ABSTRACT

RIGGAN, BENJAMIN SCOTT. Study of Morphology and Artificial Neural Networks for Detecting Microcalcifications. (Under the direction of Dr. Wesley Snyder.)

X-ray mammography is the most widely used screening process to find early indications of breast cancer. Radiologists have a systematic method using low dose x-rays to analyze the breast densities. Although this method of screening is very effective, radiologists have been known to make mistakes: either by misdiagnosing breast cancer or by missing early indications of it. This thesis investigates, in detail, the methods of autonomously detecting microcalcifications (one major indicator of breast cancer if located in clusters), so that mammograms can more efficiently be analyzed. The methods used to detect microcalcifications has four components: a pre-processor, a segmentation algorithm, a feature vector, and a classifier. The performances of the several classifiers are studied and compared. Also, the relationships and dependencies between the microcalcification features are studied in order to better understand why such classification techniques work.

The pre-processor which best finds points of interest is the h-dome transform in conjunction with morphological noise reduction techniques. The h-dome transform detects regional maxima while eliminating maxima with dynamics below a specified level. This method when used with the noise removal techniques eliminates many false positives within the image, and thus provides a good starting point for classification.

The segmentation algorithm is chosen to best extract the pixels of the microcalcifications. The algorithm chosen is the "hill climbing" algorithm which locates the points with the maximum change in brightness radially outward from the center of the region. The "hill climbing" algorithm provides a consistent boundary for the suspected regions, which allows for the features to be correctly extracted.

The feature vector is constructed using statistical measures that are related to microcalcifications. The features studied include: area, contrast, mean brightness, brightness variance, second central moments, and entropy. This feature vector helps to describe the size, shape, texture, and statistical brightness of the segmented microcalcification. Classifying these features using supervised learning techniques help identify dependencies between these features.

Various classifiers are tested using the pre-processing technique, segmentation algorithm, and feature vector described. The classifiers that are discussed include: the Mahalanobis distance, perceptron, multilayer perceptron, radial basis function networks, and support vector machines. The performances of these classifiers are investigated in order to determine how the classifiers compare with one another. This study finds that the radial basis function network with 150 cluster centers provides the best classification performance with 99.9% true positive

rate and 0.8% false positive rate on average. However, the perceptron generalizes the best.

The analysis of the features includes a study of individual features and pairs of features. The individual measures are analyzed using the sample means and variances, in order to provide an overall expectation for each feature. Then, the pairs of feature are studied both scatter plots and the sample covariance matrix. The feature study reveals the "very useful", "useful", and "not so useful features."

Study of Morphology and Artificial Neural Networks for Detecting Microcalcifications

by
Benjamin Scott Riggan

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Electrical Engineering

Raleigh, North Carolina

2011

APPROVED BY:

Dr. Joel Trussell	Dr. Ginger Yu

Dr. Wesley Snyder
Chair of Advisory Committee

DEDICATION

I would like to dedicate this work to my parents, Raymond and Susan Riggan; to my family; to Alyssa Smith; and to all of my friends who have helped and encouraged me along the way.

BIOGRAPHY

Benjamin Riggan was born in Charleston, South Carolina. He graduated from North Carolina State University with a Bachelor of Science degree in Computer Engineering, in December 2009. While working as an undergraduate teaching assistant, he gained interests in the fields of computer vision and image processing. Some initial experience in these fields were gained by working on an autonomous vehicle research project headed by Dr. Wesley Snyder. On this project, he helped write software that performed stop sign recognition and vehicle recognition. After graduating, he continued his education at North Carolina State University, pursuing a Master of Science degree in Electrical Engineering. He continues to work with Dr. Wesley Snyder, and his research currently focuses on the methods used for autonomous detection of microcalcifications.

Benjamin Riggan will further his education by pursuing a Doctor of Philosophy degree in Electrical Engineering at North Carolina State University, after completion of the Masters degree. He will continue to study under the direction of Dr. Wesley Snyder, where he will obtain a better understanding of the techniques and applications used in the fields of computer vision and image processing.

ACKNOWLEDGEMENTS

I thank my committee for their help and guidance. First, I thank Dr. Wesley Snyder for all his time and effort put into advising, teaching, and guiding me. I would also like to thank Dr. Joel Trussell for challenging me and making me a better student and researcher. Last, I would like to thank Dr. Ginger Yu, who encouraged me to go to graduate school and pursue my dreams. I also want to thank her for all the time and effort spent in reviewing this thesis.

