Summer 8-2019

# Predictive Diagnostic Analysis of Mammographic Breast Tissue Microenvironment

Dexter G. Canning
*University of Maine*, dcanning2015@gmail.com

# PREDICTIVE DIAGNOSTIC ANALYSIS OF MAMMOGRAPHIC BREAST TISSUE MICROENVIRONMENT

By

Dexter G. Canning

A Thesis Submitted in Partial Fulfillment
of the Requirements for a Degree with Honors
(Mathematics)

The Honors College

University of Maine

June 2019

Advisory Committee:

Andre Khalil, Associate Professor of Biomedical Engineering, Advisor
RW Estela, Preceptor in the Honors College
Karissa Tilbury, Assistant Professor of Biomedical Engineering
Brian Toner, Ph.D. Candidate in Computer Science
Zheng Wei, Assistant Professor of Statistics

# ABSTRACT

Improving computer-aided early detection techniques for breast cancer is paramount because current technology has high false positive rates. Existing methods have led to a substantial number of false diagnostics, which lead to stress, unnecessary biopsies, and an added financial burden to the health care system. In order to augment early detection methodology, one must understand the breast microenvironment. The CompuMAINE Lab has researched computational metrics on mammograms based on an image analysis technique called the Wavelet Transform Modulus Maxima (WTMM) method to identify the fractal and roughness signature from mammograms. The WTMM method was used to color code the mammograms based on the type of tissue present and assign the Hurst exponent (H) value to corresponding tissue: dense tissue with H greater than 0.55, fatty tissue with H less than 0.45, and disrupted tissue with H between 0.45 and 0.55, with the latter being a key trait in tumorous tissue. This analysis on the full breast was performed on 127 cases for the Medio Lateral Oblique (MLO) view. We are revisiting these data by analyzing the region behind the nipple for the MLO view and the region outside the nipple area. After performing the WTMM analysis on each breast, non-parametric statistical analysis methods were performed to determine the level of significance between normal, benign, and cancerous cases. Furthermore, we utilized logistic models to assess the predictability of these metrics for future datasets.

# ACKNOWLEDGEMENTS

In my years as an undergrad, I learned great mentors not only inspires confidence in their abilities, but they also enable me to become confident in mine as well. I'm privileged to have been able to intern in the CompuMAINE Lab under the direction of a great mentor in my undergrad career, Professor Andre Khalil. I found not only my academic interests, but also the right direction in life. Furthermore, I would not have succeeded in this massive undertaking without the unwavering assistance from my past supervisor Zach Marin and my current supervisor Brian Toner, and my thesis advisor Professor Andre Khalil. The combined research angles from the 2017 paper on microenvironment changes in tumorous breasts is the motivation behind the thesis. Also, I would like to thank Professor Zheng Wei, another great mentor, who has helped me immensely with the statistics and also agreed to be my advisor for the duration of my Master's Program starting fall of 2019. Lastly, I'd like to thank all my friends and family members who have helped along the way. Also, a special thanks to Matt Ryckman and Amber Hathaway, who have been gracious enough to read my thesis and offer corrections.

TABLE OF CONTENTS

# 1. INTRODUCTION

## 1.1 Motivation Behind The Research

Breast cancer is a common type of cancer among women in the US. It is estimated that there will be around 268,600 breast cancer cases in 2019 in the US alone, and the mortality rate for breast cancer patients in the US is around 15.5% [1]. Breast cancer is categorized as localized, regional, or distant. For patients with access to proper early treatment, survival rates are relatively high for individuals in localized or regional stages, but low for individuals in the distant stage. The survival rate for localized cancer in a five-year period is around 98.8%, while for patients diagnosed during the regional or distant stages, the survival rates are about 85.5% and 27.4%, respectively. In total, the chance of survival in the 5-year period for any given breast cancer patient with access to proper medical care irrespective of stage is around 89.9% [1].

Early cancer diagnostic tools including digital mammograms and Computer Aided Detection/Diagnostic (CAD) software have helped prevent deaths caused by breast cancer. However, current CAD software has a false positive rate of around 15% given the specificity of CAD is around 85%. False positive rate in this context means being diagnosed with breast cancer, when a patient does not have breast cancer, and specificity in this context means correctly identifying breast cancer patients [2]. False positive rates incur an unnecessary financial burden and unnecessary stress on patients. In 2012 and 2013 alone, the false positive diagnoses cost the US around 4 billion dollars per year [3]. However, because of the potential cost savings and improved diagnostics, more accurate CAD technology is in high demand.

## 1.2 Current Mammography Screening Techniques

### 1.2.1 Who Receives Mammograms?

The primary function of mammograms is that of early detection. Since early detection plays such an important role in the survival rate of breast cancer, the American College of Radiology (ACR) recommend that women forty-five and above receive annual screenings. Further, the ACR also advises women who have additional risk factors (i.e. have a family history of breast cancer, were previously diagnosed, etc) to participate in annual screenings [4].

1.2.2 Mammography Procedure

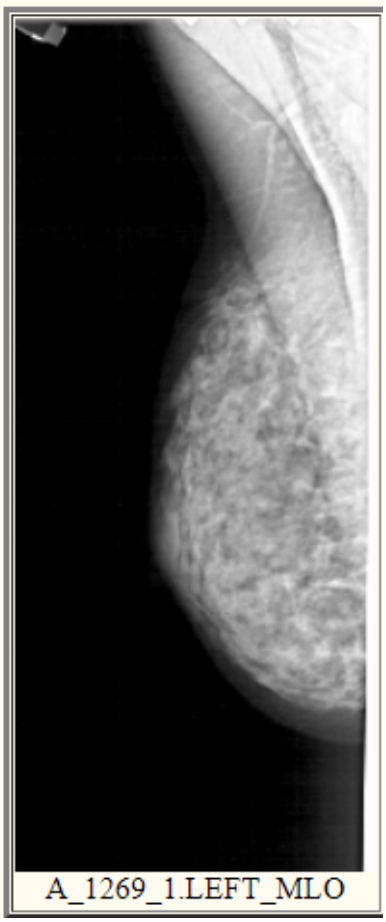


**Figure 1.1**

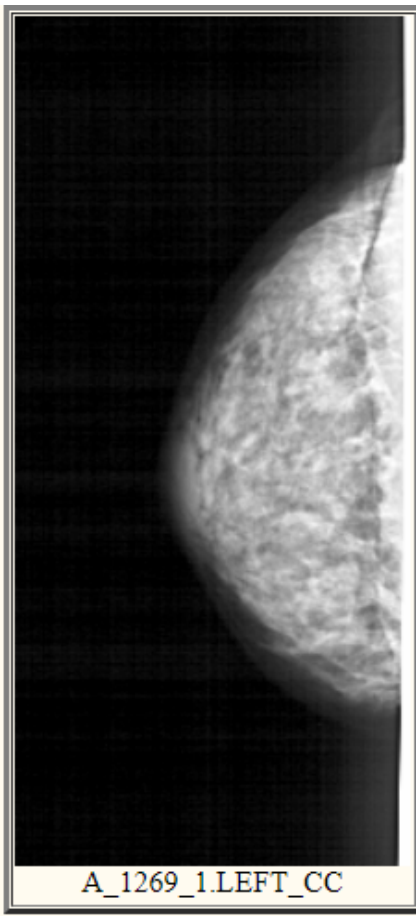


**Figure 1.2**

Figure 1.1 and 1.2 are taken from [5].

Two-dimensional (2D) mammography imaging uses two different views, the Medio Lateral Oblique (MLO) view shown in **Figure 1.1** and Cranial Caudal (CC) view shown in **Figure 1.2**. The MLO view depicts the breast at an angle and the CC view is a projection of the breast from above. Both views are used because each view has crucial information that the other view lacks. In this research, we utilized only the MLO.

The patient's breast is placed on a flat surface while the paddle compresses the breast. The x-ray machine then emits a small dosage of x-ray radiation. The dark regions of the mammogram represent the places where the x-rays reached the detector, while the light places are the places where the breast tissue absorbed the x-rays [6]. It is common

practice to take two different views of each breast. The Medio Lateral Oblique (MLO) view is taken at an 45° angle with respect to the side of the breast whereas the Cranial Caudal (CC) view is obtained by positioning the source vertically above the breast and then the detector below. Both views are used to aid in breast cancer detection [7].

### 1.2.3 Benefits of Mammography

As mentioned in Section 1.1, early detection of breast cancer almost guarantees survival. Since the breasts are only exposed to low dosage of x-ray radiation during the mammogram, it is unlikely that the patient will develop breast cancer from excess exposure to ionizing radiation. There is a small risk, however, and that is why the ACR and NCCN recommend annual rather than more frequent screenings [8, 9].

## 1.3 Information Gathered From the Mammogram

A distinction can be made between breasts in terms of tissue density. Breast tissue density can be categorized as one of four types. According to the National Cancer Institute, the densities are defined (from lowest density to highest density) as follows:

- The breasts are almost entirely fatty
- There are scattered areas of dense glandular tissue and fibrous connective tissue (together known as fibroglandular density)
- The breasts are heterogeneously dense, which means they have more of these areas of fibroglandular density. This may make it hard to see small masses in the breast tissue on a mammogram.

- The breasts are extremely dense, which makes it hard to see tumors in the breast tissue on a mammogram [10].

## 1.4 Computer Aided Detection/Diagnostic

In an effort to reduce healthcare costs, many institutions are implementing digital screening Computer Aided Detection/Diagnostic (CAD) software. CAD is utilized by around 92% of all the digital screening facilities in the US in 2016 [11]. As mentioned in Section 1.1, the false positive rates for CAD is higher than would be ideal. In order to augment CAD methodologies, one must have an understanding of the breast microenvironment. In this research, four metrics were used to analyze the full breast area and identify differences in breast microenvironment for normal and cancer cases, normal and benign cases, and normal and tumorous cases. In this context, tumorous refers to both benign and cancerous cases.

## 1.5 Review of Marin et al. (2017)

In previous research, identifying tumors in a mammogram has often the main tool for cancer detection, but research done by Marin et al (2017) [12] hypothesized that there is a correlation between breast microenvironment and the development of breast cancer. If the tissue structure of the breast is not disrupted, the patient is not likely to have tumorous breast tissue. If the tissue structure is disrupted, that is, if the tissue exhibits a high level of entropy, then the breast tissue will enable tumor growth. In this research, breast tissue is categorized into three groups: fatty, disrupted, and dense. The Hurst Exponent (H) is used to quantify the roughness of breast tissue and Marin et al (2017)

[12], and this information is used to categorize breast tissue in this research as well. Fatty tissue has an H value between 0 and 0.45, disrupted tissue has an H value between 0.45 and 0.55, and dense tissue has an H value greater than 0.55. When H is 0.5, this means that the tissue growth is random, which is associated with unhealthy tissue growth. The ranges of H have already been predetermined in Marin et al (2017) [12]. Examples in Chapter 2 will illustrate that cancerous cases typically exhibit more disrupted tissue than normal cases.

## 1.6 Brief Description of Image Analysis Methodology

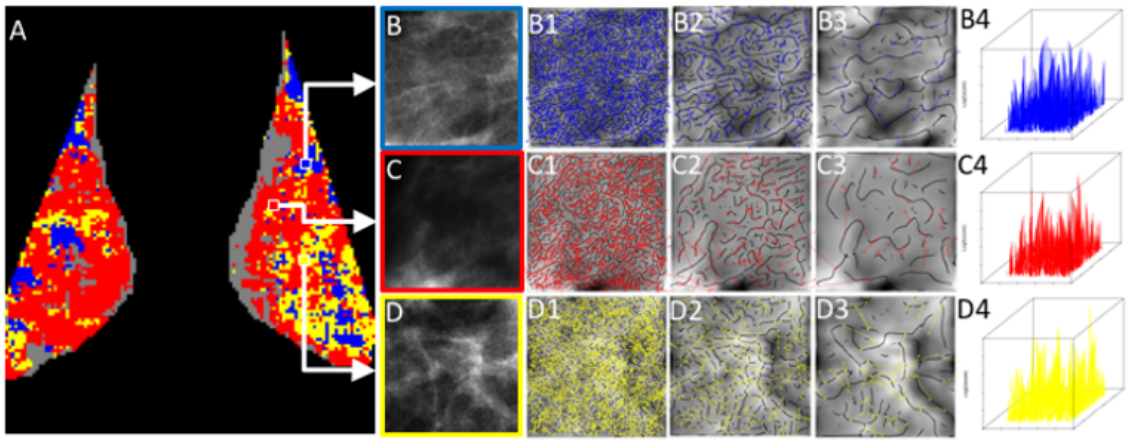


**Figure 1.3:** The above figure is an RGB file, which provides an example of the 2D Wavelet Transform Modulus Maxima (WTMM) method and the Window Sliding method applied to the two mammograms. The blue represents fatty tissue, with $H < 0.45$, red represents dense tissue, with $H > 0.55$, and yellow represents disrupted tissue, with $0.45 < H < 0.55$. Small size scale is represented in B1, C1, and D1. Medium size scale is represented in B2, C2, and D2. Large size scale is represented in B3, C3, and D3. The Wavelet Skeleton of results from all size scales are represented in B4, C4, and D4. The above figure is taken from Marin et al (2017) [12].

