R
3006	88160.6	104512.6	L	L
3008	123025.4	125481.0	L	L
3009	249380.8	312646.6	L	L
3011	166453.2	206573.4	L	L

Table 2: Means of the counts of the top 5 peak intensity frequencies before normalization (red indicates an incorrect prediction)

However, after taking some time to look over the algorithm, we realized that it would be a better idea to use the normalized means when determining which breast contained the mass. This is because we realized that the differently sized images resulted in some of the peaks having large values that disrupted the consistency of the algorithm. Here is the table of the new values:

File	L_mean	R_mean	prediction	actual
Benign				
2002	0.090347	0.100650	L	L
2003	0.091115	0.099387	L	L
2004	0.077246	0.139343	L	L
2005	0.061975	0.077009	L	L
2006	0.091128	0.073259	R	R
2007	0.095590	0.069263	R	L
2008	0.034668	0.050054	L	L
Malignant				
3001	0.048990	0.062662	L	L
3002	0.069422	0.082448	L	L
3004	0.103057	0.100928	R	R
3006	0.065383	0.067233	L	L
3008	0.073904	0.172703	L	L
3009	0.106677	0.125964	L	L
3011	0.093506	0.184430	L	L

Table 3: Means of the counts of the top 5 peak intensity frequencies after normalization (red indicates an incorrect prediction)

As seen from above table, we were able to improve the prediction upon normalizing the means, with only one of the samples being predicted incorrectly.

5.3 Benign or Malignant

For this algorithm, we used a threshold value. If the standard deviation was less than the threshold, the sample was diagnosed as benign; otherwise, it was diagnosed as malignant.

Std of Benign Sample	Prediction < 1.65e5 = B	Std Malignant Sample	Prediction > 1.65e5 = M
2002_L: 8.85e4	B	3001_L: 1.62e6	M
2003_L: 7.27e4	B	3002_L: 1.74e5	M
2004_L: 1.64e5	B	3004_R: 1.38e5	B
2005_L: 1.16e5	B	3006_L: 7.20e4	B
2006_R: 1.23e5	B	3008_L: 2.26e5	M
2007_L: 1.20e5	B	3009_L: 1.71e5	M
2008_L: 1.05e5	B	3011_L: 1.66e5	M

Table 4: Standard Deviation of the Infected Breast (red indicates an incorrect prediction)

As seen in the table above, all of the benign tumors were predicted correctly while only two of the malignant tumors were predicted incorrectly, meaning that out of the 14 samples, only 2 were thought to be benign when they were actually malignant. Although the accuracy of this prediction is very high, in an actual medical environment, the predictions above can lead to disastrous results. This is because a patient with a malignant tumor will be thought to have a benign tumor and will not receive the immediate treatment (s)he may have received if the prediction had been correct to begin with.

5.5 Mask for Mass

Shown below are the mask images obtained when examining one benign sample (2008) and one malignant sample (3008) with each respective step labeled:

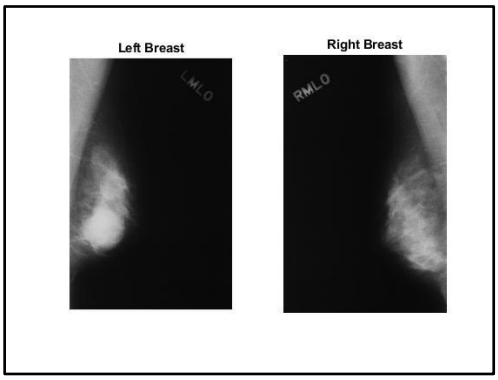

Figure 4a: Benign Sample (2008)

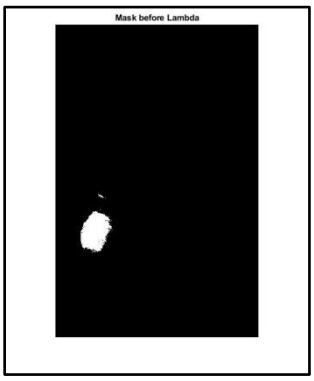

Figure 4b: Initial Mask

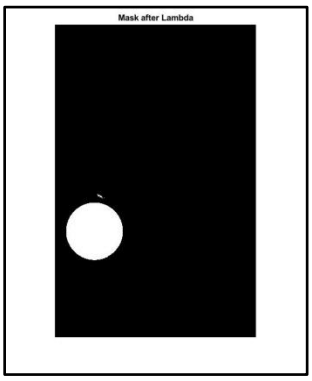

Figure 4c: Final Mask

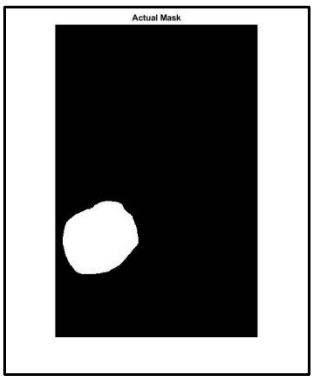

Figure 4d: Final Mask

As seen above, the most concentrated region is first singled out. After the center is found, a lambda constant is used as the radius to create a circular area around the center point.