Secondly, I thank the imaging group directed by Dr. Wesley Snyder who have helped me along way, both inside and outside the classroom. I thank Dr. Rachana Gupta, Theju Jacobs, Yongqiang (Garfield) Huang, and Bo Liu for the all of their support, encouragement, and aid.

Lastly, I thank my brother, Mark Riggan, for his help and expertise in Adobe Photoshop and Illustrator, which helped in the generation of models, graphs, and figures.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

In this study, the results of various microcalcification detection techniques are compared in order to find which methods perform well under different conditions. The study analyzes the results from various classifiers and determines which performs the best, given the extracted features. The comparison of these methods provides better insight into the problem of detecting calcifications. The detection of microcalcifications is not trivial due to the complexity of the image and effects from noise. These complications cause feature selection, segmentation, and classification to be difficult.

1.1 Microcalcifications

1.1.1 What is a Microcalcification?

"Microcalcifications are tiny granular deposits of calcium that appear on the mammogram as small bright spots" [10]. Clusters of microcalcifications (or MCs) in mammograms may be an indication of breast cancer, and the identification of these clusters currently requires the service of an expert radiologist. An example of a cluster of MCs is shown in Figure 1.1. MCs alone are typically very small. "According to standard literature on microcalcifications, they are up to about $0.7mm$ in size having an average diameter of $0.3mm$. Objects smaller than $0.1mm$ can hardly been seen in the mammogram because of the tissue background texture and noise" [7]. Radiologists have been known to misdiagnose breast cancer and even miss early signs of cancer because of their small size, the amount of noise present, and also varying type of breast tissue densities. In some studies, signs of breast cancer were present two years prior to the diagnosis [10]. The earlier that breast cancer can be diagnosed, generally the better the prognosis [33].

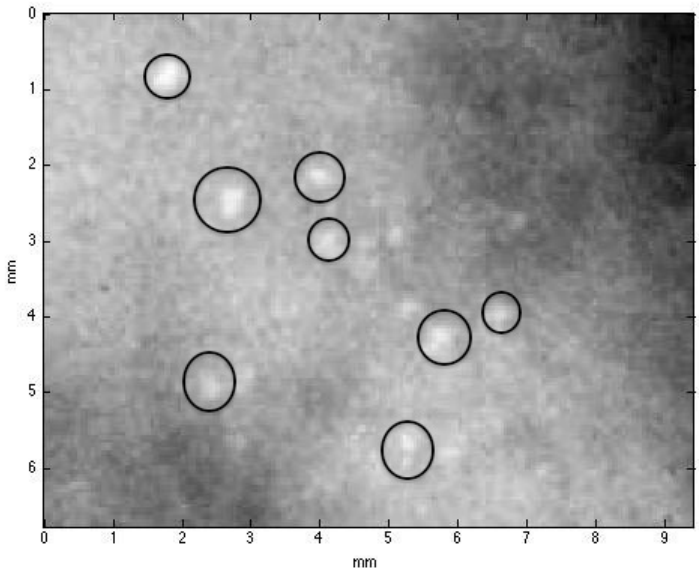


Figure 1.1: A small cluster of MCs.

1.1.2 How are Microcalcifications Detected?

Mammography is the most widely used technique for breast cancer screening, and for non-invasively finding and analyzing indicators of breast cancer. Mammography is more formally defined as the "[Radiographic] examination of the breast by means of x-rays, ultrasound, nuclear magnetic resonance, etc" [14]. Expert radiologists analyze the mammograms, looking for signs of cancer such as clustered MCs or dense regions corresponding to tumors or lesions in the breast. They analyze the MCs shape, type, and cluster distribution in order to estimate the malignancy of the tissue. If any indicators (or potential indicators) of cancer are found, the radiologists compare the current mammograms with previous annual screenings in order to better determine the need for further testing or biopsy. The most widely used form of mammography for breast cancer screening is x-ray mammography, opposed to infrared (thermography) or ultrasound. An example of an x-ray mammogram is shown in Figure 1.2.