By understanding breast microenvironment and tissue disruption, CAD augmentation is possible via the 2D Wavelet Transform Modulus Maxima (WTMM) method and Window Sliding method, implemented in Marin et al. paper (2017) [12]. The 2D WTMM is used to identify the H value of any given region and the Window Sliding method is used to analyze the image at one subregion at a time. The mammograms used in this research were analyzed by the 2D WTMM method in conjunction with the Window Sliding Method.

The 2D WTMM uses the Gaussian function to examine where the gradient or slope, is maximal on the mammogram. This process is done for all size scales and all the results are combined into a 3D plot called the wavelet skeleton. The results of the wavelet skeleton are fed into a partition function. The output of the partition function is the Hurst exponent. The Hurst exponent, as mentioned in Section 1.5, classifies the type of tissue present in the given subregion. Hurst exponents can be calculated only if the subregion analyzed is monofractal in nature, meaning the roughness is homogeneous throughout the whole subregion. Otherwise, if the subregion analyzed is not monofractal, the sub region would be colored gray, as shown in Figure 1.3 [12].

The mammograms are analyzed from left to right and top to bottom. Each mammogram is sectioned off into 360 * 360 pixel subregions, and a 256 * 256 pixel subregion is placed at the center of a 360 x 360 subregion. The reason for having 360 * 360 pixel area subregion with the 256 * 256 pixel area subregion in the center is to avoid edge effects from the 2D WTMM method [12]. The 2D WTMM method is used to analyze the 256 * 256 pixel subregion. After this is done, the Window Sliding method moves the 256 * 256 region to the next adjacent subregion or to the next row and

employs the 2D WTMM method again. This process is utilized until all subregions of the breast have been analyzed.

## 1.7 Research Question

As mentioned in Section 1.5, Marin et al. (2017) [12] established a link between loss of normal tissue structure and tumor growth. In this research, we hope to learn more about the breast microenvironment by examining other regions of the breast such as the 1024 x 1024 pixel area nipple region (nipple region) and the area excluding the 1024 x 1024 pixel area nipple region (exclusion zone). Further, we hope to utilize additional statistical tools to help construct models to better understand the nuances of the breast microenvironment as well as differentiate between normal and tumorous breasts.

## 1.8 Paper Overview

In Chapter 2, the primary focus is to determine whether the metrics in Marin et al (2017) [12] can be used to determine significant differences between the four subgroups, normal, benign, cancerous, and tumorous, for the nipple region and the exclusion zone. For this research, the significance level was $p<0.05$. The Wilcoxon Rank Sum Test was utilized to determine whether significant differences exist between the four subgroups. In Chapter 3, we outline the Logistic modeling methodology and the interpretation of $\beta$'s and how it contributes to our understanding of breast microenvironment. The $\beta$'s are the slope of the logistic model. Furthermore, the methodology behind model selection is also outlined. The comparison done in Chapter 3 augments our understanding of breast

microenvironment by exploring how other variables such as density score and age contribute to distinctions between tumorous and normal cases.

## 2. ANALYZED RESULTS USING NON-PARAMETRIC STATISTICS

### 2.1 Data: (For all MLO cases analyzed)

In this research, a total of 127 cases were analyzed. The cases were categorized follows: 43 cases from patients with normal tissue, 35 cases from patients with benign tumors, and 49 cases from patients with cancerous tumors. These mammograms were used for the full breast analysis, nipple region analysis, and the exclusion zone analysis.

## 2.2 Left versus Right Breast Image Registration

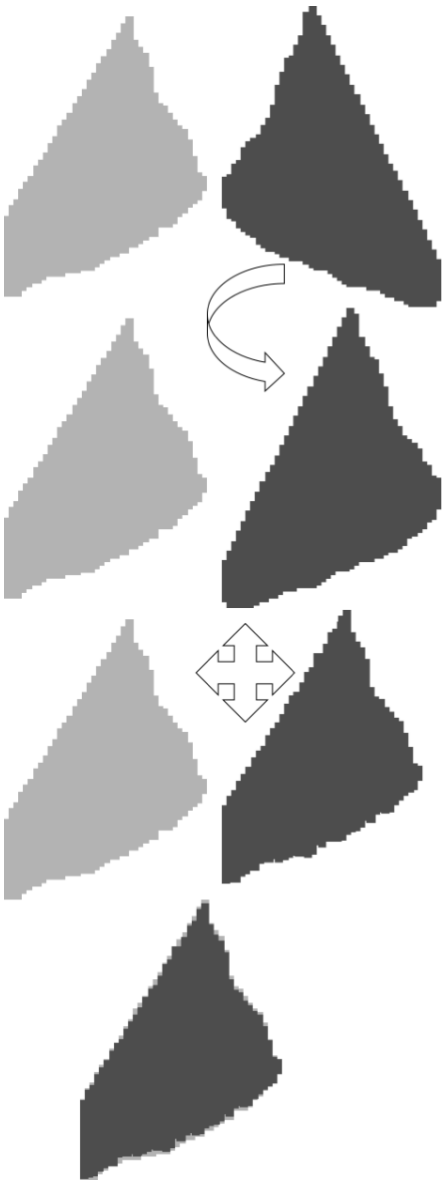

**Figure 2.1:** Image registration of a breast to the opposite breast. The above figure is taken from Marin et al (2017) [12].

The registration process is done so the transitioned squares metric for the full breast can be performed. The transitioned squares metric will be outlined in Section 2.3.1. First, both the mammograms for the left breast, shown in dark gray in Figure 2.1, and the right breast, shown in light gray, must be obtained. Then, the left breast

mammogram must be flipped horizontally. After this is done, the left breast mammogram is fitted to that of the right breast so they are roughly the same size, in terms of both area and perimeter [12]. Once the registration process was completed on the mask files, the RGB image files were then transformed to match the corresponding mask. These RGB image files represent the different types of breast tissue as categorized in Section 1.5, where red represents dense tissue, green represents disrupted tissue, blue represent fatty tissue, and gray represent tissue which was unable to be classified properly.

## 2.3 Metrics Used

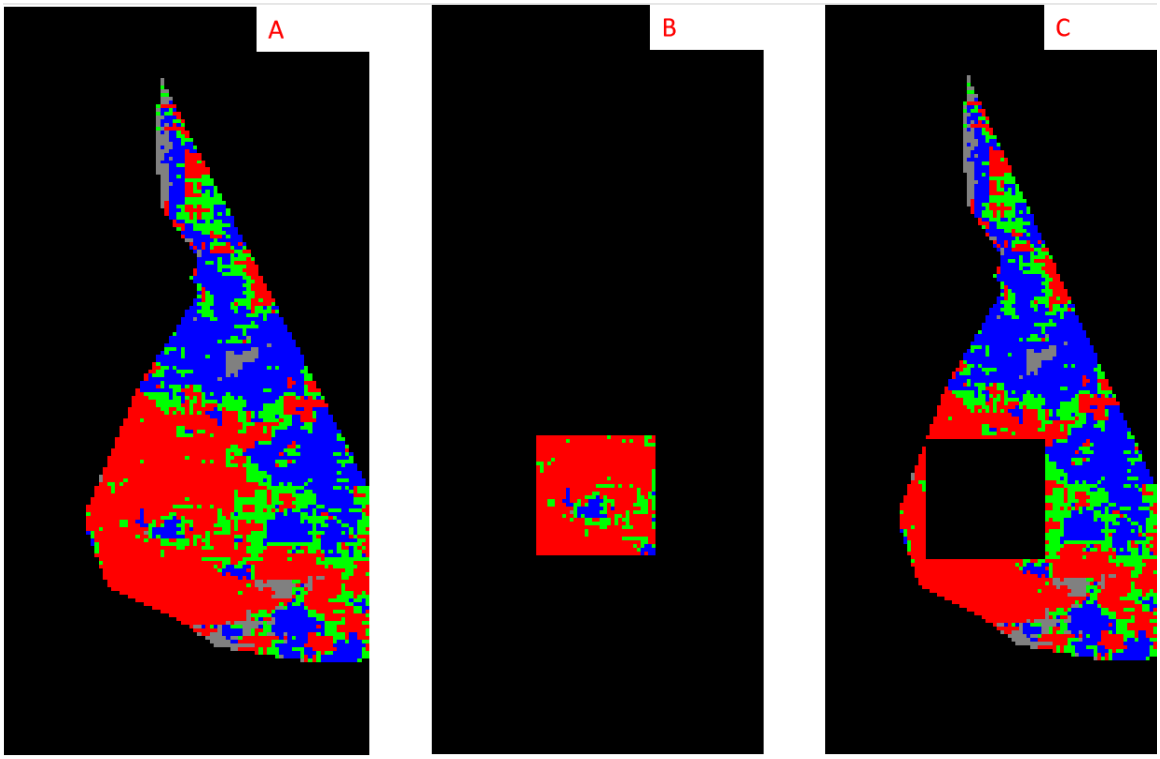

**Figure 2.2 Is a** representation of the Full Breast [$\lambda$] analysis (**A**), Nipple Region [$\xi$] analysis (**B**), and Exclusion Zone [$\phi$] analysis (**C**). We first use the RGB file we have as a result of the 2D WTMM and window sliding analysis to run the metrics we have developed for the full breast analysis.

2.3.1 Full Breast [$\lambda$ Analysis (MLO)

The full breast analysis, which will be referred to as the $\lambda$ analysis, examines the full breast area of the mammogram. An example of this can be seen in Figure 2.2A. The research performed in the Marin et al. (2017) [12] paper identified computational metrics which were utilized for the $\lambda$ analysis. To use these metrics, masks of full breasts were drawn for every single MLO view mammogram. From here, the respective masks and RGB files were opened in RStudio, where each computational metric for the $\lambda$ analysis was computed to compare benign and normal breast tissue, cancerous and normal breast tissue, and tumorous and normal breast tissue.

The following computational metrics were used: the sum of yellow squares, the sum of yellow clusters, the sum of transitioned squares, and the combination metric. The combination metric as established in March et al (2017) is computed as

$$\Xi(\alpha) = \frac{\sum yellow\ squares * \sum yellow\ clusters * \sum transitioned\ squares}{(\sum All\ squares)^{a}}, \qquad \text{[Equation 2.1]}$$

with a= 1.5 [12]. For the sum of yellow clusters metric, clusters are defined as 45 or more connected squares. For the sum of transitioned squares metric, the computer examines the same pixel location for the left and right breast of the same case and compares whether both pixels are the same color or not.

After each metric was performed on the RGB files, the Wilcoxon Rank Test (Two-tailed) was applied to compare the data sets from the four subgroups: normal, benign, cancerous, and tumorous cases. The Wilcoxon Rank Test (Two-tailed) was best suited for this dataset since the distribution of the data was unknown [12]. Afterwards, to determine whether benign, cancerous, and tumorous cases exhibited certain patterns that normal cases do not, the Wilcoxon Rank sum test (Upper-tail) was applied between groups to the data.

2.3.2 Nipple Region [$\xi$] Analysis (MLO)

The nipple region analysis, which will be referred to as the $\xi$ analysis, examines the nipple region area of the mammogram depicted in Figure 2.2B. The motivation behind utilizing the 1024 * 1024 region was to replicate experiments performed by other academic institutions. After masks of 1024 x 1024 pixel area were drawn and placed as close to the nipple region inside the breast as possible, the respective masks and RGB files were opened in RStudio. Each computational metric for the $\xi$ analysis was computed to compare normal and benign breast tissue, normal and cancerous breast tissue, and normal and tumorous breast tissue. The metrics performed were the number of yellow squares behind the nipple, the number of yellow clusters behind the nipple (cluster sizes from 1 and greater), and the number of transitioned squares.