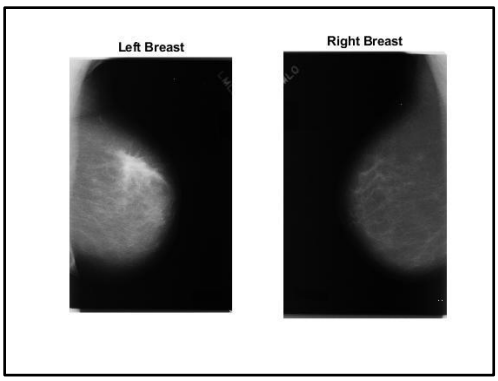

Figure 5a: Benign Sample (3008)

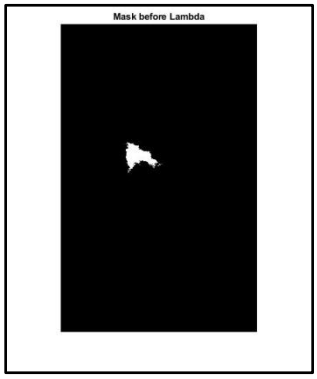

Figure 5b: Initial Mask

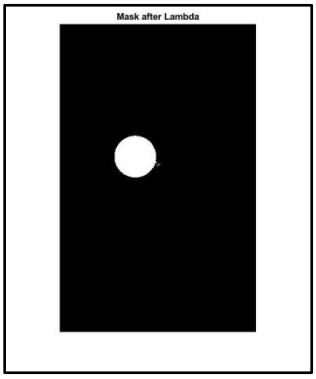

Figure 5c: Final Mask

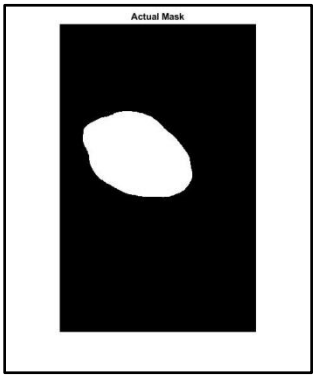

Figure 5c: Actual Mask

As seen in the masks above, the initial mask is usually able to pinpoint the general area of the cancer's location; however, when visually comparing the finals masks with the actually masks, we can see that there is not a consistent overlap. Rather, it seems that the lambda radius used is not large enough to take into account the entirety of the actual mask. However, the lambda value used, which in this case is 500, was brought about after experimenting with various lambda values, and 500 was determined to produce the most accurate masks.

6 Discussion

6.1 Review

After having completed out implementation, we ended up with the following results when running our implementation on the training set:

Mean mask overlap (Left side):
0.6578
Mean mask overlap (Right side):
0.91337
Sensitivity of cancer:
0.85714
Specificity of cancer:
0.97143
Diagnostic accuracy:
0.90476

Initially, we feared that we may have overfit to the training data as most of our algorithm decisions were based upon first observing how all of the training data looked. One possible solution to this fear would have been to implement normalization more thoroughly throughout our implementation to ensure that all of the calculated values would reflect a more balanced representation of the data as each value would be proportional to its own image.

6.2 Other Improvements

In addition, during the masking step, instead of using a lambda constant as the radius for the circular area around the mask's center, we could have used region growing instead. In addition, we could have brought about a more exact separation of parts in the mammogram if we implemented multiple algorithms instead of focusing solely on intensity frequency distribution, even if that meant spending more on computational time.