The size of the clusters was determined by computing the p-values for the clusters of sizes 1 and greater to clusters of size 50 and greater. The p-values were calculated by employing the Wilcoxon Rank Sum Test (Two-tailed) to compare the normal and cancerous data sets. The p-values for normal and cancerous tissue were used for this comparison since this is how the computational metrics were calibrated in Marin et al

(2017) [12]. Out of the fifty p-values calculated, the smallest p-value was selected and the cluster size at this p-value were used, which in this experiment was cluster size 1 and greater.

The transitioned squares metric for the $\xi$analysis is different from the lambda analysis because the images were not registered. Instead, the area behind the nipple of one breast was flipped horizontally and both 1024 x 1024 areas were aligned. The combination metric was not attempted since the sum of yellow squares and sum of transitioned square metric returned p-values greater than 0.05 for the normal versus benign, normal versus cancerous, and normal versus tumorous comparison. The Wilcoxon Rank Sum Test (Two-tailed) was used to determine if each metric could be used to differentiate between the four subgroups. Finally, the Wilcoxon Rank Sum Test (Upper-tail) was used between groups to determine if there was a significant difference between normal and benign, normal and cancerous, and normal and tumorous cases.

### 2.3.3 Exclusion Zone [$\phi$] Analysis (MLO)

The exclusion zone analysis, which will be referred to as the $\phi$ analysis, examines the region of the mammogram depicted in Figure 2.2C. First, masks for the $\xi$analysis and the full masks for each case were opened, because the exclusion zone is the area of full mask minus the area of the masks used in the $\xi$analysis. Afterwards, the respective RGB files were opened in RStudio, and the RGB file was mapped with the new area. Each computational metric for the $\phi$analysis was tested to determine whether there were significant differences between normal and benign breast tissue, normal and cancerous breast tissue, and normal and tumorous breast tissue.

The metrics performed were the number of yellow squares in the exclusion zone, the number of yellow clusters in the exclusion zone (cluster sizes 45 and greater), and the new combination metric,

$$\chi = \sum yellow\ squares * \sum yellow\ clusters(size\ 45+)$$

[Equation 2.2]

The new combination metric, outlined in Equation 2.2, is the number of yellow squares times the number of yellow clusters. The best size to establish any significant differences between the three subgroups was determined to be 45 and above. The size of the clusters was determined by the process outlined in 2.3.2. After each metric was performed on the RGB files, the same statistical methodology outlined in section 2.3.1 was utilized to analyze the data sets generated by each metric for the four subgroups.

## 2.4 Wilcoxon Rank Sum Test

### 2.4.1 Brief Description

The Wilcoxon Rank Sum Test is the non-parametric equivalent of the paired t-test. The Wilcoxon Rank Sum Test is used because it can be applied to samples of size 10 or greater and it can be performed on samples with unknown distributions. The null hypothesis of the Wilcoxon Rank Sum Test is that there are no differences in means between two data sets.

$$H_0: \Delta = 0$$

[Equation 2.3]

For Equation 2.3, the delta sign is the difference in means for the two data sets that are being compared. The W statistic is defined as

$$W = \sum_{j=1}^{n} R_j$$

[Equation 2.4]

If one were to calculate the W statistic by hand, the W statistic is calculated by entering data sets of both subgroups into the same vector and ordering the elements from least to greatest. The $R_j$ value of each element is determined by ordering the elements in the vector from smallest to largest. The ranks associated with each data point are then summed. The letter j denotes the jth data point of the dataset and n is the total number of data points in one dataset [13]. For this research, we consider $p<0.05$ to be significant.

2.4.2 Assumptions for the Wilcoxon Ranks Sum Test

Assumption 1:

In this research, it is assumed that there is a difference between the normal and benign, cancerous, and tumorous breast tissue data sets.

Assumption 2:

The data points from each dataset are randomly selected from the total population.

Assumption 3:

Each data point comes from the total population.

A list of assumptions and how they were derived can be seen in reference [13].

2.4.3 Wilcoxon Rank Sum Test (Two-tailed)

For the two-tailed test, the null hypothesis is the same as the one outlined in Section 2.4.1. The alternative hypothesis for the two-tailed test states that there is a difference in means for the two data sets compared.

$$H_A: \Delta \neq 0$$

[Equation 2.5]

In the equations below, $\alpha_2$is the lower tail probability, whereas $\alpha_1$is the upper tail probability. Also the table of W calculations can be consulted in reference [13] if one wished to calculate W by hand. In the equations given in the remainder of this section, n and m are the sizes of the two data sets that are being compared.

If the null hypothesis were to be rejected, the following condition needs to hold true:

$$W \geq w(\alpha_2, m, n) \text{ or } W \leq [n(m+n+1) - w(\alpha_1, m, n)]$$

[Equation 2.6]

If the null hypothesis cannot be rejected, then the following condition must hold true:

$$[n(m+n+1) - w(\alpha_1, m, n)] \leq W \leq w(\alpha_2, m, n).$$

[Equation 2.7]

[13]

2.4.4 Wilcoxon Rank Sum Test (Upper-tail)

For the upper-tail test, the null hypothesis is the same as the one outlined in Section 2.4.1. The alternative hypothesis for the upper-tail test states that one of the two data sets being compared, has a higher mean than the other in the comparison.

$$H_A: \Delta > 0$$

[Equation 2.8]

If the null hypothesis were to be rejected, then the following condition must hold true:

$$W \geq w(\alpha, m, n)$$

[Equation 2.9]

If the null hypothesis cannot be rejected, then the following condition must hold true:

$$W < w(\alpha, m, n).$$

[Equation 2.10]

[13]

2.4.5 Large Sample Approximation

In this research, the Large Sample Approximation is utilized because all data sets have more than 10 elements.

$$W^* = \frac{W - E_0(W)}{[var_0(W)]^{1/2}} = \frac{W - [\frac{n(m+n+1)}{2}]}{[\frac{mn(m+n+1)}{12}]^{1/2}}$$

[Equation 2.11]

The hypothesis testing mechanism utilizes the normal theory approximation. In the normal theory approximation below, $z_\alpha$means the z-score at the given level of significance, $\alpha$.

*Normal Theory Approximation:*

If the null hypothesis were to be rejected, then the following condition must hold true:

$$W^* \geq z_\alpha$$

[Equation 2.12]

If the null hypothesis cannot be rejected, then the following condition must hold true:

$$W^* < z_\alpha.$$

[Equation 2.13]

[13]

2.4.6 Treatment of Ties

When there are ties in rank for different elements in the same vectors, we utilize the treatment of ties to compute the W statistic,

$$Var_0(W) = \frac{mn}{12}\left[m + n + 1 - \frac{\sum_{j=1}^{g} t_j\left(t^2{}_j - 1\right)}{(m+n)(m+n+1)}\right]$$

[Equation 2.14]

In the above formula, g is the total number of tied groups and $t_j$ is the size of the groups [13]. Equations 2.3 – Equations 2.14 are taken from [13].

## 2.5 Wilcoxon Rank Sum [Two-tailed] Test Analysis of all three regions

In this research, our hypothesis was that the breast microenvironment for normal breast tissue is different from benign, cancerous, and tumorous breast tissue. If $p<0.05$, one can reject the null hypothesis. The p-values yielded in Tables 2.1 to 2.4 are produced by utilizing the Wilcoxon Rank Sum (Two-sided) Test. The only conclusion that can be arrived at when utilizing the Wilcoxon Rank Sum (Two-sided) Test is whether there are any differences between two data sets.

### 2.5.1 $\lambda$ Analysis (MLO)

The results for the sum of yellow squares metric from Table 2.1 show the p-value for normal and cancer is 0.01692, the p-value for normal and benign is 0.0002238, the p-value for benign and cancer is 0.1408, and the p-value for tumor and normal is 0.0008336. The p-values for normal versus cancer, normal versus benign, and normal

versus tumorous cases suggests there is a difference between normal versus, benign cases, normal versus cancerous cases, and normal versus tumorous cases. As shown in the middle box plot of Figure 2.6, the yellow squares medians for normal and tumorous cases are different from each other. Since the p-values for normal versus benign, normal versus cancer, and normal versus tumor group are small, we decided to utilize the sum of yellow squares as a metric to help differentiate between normal versus benign, normal versus cancer, and normal versus tumor group.

The results for the sum of yellow clusters metric from Table 2.2 show the p-value for normal and cancer is 0.009061, the p-value for normal and benign is 0.002951, the p-value for benign and cancer is 0.4313, and the p-value for tumor and normal is 0.001469. The p-values for normal versus cancer, normal versus benign, and normal versus tumorous cases are smaller than 0.05, this suggests that there is a difference between normal versus, benign cases, normal versus cancerous cases, and normal versus tumorous cases. As shown in the middle box plot of Figure 2.7, the yellow clusters medians for normal and tumorous cases are different from each other. Since the p-values for normal versus benign, normal versus cancer, and normal versus tumor group are small, we decided to utilize the sum of yellow clusters as a metric to help differentiate between normal versus benign, normal versus cancer, and normal versus tumor group.

The results for the sum of transitioned squares metric from Table 2.3 show the p-value for normal versus cancer is 0.02471, the p-value for normal and benign is 0.0008767, the p-value for benign and cancer is 0.1533, and the p-value for tumor and normal is 0.001665. The p-values for normal versus cancer, normal versus benign, and normal versus tumorous suggests that there is a difference between normal versus, benign

cases, normal versus cancerous cases, and normal versus tumorous cases. As shown in the left box plot of Figure 2.8, the transitioned squares medians for normal and tumorous cases are different from each other. Since the p-values for normal versus benign, normal versus cancer, and normal versus tumor group are small, we decided to utilize the sum of transitioned squares as a metric to help differentiate between normal versus benign, normal versus cancer, and normal versus tumor group.

The results for the combination metric from Table 2.4 show the p-value for normal and cancer is 0.003425, the p-value for normal and benign is 0.001647, the p-value for benign and cancer is 0.5854, and the p-value for tumor and normal is 0.0005389. The p-values for normal versus cancer, normal versus benign, and normal versus tumorous cases suggests that there is a difference between normal versus, benign cases, normal versus cancerous cases, and normal versus tumorous cases. As shown in the left box plot of Figure 2.9, the combination metric medians for normal and tumorous cases are different from each other. Since the p-values for normal versus benign, normal versus cancer, and normal versus tumor group are small, we decided to utilize the combination score as a metric to help differentiate between normal versus benign, normal versus cancer, and normal versus tumor group.

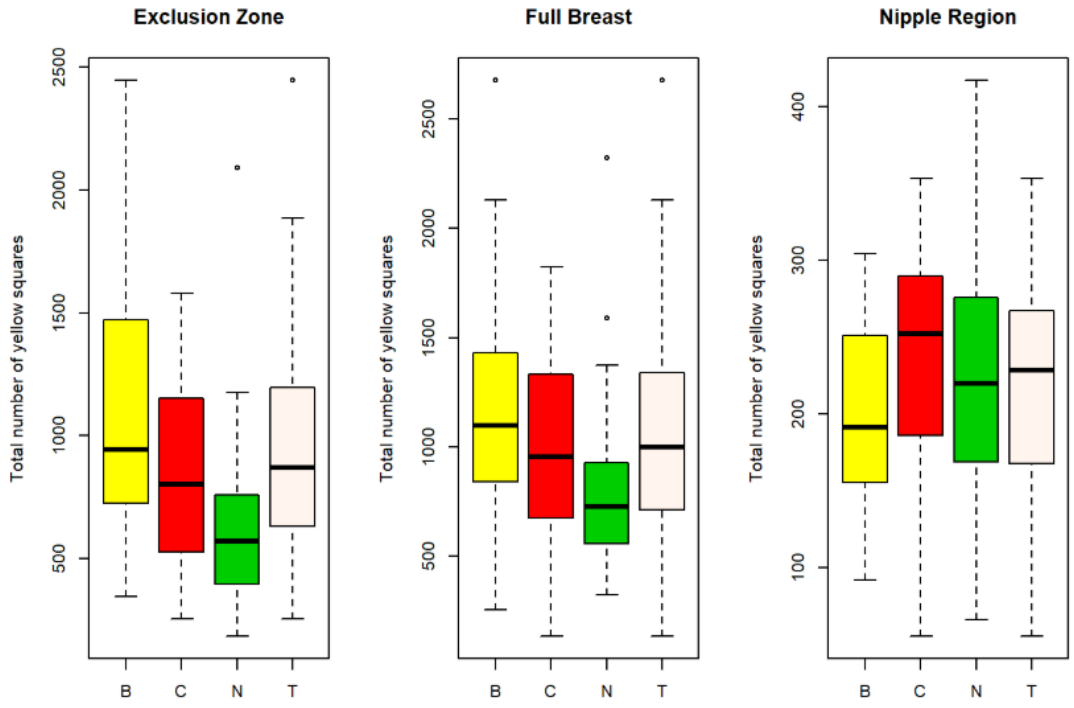


Figure 2.6: Box plot of results generated by the sum of yellow squares metric. The yellow represents the benign cases, the red represents the cancer cases, green represents the normal cases, and the seashell color represents the tumorous cases (benign and cancer).

Table 2.1: Table of p-values for the yellow square metric

| | Full Breast | Nipple Region | Exclusion zone |
|---|---|---|---|
| Benign versus Normal | 0.0002238 | 0.1052 | 9.512e-06 |
| Cancer versus Normal | 0.01692 | 0.5475 | 0.002083 |
| Benign versus Cancer | 0.1408 | 0.01978 | 0.08598 |
| Tumor versus Normal | 0.0008336 | 0.6699 | 3.25e-05 |

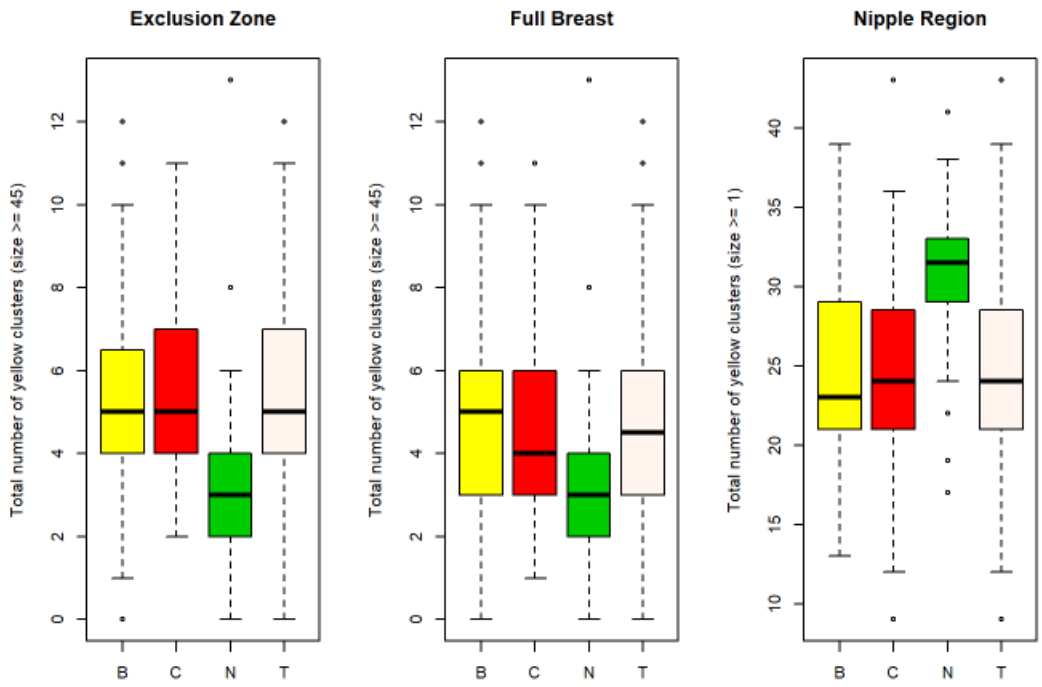


Figure 2.7: Box plot of results generated by the sum of yellow clusters metric. The yellow represents the benign cases, the red represents the cancerous cases, green represents the normal cases, and the seashell color represents the tumorous cases (benign and cancer).

Table 2.2: Table of p-values for the yellow clusters metric

| | Full Breast | Nipple Region | Exclusion zone |
|---|---|---|---|
| Benign versus Normal | 0.002951 | 7.79e-04 | 1.58e-03 |
| Cancer versus Normal | 0.009061 | 1.53e-05 | 2.52e-03 |
| Benign versus Cancer | 0.4313 | 8.74e-01 | 4.85e-01 |
| Tumor Versus Normal | 0.01461 | 4.77e-06 | 3.54e-04 |

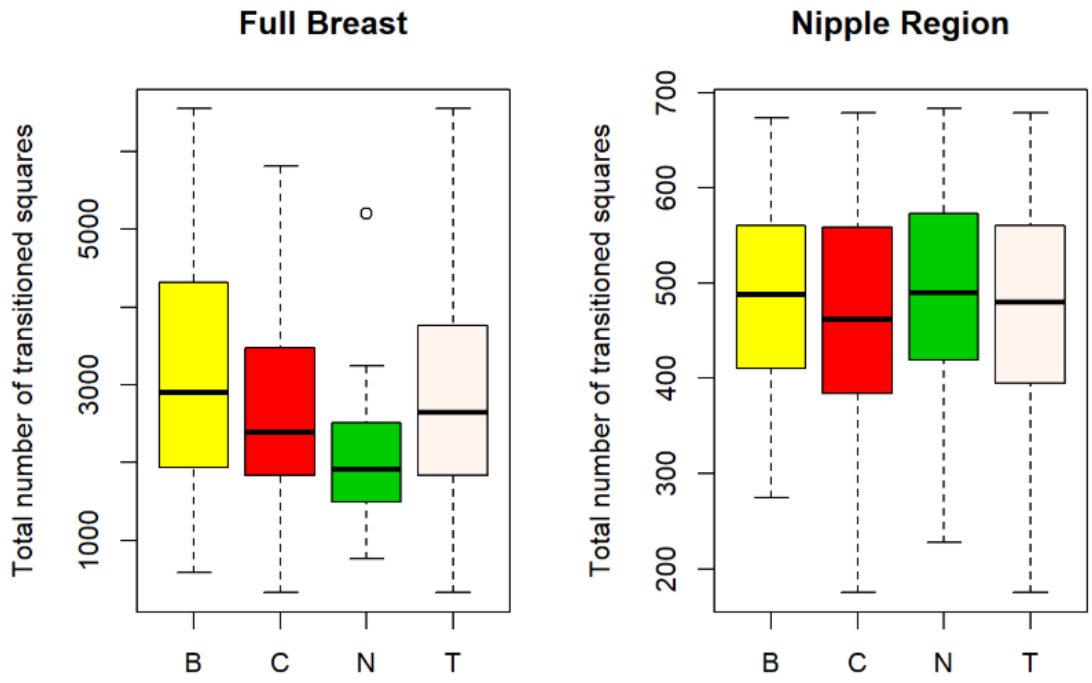


Figure 2.8: Box plot of results generated by the sum of transitioned squares metric. The yellow represents the benign cases, the red represents the cancer cases, green represents the normal cases, and the seashell color represents the tumorous cases (benign and cancer).

Table 2.3: Table of p-values for the transitioned squares metric

| | Full Breast | Nipple Region |
|---|---|---|
| Benign versus Normal | 0.0008767 | 0.7895 |
| Cancer versus Normal | 0.02471 | 0.2829 |
| Benign versus Cancer | 0.1533 | 0.4775 |
| Tumor Versus Normal | 0.001665 | 0.3965 |

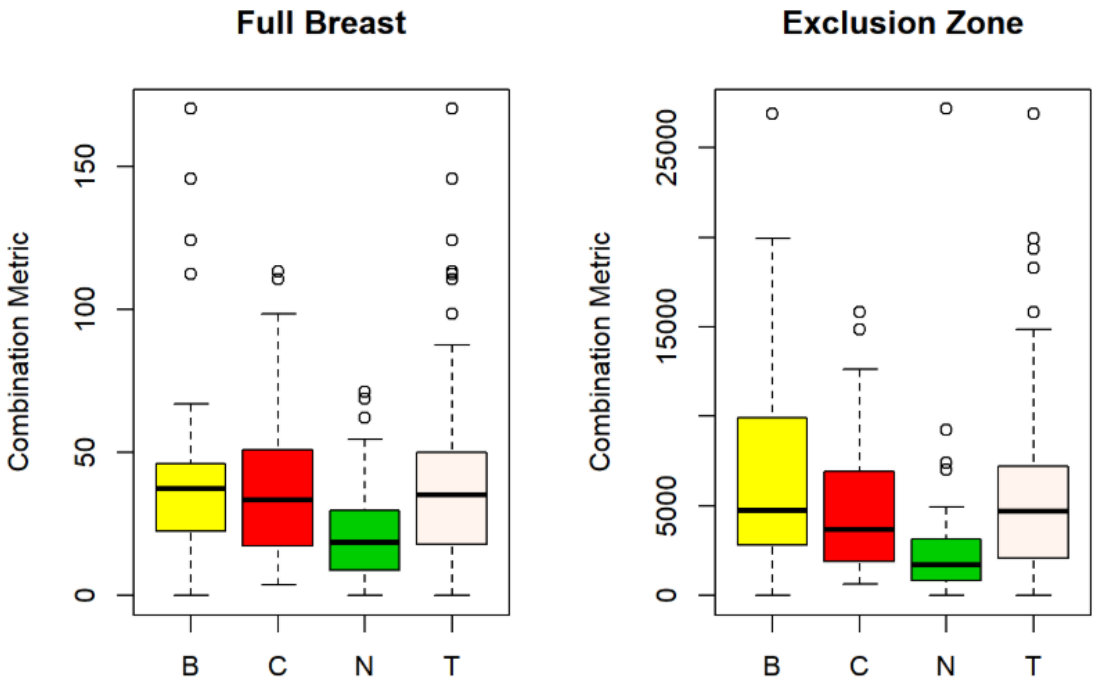


Figure 2.9: Box plot for results generated by the combination metric. The yellow represents the benign cases, the red represents the cancer cases, green represents the normal cases, and the seashell color represents the tumorous cases (benign and cancer).

Table 2.4: Table of p-values for the combination metric

| | Full Breast | Exclusion zone |
|---|---|---|
| Benign versus Normal | 0.001647 | 0.0001282 |
| Cancer versus Normal | 0.003425 | 0.001393 |
| Benign versus Cancer | 0.5854 | 0.2383 |
| Tumor Versus Normal | 0.0005389 | 6.427e-05 |

2.5.2 ξ Analysis (MLO)

The results for the sum of all yellow squares metric from Table 2.1 show the p-value for normal and cancer is 0.5475, the p-value for normal and benign is 0.1052, the p-value for benign and cancer is 0.01978, and the p-value for tumor and normal is 0.6699. The p-values for normal versus cancer, normal versus benign, and normal versus tumorous cases suggests that there is no difference between normal versus benign cases, normal versus cancerous cases, and normal versus tumorous cases. As shown in the right box plot of Figure 2.6, the yellow squares medians for normal and tumorous cases are not seemingly different from each other. Since the p-values for normal versus benign, normal versus cancer, and normal versus tumor group are relatively large, we decided to not utilize the sum of yellow squares as a metric to help differentiate between normal versus benign, normal versus cancer, and normal versus tumor group.

The results for the sum of all yellow clusters metric from Table 2.2 show the p-value for normal and cancer is 1.53e-05, the p-value for normal and benign is 7.79e-04, the p-value for benign and cancer is 8.47e-01, and the p-value for tumor and normal is 4.77e-06. The p-values for normal versus cancer, normal versus benign, and normal versus tumorous cases suggests that there is a difference between normal versus, benign cases, normal versus cancerous cases, and normal versus tumorous cases. As shown in the right box plot of Figure 2.7, the yellow clusters medians for normal and tumorous cases are not seemingly different from each other. Since the p-values for normal versus benign, normal versus cancer, and normal versus tumor group are small, we decided to utilize this as a metric to help differentiate between normal versus benign, normal versus cancer, and normal versus tumor group.

The results for the sum of transitioned squares metric from Table 2.3 show the p-value for normal and cancer is 0.2829, the p-value for normal and benign is 0.7895, the p-value for benign and cancer is 0.4775, and the p-value for tumor and normal is 0.3965. The p-values for normal versus cancer, normal versus benign, and normal versus tumorous cases suggests that there is no difference between normal versus, benign cases, normal versus cancerous cases, and normal versus tumorous cases. As shown in the right box plot of Figure 2.8, the transitioned squares medians for normal and tumorous cases are not seemingly different from each other. Since the p-values for normal versus benign, normal versus cancer, and normal versus tumor group are relatively big, we decided to not utilize this as a metric to help differentiate between normal versus benign, normal versus cancer, and normal versus tumor group.