6.3 Final Presentation

However, our fears ended up being unfounded as our implementation ended up ranking second among all of the 11 groups' implementation. The scores are shown below:

	Time		Dice (left)		Dice (right)		Sensitivity to caner		Specificity to cancer		Diagnostic accuracy		Combined	Overall
Team No.	Value	Ranking	Value	Ranking	Value	Ranking	Value	Ranking	Value	Ranking	Value	Ranking	Ranking	Ranking
1	138.7	7	0.5066	7	0.7778	2	0.33	6	1.00	1	0.6667	3	26	**4**
2	87.6	6	0.6385	2	0.5556	8	0.67	3	0.87	7	0.6667	3	29	**6**
3	5403.2	11	0.5152	6	0.6208	7	0.00	7	0.87	7	0.5556	8	46	**9**
4	20.7	3	0.5600	5	0.6667	6	0.00	7	1.00	1	0.6667	3	25	**3**
5	203.4	8	0.4444	9	0.5556	8	0.00	7	0.80	9	0.5000	10	51	**10**
6	569.3	9	0.6289	3	0.7778	2	0.00	7	0.93	3	0.6667	3	27	**5**
7	32.1	4	0.4423	10	0.7556	5	1.00	1	0.67	10	0.6111	7	37	**7**
8	41.5	5	0.5041	8	0.5553	10	0.00	7	0.93	3	0.5556	8	41	**8**
9	1043.2	10	0.0595	11	0.0102	11	1.00	1	0.13	11	0.1667	11	55	**11**
10	15.6	2	0.7210	1	0.8332	1	0.67	3	0.93	3	0.7778	1	11	**1**
11	9.4	1	0.6130	4	0.7778	2	0.67	3	0.93	3	0.7778	1	14	**2**

The two categories in which our implementation excelled the best in was in time and diagnostic accuracy, the first of which can be attributed to our use of algorithms that did not involve complicated computational power, and the latter revealing that our algorithm, despite its low runtime, remains somewhat accurate in its diagnosis.

One reason I came up with for why our algorithm might have done so well is that both the training and testing set may have originated from the same original set and was arbitrarily divided into the training and testing set in order to fulfill the assignment's requirements. This can possibly explain why overfitting to the training set allowed us to obtain such accurate results.

7 Conclusion

This project attempts to find abnormities in the provided mammograms in order to make an accurate diagnosis of whether the sample is healthy or suspicious, whether the cancerous mass is located on the left or right breast, whether the cancerous mass is benign or malignant, and to use all of the information gathered from all the previous steps to create a mask of the mass.

Despite implementing algorithms in each step that did not require much computational power, we were able to obtain quick results with high accuracy. Although this can be attributed to possible overfitting to data from the same original dataset, this project helped show us that even unestablished methods can bring about results that can compete with the established methods.

8 Team Evaluations

My contributions to this project:

- Set up the entirety of the initial code that was submitted before the final submission
    - Diagnose as healthy / suspicious based on mean intensity (failed)
    - Create a mask based on how unique a peak intensity was (somewhat successful)
    - Diagnose as benign / malignant based on distribution of peak intensities (failed)
- Coded for determining whether the left or right breast contained the mass (same as initial)
- Coded for the masking of the masses using polar histograms instead (successful)
- Using the mistakes I learned from the initial code, I provided my partner with various other options to implement that I thought might help bring about better results
- Setup the PowerPoint that was used for our presentation

Laura's contributions to this project:

- Took the time to examine all of the mammograms to find a trend among them
- Came up with an accurate method for diagnosing as healthy or suspicious
- Came up with a somewhat accurate method for diagnosing as benign or malignant
- Contributed in our discussions for finding the most optimal methods

- Worked on presentation slides that reflected her tasks

Ganesh and Shayan provided us some hints when we asked questions, but for the most part, the two of them focused on completing project 2. Everyone put in the time and effort to make sure that all of our work was submitted on time. They always made sure to show up when we decided to meet up and did not make an excuse to run away from the work. I believe that everyone deserves full credit for the work they put in.

Bibliography

https://www.cancer.gov/types/common-cancers

https://www.healthline.com/health/breast-cancer/how-does-breast-cancer-affect-the-body#symptoms

https://blackboard.jhu.edu/bbcswebdav/pid-5119365-dt-content-rid-22654810_2/courses/EN.520.433.01.SP18/2018-03-Mammography-A-Introduction.pdf

https://blackboard.jhu.edu/bbcswebdav/pid-5119365-dt-content-rid-23161886_2/courses/EN.520.433.01.SP18/2018-04-Mammography-B-Enhancement.pdf

https://blackboard.jhu.edu/bbcswebdav/pid-5119365-dt-content-rid-23161887_2/courses/EN.520.433.01.SP18/2018-05-Mammography-C-Features.pdf

https://deepblue.lib.umich.edu/bitstream/handle/2027.42/134888/mp6640.pdf?sequence=1&isAllowed=y

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5379115/