### 2.5.3 $\phi$Analysis (MLO)

The results for the sum of yellow squares metric from Table 2.1 show the p-value for normal and cancer is 0.002083, the p-value for normal and benign is 9.512e-06, the p-value for benign and cancer is 0.08598, and the p-value for tumor and normal is 3.25e-05. The p-values for normal versus cancer, normal versus benign, and normal versus tumorous cases suggests that there is a difference between normal versus, benign cases, normal versus cancerous cases, and normal versus tumorous cases. As shown in the left box plot of Figure 2.6, the yellow squares medians for normal and tumorous cases are different from each other. Since the p-values for normal versus benign, normal versus cancer, and normal versus tumor group are small, we decided to utilize this as a metric to

help differentiate between normal versus benign, normal versus cancer, and normal versus tumor group.

The results for the sum of yellow clusters metric from Table 2.2 show the p-value for normal and cancer is 2.52e-03, the p-value for normal and benign is 1.58e-03, the p-value for benign and cancer is 4.85e-01, and the p-value for tumor and normal is 3.54e-04. The p-values for normal versus cancer, normal versus benign, and normal versus tumorous cases suggests that there is a difference between normal versus, benign cases, normal versus cancerous cases, and normal versus tumorous cases. As shown in the left box plot of Figure 2.7, the yellow clusters medians for normal and tumorous cases are different from each other. Since the p-values for normal versus benign, normal versus cancer, and normal versus tumor group are small, we decided to utilize the sum of yellow clusters as a metric to help differentiate between normal versus benign, normal versus cancer, and normal versus tumor group.

The results for the combination metric from Table 2.4 show the p-value for normal and cancer is 0.001393, the p-value for normal and benign is 0.0001282, and the p-value for benign and cancer is 0.2383, and the p-value for tumor and normal is 6.427e-05. The p-values for normal versus cancer, normal versus benign, and normal versus tumorous cases are smaller than 0.05, this suggests that there is a difference between normal versus, benign cases, normal versus cancerous cases, and normal versus tumorous cases. As shown in the right box plot of Figure 2.9, the combination metric medians for normal and tumorous cases are different from each other. Since the p-values for normal versus benign, normal versus cancer, and normal versus tumor group are small, we

decided to utilize this as a metric to help differentiate between normal versus benign, normal versus cancer, and normal versus tumor group.

## 2.6 Wilcoxon Rank Sum [Upper-tail] Test Analysis of All Three Regions

### 2.6.1 $\lambda$Analysis (MLO)

Section 2.5.1 suggests that all the metrics for the $\lambda$ analysis can be used to differentiate benign, cancerous, and tumorous groups from the normal group because the p-values for all of those comparisons are less than 0.05. The question we would like to answer is, what makes the benign, cancerous, and tumor groups different from normal breast groups?

For the sum of yellow squares metric, the median count from Figure 2.6 shows that benign, cancerous, and tumorous breasts have a higher median than normal breasts. The p-value for benign vs normal is 0.0001119, the p-value for cancer and normal is 0.008458, and the p-value for tumor and normal is 0.0004168. Since the p-values are smaller than 0.05, this suggests that benign, cancer, and tumorous groups have a higher yellow square count than the normal group.

The median count for the sum of yellow clusters from Figure 2.7 shows that benign, cancerous, and tumor group have a higher median than normal breasts. After performing the upper-tail test, I found the p-value for benign vs normal is 0.001446, the p-value for cancer and normal is 0.00453, and the p-value for tumor and normal is 0.0007346. Since the p-values are smaller than 0.05, this suggests that benign, cancerous, and tumorous groups, in general, have a higher yellow cluster count than the normal group.

The median count for transitioned squares from Figure 2.8 shows that benign, cancerous, and tumorous groups have a higher median than the normal group. After performing the upper-tail test, I found the p-value for benign vs normal is 0.0004384, the p-value for cancer and normal is 0.01235, and the p-value for tumor and normal is 0.0008324. Since the p-values are smaller than 0.05, this suggests that benign, cancerous, and tumorous groups, in general, have a higher transitioned square count than the normal group.

The median count for the combination metric from Figure 2.9 shows that benign, cancerous, and tumorous groups have a higher median than the normal group. After performing the upper-tail test, I found the p-value for benign vs normal is 0.0008233, the p-value for cancer and normal is 0.001712, and the p-value for tumor and normal is 0.0002694. Since the p-values are smaller than 0.05, this suggests that benign, cancerous, and tumorous groups, in general, have a higher combination metric count than the normal group.

### 2.6.2 $\xi$ Analysis (MLO)

Section 2.5.2 suggests that the yellow squares metric and the transitioned squares metric are ineffective in the $\xi$ analysis because the p-values for those comparisons are greater than 0.05. We need to examine if the tumorous, cancerous, or benign cases will exhibit more clusters than the normal cases.

The median count for yellow clusters from Figure 2.7 shows that benign, cancerous, and tumorous groups have a lower median than the normal group. After performing the upper-tail test, I found the p-value for benign vs normal is 0.0003897, the

p-value for cancer and normal is 7.65e-06, and the p-value for tumor and normal is 2.385e-06. Since the p-values are smaller than 0.05, this suggests that benign, cancerous, and tumorous groups, in general, have a lower yellow cluster count than the normal group.

### 2.6.3 $\phi$ Analysis (MLO)

Section 2.5.3 suggests that for the $\phi$ analysis, all the metrics were able to differentiate between benign, cancerous, and tumorous groups from the normal group because the p-values for those comparisons are less than 0.05.

The median count for yellow squares from Figure 2.6 shows that benign, cancerous, and tumorous groups have a higher median than the normal group. After performing the upper-tail test, I found the p-value for benign vs normal is 4.756e-06, the p-value for cancer and normal is 0.001041, and the p-value for tumor and normal is 1.625e-05. Since the p-values are smaller than 0.05, this suggests that benign, cancerous, and tumorous groups, in general, have a higher yellow cluster count than the normal group.

The median count for yellow clusters from Figure 2.7 shows that benign, cancerous, and tumorous groups have a higher median than the normal group. After performing the upper-tail test, I found the p-value for benign vs normal is 0.0007887, the p-value for cancer and normal is 0.00126, and the p-value for tumor and normal is 0.000177. Since the p-values are smaller than 0.05, this suggests that benign, cancerous, and tumorous groups, in general, have a higher yellow cluster count than the normal group.

The median combination metric count from Figure 2.9 shows that benign, cancerous, and tumorous groups have a higher median than the normal group. After performing the upper-tail test, I found the p-value for benign vs normal is 6.408e-05, the p-value for cancer and normal is 0.0006967, and the p-value for tumor and normal is 3.213e-05. Since the p-values are smaller than 0.05, this suggests that benign, cancerous, and tumorous groups, in general, have a higher combination metric count than the normal group.

# 3. STATISTICS

## 3.1 Logistic Regression Methodology

### 3.1.1 Purpose of the Logistic Model

The logistic modeling methodology was utilized due to its ability to quantify the probability that a predictive variable will correctly identify tumorous breast tissue. One purpose of the logistic model is to answer classification problems. For this research, it is utilized to see if the metrics from sections 2.3.1, 2.3.2, and 2.3.3 will have high odds of differentiating between the tumor and normal groups.

### 3.1.2 Relation Between Probability And Odds (For Logistic Regression With Single Explanatory Variable)

The logit model (linearized logistic model) has two different outputs for its response variables, 0 and 1. The probability P can be calculated for the two response variables. In this research, $\pi(x)$ is the probability that a mammogram in the sample shows evidence of tumorous tissue, Y is the binary response variable tumorous or normal, and X is any given metric, the predictive variable. Then

$$\pi(x) = P(Y = 1|X = x) = 1 - P(Y = 0|X = x)$$

[Equation 3.1]

$$\pi(x) = \frac{exp\,(\alpha+\beta x)}{1+exp\,(\alpha+\beta x)}$$

[Equation 3.2]

In Equation 3.2, $\pi(x)$, Y, and X are the same variables as defined in Equation 3.1. Also, $\alpha$ is the intercept of the logit model, and $\beta$ is the slope of the logit model. The logit model has a linear relationship between the predictive variable and its response variable.

$$logit[\pi(x)] = log\left(\frac{\pi(x)}{1-\pi(x)}\right) = \alpha + \beta x.$$

[Equation 3.3]

Equations 3.1 – 3.3 are taken from [14].

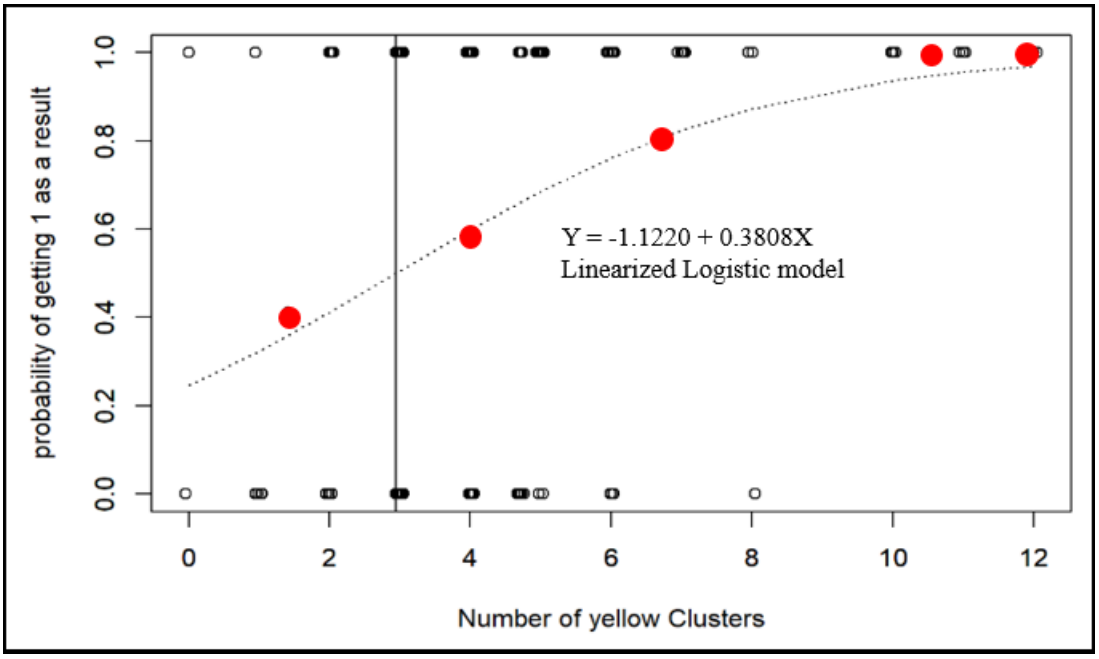


Figure 3.1: Logistic curve example.

The example above depicts the probability that a mammogram shows evidence of tumorous tissue as a function of the number of yellow clusters in the image. The two horizontal scatter plots at y=1 and y=0 show how many images in each cluster category (e.g., the data points along the line x=1 corresponds to the images containing one yellow cluster) showed evidence of the tumorous tissue and how many did not, respectively. The logistic curve with the red dots shows a correlation between the number of yellow

clusters and the likelihood that an image in a given cluster category will depict tumorous tissue. The vertical line in the figure is the median effective level, and that is where the probability that an image in a given cluster category will show tumorous tissue is 50 percent. The value of the median effective level is $\frac{-\alpha}{\beta}$. For the analysis done in Figure 3.1, $\alpha$ is -1.12 and $\beta$ is 0.381. This yields 2.94 for the median effect level.

### 3.1.3 Interpretation of the $\beta's$

In this research, the response variable is tissue type, namely healthy versus tumorous, and the predictive variables are the metrics outlined in sections 2.3.1, 2.3.2, and 2.3.3. As outlined in section 3.1.2, the logit model of logistic regression is linear. Therefore, $\beta_j$ represents the slope of the relationship between the predicted diagnosis and the metric. $e^{\beta j}$ is the odds ratio between X = x+1 and X = x. This means if $\beta_j$ is positive, the odds that an image depicts tumorous tissue increase with each x. If $\beta_j$ is negative, the odds of that an image depicts tumorous tissue decreases with each increase in x. If $\beta_j$ is zero, this means that for the given metric, there is likely no correlation between the chosen metric and the probability that an image will show evidence of tumorous tissue [14].

### 3.1.4 Receiver Operating Characteristic (ROC) Curve

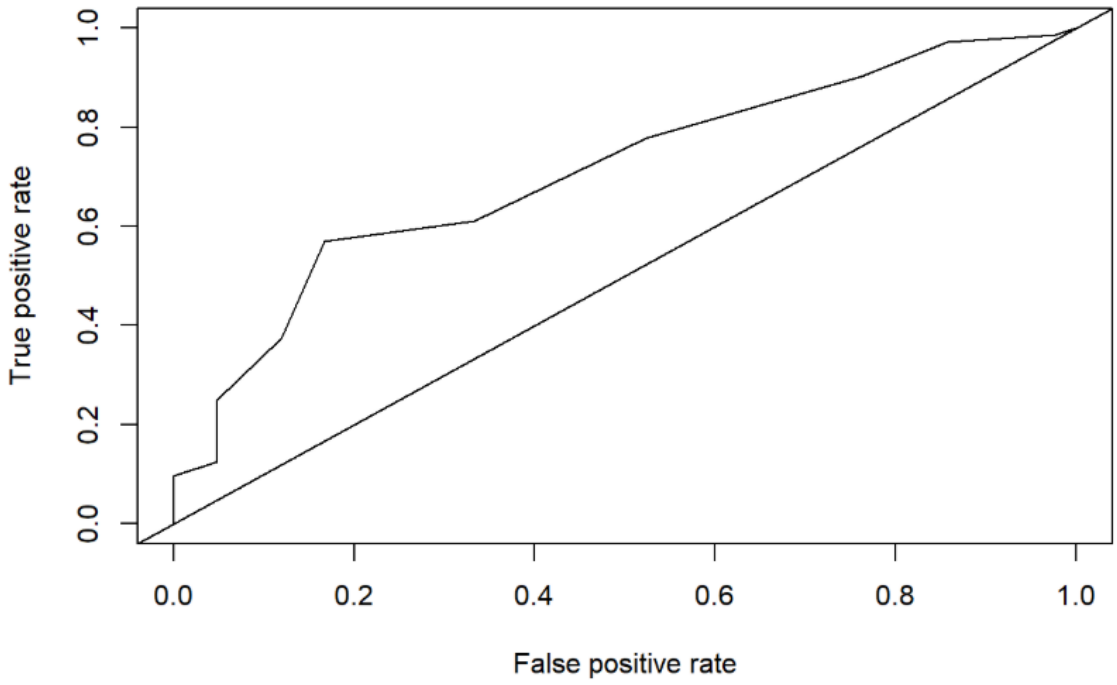


Figure 3.2: Example of ROC Curve.

The receiver operating curve (ROC) plot is a plot of sensitivity versus 1-specificity. Sensitivity is the true positive rate, meaning the percentage of correctly predicted tumorous diagnoses, whereas specificity is the probability that normal tissue will be predicted correctly. The ROC curve, then, is a plot of true positive rate versus false positive rate as shown in Figure 3.2.

$$sensitivity = P(y_{hat} = 1 \mid y = 1)$$

[Equation 3.4]

$$specificity = P(y_{hat} = 0 \mid y = 0)$$

[Equation 3.5]

[Equation 3.5]

In the above equations, $y_{hat}$ is the prediction that the data point has a value of 1, and y is the actual value (0 or 1) of the data points. The ROC curve usually has a concave shape connecting the points (0, 0) and (1, 1). The higher the area under the ROC curve, the better the predictions of the given model. In other words, the ROC curve is used to summarize the predictive power of a given model.

In Figure 3.2, the area under the ROC curve is 0.71, meaning the predictive power of the model is 0.71. The predictive power of the model in Figure 3.2 to determine if the mammogram depicts tumorous tissue is greater than the predictive power of CAD today, which is at 0.63 [15]. Predictive power is the ability to make an accurate assessment of potential future data. The lower triangle represents 50% predictive power, which would be the same as random guessing. If the area under the ROC curve matched the area of the lower triangle, it could mean that the model is no better at predicting than random guessing [14]. Equation 3.4 and 3.5 are taken from [14].

### 3.1.5 Goodness of Fit Tests

Possessing predictive power beyond 0.50 does not mean the model should be utilized. The model should also fit the data well in addition to possessing high predictive power. In order to compute how well the model fits the data, one must utilize goodness of fit tests. There are two common types of the goodness of fit tests, one being the Pearson residual test and the other being the $G^2$ test. The Person residual test is defined as

$$X^2 = \sum_{i=1}^{N} e_i^2.$$

[Equation 3.6]

In Equation 3.6, the $e_i$ stands for Pearson residuals, and the sum of the Pearson residuals provides the Pearson statistic, which is used to determine if a model fits the data poorly. The deviance residual, which is used in the $G^2$ test, is defined as

$$\sqrt{d_i} * sign(y_i - n_i\pi_{i_{hat}})$$

[Equation 3.7]

In Equation 3.7, the $d_i$ stands for deviance, "$y_i$ denote the binomial variate for $n_i$ trials" [14], and "$n_i\pi_{i_{hat}}$ is the fitted number of successes" [14]. Further, the square root of $d_i$ is multiplied by the sign of $y_i - n_i\pi_{i_{hat}}$ to get the deviance residual which is used in the $G^2$ test. Both are used this research. If both goodness of fit tests return p-values greater than 0.05 for the current model, this suggests the current model fits the data. However, if at least one goodness of fit test returns a p-value less than 0.05 for the current model, this suggests the current model fits the data poorly. Should the model fit the data poorly, the model is rejected in favor of an alternative model [14].

### 3.1.6 Logistic Regression with Categorical Predictive Variable

The motivation behind using categorical predictive variable is that the density score may contribute to better fitting of the data in the amended models. A categorical variable is a variable that is not quantitative, for example, a 1 to 10 rating system, or type 1 versus type 2 diabetes. The numerical representation of categorical data is either ordinal or nominal, meaning data with a ranking order (e.g. race times) or data without a ranking order (e.g. colors), but not a representation of quantity.

The method in which categorical variables are analyzed in this research is by the use of indicated variables. The model is in the form of

$$logit(\pi_i) = \alpha + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_I x_I \ ,$$

[Equation 3.6]

where i ={2,3,…,I} because this is the default setting in RStudio, where I is the total number of levels (e.g., density score ranges from 1 to 4; therefore I in this context is 4) [14].

Equation 3.6 is taken from [14].

### 3.1.7 Model Selection via Backwards Elimination

In order to maximize predictive power of models, the complex models are examined for goodness of fit. The most complex model in this context means a model including all possible predictive variables with all interaction terms. The motivation behind the backwards elimination method is to find the most simple model which also fits the data well. The backward elimination algorithm starts with the most complex model and the algorithm eliminates the least significant terms one at a time until goodness of fit is compromised [14].

## 3.2 Logistic Regression Results for Single Predictive Variable

### 3.2.1 $\lambda$ Analysis (MLO)

The model utilizing the yellow square metric as a sole predictive variable has a predictive power of 0.674, or around 67.4%. The mathematical definition of the model is given as:

$Y = -0.6996 + 1.4481X$, where X is the yellow squares metric result divided by 1000.

The odds ratio is 4.255, which is greater than 1. This suggests that for every additional

1000 yellow squares, the odds that an image depicts tumorous tissue increases. The Pearson residuals test has a p-value of 0.356, and the $G^2$test has a p-value of 0.043. Since the model failed to pass at least one goodness of fit test, this indicates that the model utilizing the sum of yellow squares divided by 1000 is not suitable to gauge whether an image shows tumorous tissue.

The model utilizing the results from the yellow clusters metric has a predictive power of 0.661, or around 66.1%. The mathematical definition of the model is given as: $Y = -0.3109 + 0.2339X$, where X is the yellow clusters result. The odds ratio is 1.2635, which is greater than 1. This suggests that for each additional yellow cluster, the probability that an image depicts tumorous tissue increases. The Pearson residuals test has a p-value of 0.293, and the $G^2$test has a p-value of 0.035. Since the model failed to pass at least one goodness of fit test, this indicates that the model utilizing the sum of yellow clusters is not suitable to gauge whether an image shows tumorous tissue.

The model utilizing only the results from the transitioned square metric has a predictive power of 0.672, or around 67.2%. The mathematical definition of the model is given as:

$Y = -0.7822 + 0.5899X$, where X is the transitioned squares metric results divided by 1000.

The odds ratio is 1.8038, which is greater than 1. This suggests that for every additional 1000 transitioned squares, the probability that an image depicts tumorous tissue increases. The Pearson residuals test has a p-value of 0.526, and the $G^2$test has a p-value of 0.057. Since the model passed both goodness of fit tests, this indicates that the model

utilizing the sum of transitioned squares divided by 1000 is suitable to gauge whether an image shows tumorous tissue.

The model utilizing the results from the combination metric has a predictive power of 0.677, or around 67.7%. The mathematical definition of the model is given as: $Y = -0.2627 + 2.9975X$, where X is the combination metric result divided by 100. The odds ratio is 20.035, which is greater than 1. This suggests for each additional 100 combination metric count, the probability that an image depicts tumorous tissue increases. The Pearson residuals test has a p-value of 0.351, and the $G^2$test has a p-value of 0.056. Since the model passed both goodness of fit tests, this indicates that the model using the combination metric is suitable to gauge whether an image shows tumorous tissue.

### 3.2.2 $\xi$ Analysis (MLO)

The model utilizing the results from the yellow square metric has a predictive power of 0.519, or around 51.9%. The mathematical definition of the model is given as: $Y = 0.9736 - 2.1196X$, where X is the yellow squares metric result divided by 1000. The odds ratio is 0.1201, which is less than 1. This suggests that for each additional 1000 yellow squares, the probability that an image depicts tumorous tissue decreases. The result shows that the yellow squares metric is not much better at differentiating tumorous breasts than random assignment. The Pearson residuals test has a p-value of 0.429, and the $G^2$test has a p-value of 0.009. Since the model failed to pass at least one goodness of fit test, this indicates that the model using the sum of yellow squares divided by 1000 is not suitable to gauge whether an image shows tumorous tissue.

The model utilizing the results from the sum of yellow clusters metric has a predictive power of 0.756, or around 75.6%. The mathematical definition of the model is given as:

$Y = 4.789 - 15.588X$, where X is the sum of yellow clusters metric result divided by 100. The odds ratio is 1.669e-07, which is less than 1. This suggests that for each additional 100 yellow clusters, the probability that an image depicts tumorous tissue increases. The Pearson residuals test has a p-value of 0.349, and the $G^2$test has a p-value of 0.101. Since the model passed both goodness of fit tests, this indicates that the model using the sum of yellow clusters is suitable to gauge whether an image shows tumorous tissue.

The model utilizing the results from the transitioned square metric has a predictive power of 0.548, or around 54.8%. The mathematical definition of the model is given as:

$Y = 1.105 - 1.2716X$, where X is the transitioned squares metric result divided by 1000. The odds ratio is 0.2804, which is less than 1. This suggests that for each additional 1000 transitioned squares, the probability that an image depicts tumorous tissue decreases. The result shows that the transitioned squares metric is not much better at differentiating tumorous breasts than random assignment. The Pearson residuals test has a p-value of 0.428, and the $G^2$test has a p-value of 0.009. Since the model failed to pass at least one goodness of fit test, this indicates that the model using the sum of transitioned squares divided by 1000 is not suitable to gauge whether an image shows tumorous tissue.

3.2.3 $\phi$ Analysis (MLO)

The model utilizing the results from the yellow square metric has a predictive power of 0.718, or around 71.8%. The mathematical definition of the model is given as: $Y = -1.5880 + 2.7827X$, where X is the yellow squares metric result divided by 1000. The odds ratio is 16.163, which is greater than 1. This suggests that for each additional 1000 yellow squares, the probability that an image depicts tumorous tissue increases. The Pearson residuals test has a p-value of 0.066, and the $G^2$ test has a p-value of 0.154. Since the model passed both goodness of fit tests, this indicates that the model using the sum of yellow squares divided by 1000 is suitable to gauge whether an image shows tumorous tissue.

The model utilizing the results from the yellow clusters metric has a predictive power of 0.710, or around 71.0%. The mathematical definition of the model is given as: $Y = -1.1220 + 0.3808X$, where X is the yellow clusters metric result. The odds ratio is 1.463, which is greater than 1. This suggests that for each additional yellow cluster, the probability that an image depicts tumorous tissue increases. The Pearson residuals test has a p-value of 0.492, and the $G^2$ test has a p-value of 0.066. Since the model passed both goodness of fit tests, this indicates that the model the sum of yellow clusters is suitable to gauge whether an image shows tumorous tissue.

The model utilizing the results from the combination metric has a predictive power of 0.712, or around 71.2%. The mathematical definition of the model is given as: $Y = 0.53597 + 0.27711X$, where X is the combination metric result divided by 1000. The odds ratio is 1.3193, which is greater than 1. This suggests that for each additional 1000 combination metric count, the probability that an image depicts tumorous tissue

increases. The Pearson residuals test has a p-value of 0.595, and the $G^2$test has a p-value of 0.082. Since the model passed both goodness of fit tests, this indicates that the model using combination metric is suitable to gauge whether an image shows tumorous tissue.

## 3.3 Logistic Regression Results with Categorical Predictive Variable

After fitting models with each metric from sections 2.3.1, 2.3.2, and 2.3.3, we decided to fit models with two predictive variables at a time. One was the categorical variable, density score of the breast, which we used in conjunction with each metric to see how this new variable affected the predictive power of the model. The age of the patient as a predictive variable was not considered at first because in the Marin et al. (2017) [11] paper, age was shown not to be significant.

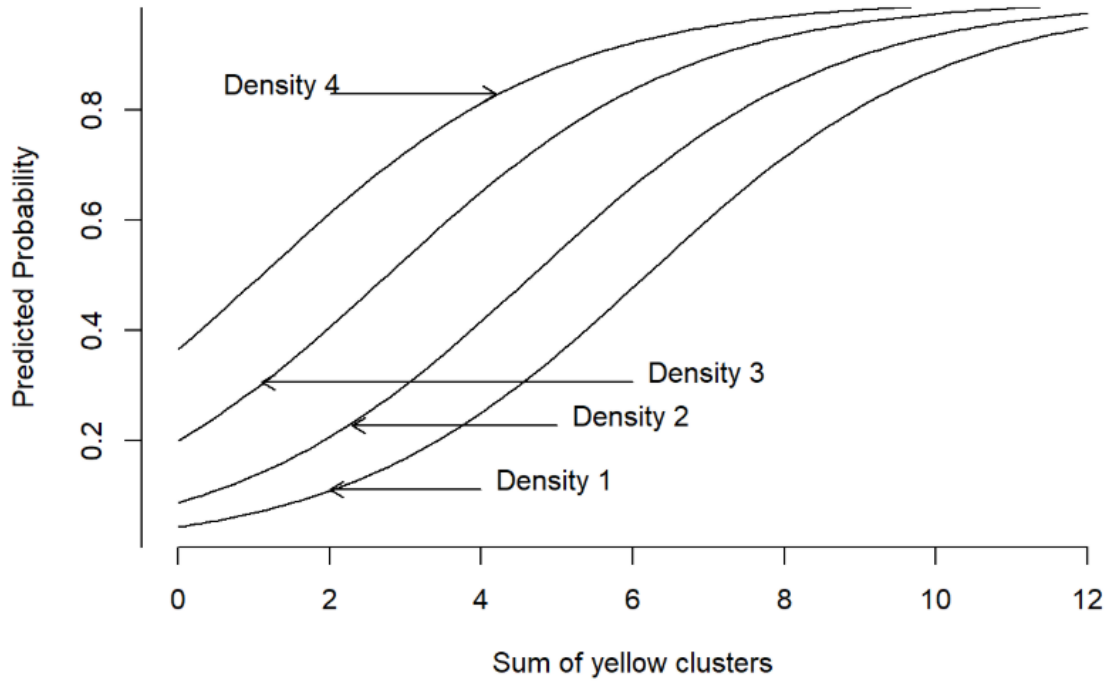


Figure 3.3: Logistic Curve of the model using two predictive variables, one of which a categorical variable.

The density score seems to have an inverse relationship with the median effect score. The higher the density score, the lower the median effect level. For higher density scores, as shown in Figure 3.3, the median effect score is lower, meaning a mammogram depicts dense breast tissue with a high density score, the number of clusters for differentiating mammograms with tumors from normal mammograms is lower compared to mammograms with lower density scores. In the future, one should not overlook the importance of breast density.

### 3.3.1 $\lambda$ Analysis (MLO)

The model utilizing density score as a categorical variable and the yellow squares metric as variables has a predictive power of 0.776, or around 77.6%. The mathematical definition of the model is given as: $Y = -2.7456 + 1.8755X + 0.9354Z_2 + 2.0518Z_3 + 2.5089Z_4$, where X is the yellow squares metric result divided by 1000, $Z_2$ is Density score of 2, $Z_3$ is the Density score of 3, and $Z_4$ is the Density score of 4. The odds ratio is 6.524, which is greater than 1. This suggests that for each additional 1000 yellow squares, the probability that an image depicts tumorous tissue increases. The Pearson residuals test has a p-value of 0.036, and the $G^2$test has a p-value of 0.166. Since the model failed to pass at least one goodness of fit test, this indicates that the model utilizing the sum of yellow squares divided by 1000 and density score is not suitable to gauge whether an image shows tumorous tissue.

The model utilizing density as a categorical variable and the yellow clusters metric as variables has a predictive power of 0.752, or around 75.2%. The mathematical definition of the model is given as: $Y = -1.9944 + 0.2980X + 0.7715Z_2 + 1.7248Z_3 +$

$2.2808Z_4$, where X is the yellow clusters metric result, $Z_2$ is Density score of 2, $Z_3$ is the Density score of 3, and $Z_4$ is the Density score of 4. The odds ratio is 1.347, which is greater than 1. This suggests that for each additional yellow cluster, the probability that an image depicts tumorous tissue increases. The Pearson residuals test has a p-value of 0.034, and the $G^2$test has a p-value of 0.118. Since the model failed to pass at least one goodness of fit test, this indicates that the model utilizing the sum of yellow clusters and density score is not suitable to gauge whether an image shows tumorous tissue.

The model using density as a categorical variable and the transitioned squares metric as variables has a predictive power of 0.775, or around 77.5%. The mathematical definition of the model is given as: $Y = -2.8001 + 0.7151X + 1.0311Z_2 + 2.2047Z_3 + 2.4798Z_4$, where X is the transitioned squares metric result divided by 1000, $Z_2$ is Density score of 2, $Z_3$ is the Density score of 3, and $Z_4$ is the Density score of 4. The odds ratio is 2.044, which is greater than 1. This suggests that for each additional 1000 transitioned squares, the probability that an image depicts tumorous tissue increases. The Pearson residuals test has a p-value of 0.088, and the $G^2$test has a p-value of 0.193. Since the model passed both goodness of fit tests, this indicates that the model utilizing sum of transitioned squares divided by 1000 and density score is suitable to gauge whether an image shows tumorous tissue.

The model using density as a categorical variable and the combination metric as variables has a predictive power of 0.748, or around 74.8%. The mathematical definition of the model is given as: $Y = -1.8417 + 3.3971X + 0.8991Z_2 + 1.7296Z_3 + 2.2591Z_4$, where X is the combination metric result divided by 100, $Z_2$ is Density score of 2, $Z_3$ is the Density score of 3, and $Z_4$ is the Density score of 4. The odds ratio is 29.877, which

is greater than 1. This suggests that for each additional 100 combination metric count, the probability that an image depicts tumorous tissue increases. The Pearson residuals test has a p-value of 0.287, and the $G^2$test has a p-value of 0.152. Since the model passed both goodness of fit tests, this indicates that the model utilizing the combination metric count divided by 100 and density score is suitable to gauge whether an image shows tumorous tissue.

### 3.3.2 $\xi$ Analysis (MLO)

For the cases of yellow and transitioned squares, since their respective models proved to be as effective as random guessing, there was no need to explore further how density could enhance the predictability of the two models.

The model using density as a categorical variable and the yellow clusters metric as variables has a predictive power of 0.795, or around 79.5 %. The mathematical definition of the model is given as: $Y = 3.4529 - 15.9624X + 1.0995Z_2 + 1.6900Z_3 + 2.1544Z_4$, where X is the yellow clusters metric result divided by 100, $Z_2$ is Density score of 2, $Z_3$ is the Density score of 3, and $Z_4$ is the Density score of 4. The odds ratio is 1.17e-08, which is less than 1. This suggests that for each additional 100 yellow clusters, the probability that an image depicts tumorous tissue increases. The Pearson residuals test has a p-value of 0.463, and the $G^2$test has a p-value of 0.184. Since the model passed both goodness of fit tests, this indicates that the model utilizing the sum of yellow clusters divided by 100 and density score is suitable to gauge whether an image shows tumorous tissue.

### 3.3.3 $\phi$ Analysis (MLO)

The model utilizing density as a categorical variable and the yellow squares metric as variables has a predictive power of 0.812, or around 81.2%. The mathematical definition of the model is given as: $Y = -3.8353 + 3.3783X + 1.1407Z_2 + 2.1294Z_3 + 2.8386Z_4$, where X is the yellow squares metric result divided by 1000, $Z_2$ is Density score of 2, $Z_3$ is the Density score of 3, and $Z_4$ is the Density score of 4. The odds ratio is 29.321, which is greater than 1. This suggests that for each additional 1000 yellow squares, the probability that an image depicts tumorous tissue increases. The Pearson residuals test has a p-value of 0.666, and the $G^2$test has a p-value of 0.351. Since the model passed both goodness of fit tests, this indicates that the model utilizing the sum of yellow squares divided by 1000 and density score is suitable to gauge whether an image shows tumorous tissue.

The model utilizing density as a categorical variable and the yellow clusters metric as variables has a predictive power of 0.785, or around 78.5%. The mathematical definition of the model is given as: $Y = -3.1173 + 0.5043X + 0.7590Z_2 + 1.7252Z_3 + 2.5637Z_4$, where X is the yellow clusters metric result, $Z_2$ is Density score of 2, $Z_3$ is the Density score of 3, and $Z_4$ is the Density score of 4. The odds ratio is 1.656, which is greater than 1. This suggests that for each additional yellow cluster, the probability that an image depicts tumorous tissue increases. The Pearson residuals test has a p-value of 0.422, and the $G^2$test has a p-value of 0.223. Since the model passed both goodness of fit tests, this indicates that the model using the sum of yellow clusters and density score is suitable to gauge whether an image shows tumorous tissue.

The model utilizing density as a categorical variable and the combination metric as variables has a predictive power of 0.768, or around 76.8%. The mathematical definition of the model is given as: $Y = -18.5913 + 1.1846X + 15.9745Z_2 + 17.5494Z_3 + 17.9387Z_4$, where X is the combination metric result divided by 1000, $Z_2$ is Density score of 2, $Z_3$ is the Density score of 3, and $Z_4$ is the Density score of 4. The odds ratio is 3.269, which is greater than 1. This suggests that for each additional 1000 combination metric count, the probability that an image depicts tumorous tissue increases. The Pearson residuals test has a p-value of 0.414, and the $G^2$test has a p-value of 0.142. Since the model passed both goodness of fit tests, this indicates that the model using the combination metric count divided by 1000 and density is suitable to gauge whether an image shows tumorous tissue.

## 3.4 Best Fitted Models

In the best fit models, we did not attempt to put more than one metric in at a time due to collinearity concerns, because all metrics are derived from the yellow square metric. Age and density score were added because they have potential to improve the fit of the model. The best fit models start with the most complex model, and then the backwards elimination algorithm is used to find a simpler model to describe the data.

### 3.4.1 $\lambda$ Analysis

We learned that half of the potential models from section 3.2.1 fit the data poorly. After running the backward elimination algorithm on the most complex model, the results show that the number of transitioned squares metric in conjunction with other dependent

variables yielded the highest predictive power for making future predictions. The mathematical definition of the model is given as:

$$Y = 2.81727 - 0.95719X_1 - 1.76142Z_2 + 6.6626Z_3 - 9.788505Z_4 - 0.09926X_2 - 0.010861X_1Z_2$$
$$-4.93515X_1Z_3 + 4.50316X_1Z_4 + 0.02846X_1X_2 + 0.01510X_2Z_2 - 0.10270X_2Z_3 + 0.25474X_2Z_4$$
$$+0.01611X_1Z_2X_2 + 0.10168X_1Z_3X_2 - 0.09404X_1Z_4X_2,$$

where $X_1$ is the transitioned squares metric result divided by 1000, $Z_2$ is Density score of 2, $Z_3$ is Density score of 3, $Z_4$ is Density score of 4, $X_2$ is age. The above model has the highest predictive power for full breast cases, 0.846 or around 84.6%. The Pearson residuals test has a p-value of 0.082, and the $G^2$ test has a p-value of 0.365. Since the model passed both goodness of fit tests, this indicates that the model is suitable to gauge whether an image shows tumorous tissue.

3.4.2 $\xi$ Analysis

We learned from section 3.2.2 that most models fit the data poorly and therefore most models were rejected. After running the backward elimination algorithm on the most complex model, the results show that the number of yellow clusters metric in conjunction with other dependent variables yielded the highest predictive power for making future predictions. The mathematical definition of the model is given as:

$$Y = -3.3254 + 11452.06X_1 + 3333.93Z_2 + 3322.97Z_3 + 3320.29Z_4 + 48.29X_2 - 11479.01X_1Z_2$$
$$-11456.99X_1Z_3 - 11434.21X_1Z_4 - 167.05X_1X_2 - 48.35X_2Z_2 - 48.12X_2Z_3 - 48.04X_2Z_4$$

$+167.67X_1Z_2X_2 + 167.18X_1Z_3X_2 + 166.71X_1Z_4X_2$, where $X_1$ is the yellow clusters metric result divided by 100, $Z_2$ is Density score of 2, $Z_3$ is Density score of 3, $Z_4$ is Density score of 4, $X_2$ is age. The above model uses all interaction terms between the three different predictive variables. The above model has the highest predictive power for nipple region cases, 0.858 or around 85.8%. The Pearson residuals test has a p-value of 0.800, and the $G^2$test has a p-value of 0.456. Since the model passed both goodness of fit tests, this indicates that the model is suitable to gauge whether an image shows tumorous tissue.

3.4.3 $\phi$ Analysis

We learned from section 3.2.3 that none of the models fit the data poorly. However, the predictive power was not as high as it was in models with more predictive variables. After running the backward elimination algorithm on the most complex model, the results show that the number of yellow squares metric in conjunction with other dependent variables yielded the highest predictive power for making future predictions. The mathematical definition of the model is given as:

$$Y = 38172.5 - 45807.3X_1 - 38189.7Z_2 - 38183.1Z_3 - 38171.8Z_4 - 747.0X_2 + 45826.1X_1Z_2$$
$$+45815.5X_1Z_3 + 45810.6X_1Z_4 + 893.2X_1X_2 + 747.2Z_2 + 747.1X_2Z_3 - 746.9X_2Z_4$$
$$-893.4X_1Z_2X_2 + 893.3X_1Z_3X_2 - 893.1X_1Z_4X_2$$

where $X_1$ is the yellow squares metric result divided 1000, $Z_2$ is Density score of 2, $Z_3$ is Density score of 3, $Z_4$ is Density score of 4, $X_2$ is age. The above model uses all interaction terms between the three different predictive variables. The above model has the best predictive power for full breast cases, 0.868 or around 86.8%. The Pearson residuals test has a p-value of 0.715, and the $G^2$test

has a p-value of 0.532. Since the model passed both goodness of fit tests, this indicates that the model is suitable to gauge whether an image shows tumorous tissue.

# 4. CONCLUSION

The goal of this paper is to understand as much about the nuances of the breast microenvironment as possible, and also to devise new models to differentiate mammograms with tumorous tissue from normal breast tissue. The Wilcoxon Rank Sum Test (Upper-tail) demonstrated that cancerous and tumorous cases typically exhibit more disrupted tissue than normal cases. Further, when analyzing the odds ratio from the single variable logit models and two variable logit models, the results suggest that as the amount of disrupted tissue increases, so does the probability that the analyzed mammogram will demonstrate tumorous characteristics. The probability that an image will exhibit tumorous characteristics is also dependent on breast density. More specifically, the logit models predicting the probability that a given image depicts tumorous tissue is shown to be dependent on each metric in conjunction with breast density. Finally, the logit models with the highest predictive power utilize each metric in conjunction with the density score and age as mentioned in Section 3.4.

The metrics outlined in Marin et al (2017) [11] (i.e. sum of yellow clusters, sum of yellow clusters, sum of transitioned squares, and combination metric) were shown to differentiate between cancer and normal groups, benign and normal groups, and tumorous and normal groups in Section 2.3.1 by utilizing the Wilcoxon Rank Sum Test (Two-tailed). All the p-values for the cancerous versus normal cases, benign versus normal cases, and tumorous versus normal cases are less than 0.05. Since the p-values are less than 0.05, this suggests that there is a significant difference in the breast microenvironment for cancerous, benign, and tumorous groups compared to normal cases. The metrics used in the $\phi$ Analysis (i.e. the sum of yellow squares, sum of yellow

clusters, and combination metric as defined in Section 2.3.3) also were able to determine significance difference for cancerous versus normal cases, benign versus normal cases, and tumorous versus normal cases because the p-values for all these cases are less than 0.05. This alone, however, is not enough to establish a link between the loss of tissue homeostasis and tumor growth because the conclusion one can draw from the Wilcoxon Rank Sum Test (Two-tailed) is that the mean number of disrupted regions for the cancer group is different from the mean number of disrupted regions in the normal group.

Marin et al (2017) [11] hypothesized that there is a link between loss of tissue homeostasis and tumor growth. In order to determine if the cancer group has a higher mean number of disrupted regions than the normal group, one must utilize the Wilcoxon Rank Sum Test (Upper-tail). The claim is further explored in Section 2.6 for the sum of yellow squares metric in the $\lambda$ Analysis and $\phi$ Analysis. The p-values for the Wilcoxon Rank Sum Test (Upper-tail) for cancerous versus normal, benign versus normal, and tumorous versus normal cases are less than 0.05 for both the $\lambda$ Analysis and $\phi$ Analysis. This suggests that mammograms from the cancer and tumor groups typically exhibit more disrupted regions than mammograms from the normal group in the respective areas of analysis. Also, this may suggest that loss of tissue homeostasis leads to tumor growth and cancer development. This is further emphasized when examining the odds ratio of logit models because the odds ratio higher probability of a breast exhibiting tumorous characteristics for each increase in a certain characteristic (E.g. 1000 yellow squares).

The logistic modeling methodology was utilized due to its ability to quantify the probability that a predictive variable will lead to identification of tumorous tissue in a mammogram. When examining logit models with a single predictive variable, the odds

ratio for models in the $\lambda$ Analysis and $\phi$ Analysis are greater than 1. This suggests that for an increase in each characteristic (e.g., increase in number of yellow squares), the probability of an image depicting tumorous characteristics also increases. Further, when examining logit models with each metric in conjunction with breast density, breast density was shown to increase the probability that a mammogram depicts tumorous growths. More specifically, in the case of the sum of yellow clusters in $\phi$ Analysis, the median effect level for mammograms with higher breast density score is lower compared to the median effect level for mammograms with lower breast density score. In the hopes of devising models maximizing p-values, more variables were added.

We decided to create logit models from running the backwards elimination algorithm on the most complex logit model including each metric in conjunction with density score, age, and all respective interaction terms. The resulting predictive power ranged from around 85% to 87%, which is higher than the predictive power of CAD which is supposedly around 63%, and also higher than some of the prosed augmented CAD methodology at 65% predictive power [14]. The logit models yielding maximum predictive power also happen to be the most complex logit models. This suggests that all variables are required to devise the models with the highest predictive power and also all variables contribute to the probability that tumorous characteristics will be identified in an image.

## 4.1 FUTURE WORK

In the future, I would like to utilize the cross validation algorithm as a means of analyzing the predictive power of additional mammograms with the models devised in

this thesis. Further, I would like to consider other potential predictive variables such as family history and previous diagnosis. Additionally, in conjunction with the methods devised in this thesis, I would utilize longitudinal studies to examine how breast tumors develop over time and identify critical moments in early development. The end goal is to eventually implement the methodology devised in this thesis and future work in the clinical setting.

# 5. BIBLIOGRAPHY


[1] National Cancer Institute, "Female Breast Cancer - Cancer Stat Facts." [Online]. Available: https://seer.cancer.gov/statfacts/html/breast.html. [Accessed: 05-Aug-2019]

[2] C. D. Lehman, R. D. Wellman, D. S. M. Buist, K. Kerlikowske, A. N. A. Tosteson, and D. L. Miglioretti, "Diagnostic Accuracy of Digital Screening Mammography With and Without Computer-Aided Detection," *JAMA Intern Med*, vol. 175, no. 11, pp. 1828–1837, Nov. 2015 [Online]. Available: https://jamanetwork.com/journals/jamainternalmedicine/fullarticle/2443369. [Accessed: 27-Aug-2019]

[3] M.-S. Ong and K. D. Mandl, "National Expenditure For False-Positive Mammograms And Breast Cancer Overdiagnoses Estimated At $4 Billion A Year," *Health Affairs*, vol. 34, no. 4, pp. 576–583, Apr. 2015 [Online]. Available: https://www.healthaffairs.org/doi/full/10.1377/hlthaff.2014.1087. [Accessed: 26-Aug-2019]

[4] American Cancer Society, "Cancer Screening Guidelines | Detecting Cancer Early," *American Cancer Society Guidelines for the Early Detection of Cancer*. [Online]. Available: https://www.cancer.org/healthy/find-cancer-early/cancer-screening-guidelines/american-cancer-society-guidelines-for-the-early-detection-of-cancer.html. [Accessed: 27-Aug-2019]

[5] Digital Database for Screening Mammography, "Digital Database for Screening Mammography." [Online]. Available: http://www.eng.usf.edu/cvprg/Mammography/DDSM/thumbnails/benigns/benign_02/case1269/A-1269-1.html. [Accessed: 05-Aug-2019]

[6] National Institute of Biomedical Imaging and Bioengineering, "X-rays | National Institute of Biomedical Imaging and Bioengineering," *X-rays*. [Online]. Available: https://www.nibib.nih.gov/science-education/science-topics/x-rays. [Accessed: 26-Aug-2019]

[7] S. J. Shin, *A Comprehensive Guide to Core Needle Biopsies of the Breast*. Springer, 2016.

[8] National Institute of Biomedical Imaging and Bioengineering, "Mammography | National Institute of Biomedical Imaging and Bioengineering." [Online]. Available: https://www.nibib.nih.gov/science-education/science-topics/mammography. [Accessed: 05-Aug-2019]

[9] "Mammography (Mammogram)." [Online]. Available: https://www.radiologyinfo.org/en/info.cfm?pg=mammo. [Accessed: 05-Aug-2019]

[10] "Mammograms," *National Cancer Institute*, 06-Jan-2017. [Online]. Available: https://www.cancer.gov/types/breast/mammograms-fact-sheet. [Accessed: 05-Aug-2019]

[11] J. D. Keen, J. M. Keen, and J. E. Keen, "Utilization of computer-aided detection for digital screening mammography in the United States, 2008 to 2016," *Journal of the American College of Radiology*, vol. 15, no. 1, pp. 44–48, 2018.

[12] Z. Marin *et al.*, "Mammographic evidence of microenvironment changes in tumorous breasts," *Medical physics*, vol. 44, no. 4, pp. 1324–1336, 2017.

[13] M. Hollander and D. A. Wolfe, *Nonparametric statistical methods*. Wiley New York, NY, USA, 1973.

[14] A. Agresti, *Categorical Data Analysis*, 3rd ed. Hoboken, New Jersey: John Wiley & Sons Inc., 2013.

[15] S. Mirniaharikandehei, A. B. Hollingsworth, B. Patel, M. Heidari, H. Liu, and B. Zheng, "Applying a New Computer-aided Detection Scheme Generated Imaging Marker to Predict Short-term Breast Cancer Risk," Phys Med Biol, vol. 63, no. 10, p. 105005, May 2018 [Online].
Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5976448/. [Accessed: 27-Aug-2019]

# 6. AUTHOR'S BIOGRAPHY

Dexter G. Canning was born in Taiwan on August 19, 1996 and raised in central Maine. Dexter graduated from Foxcroft Academy in May 2015. He started off as a Civil Engineering student at UMaine before switching majors to Mathematics. He'll be graduatinig with a Bachelor's of Science in Mathematics with a Minor in Statistics, and he served as the President of Pi Mu Epsilon. In his term, he helped organize a math career panel to help math students network and get jobs after graduation as well as a few social events to help gain recognition for the math department.

Upon graduation, Dexter will be matriculating into the Mathematics Masters program here at UMaine in the Fall of 2019 studying under the direction of Professor Zheng Wei.