1.1.3 X-ray Mammography

X-ray mammography is performed by transmitting x-rays through the breast, and detecting the amount of x-rays output after being absorbed by the tissue. A full mammography screening is performed by x-raying each breast from two different perspectives, the craniocaudal (CC) and the mediolateral oblique (MLO). The CC view is formed by transmitting x-rays through

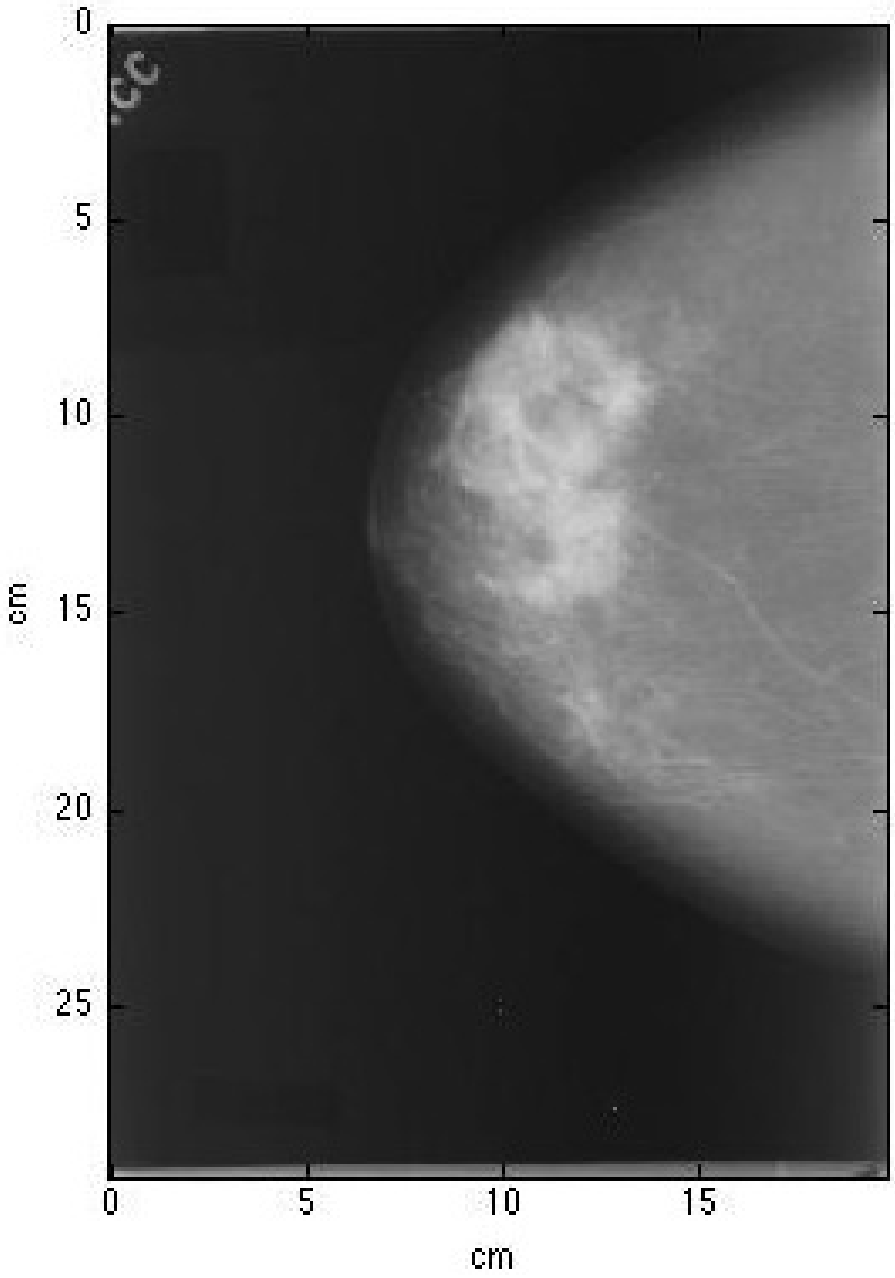


Figure 1.2: A mammogram indicating cancer.

the breast from the top through the bottom, and the MLO view is created by transmitting the x-rays through the side of the breast at an angle.

1.2 Automated Microcalcification Detection

The complexity of mammograms makes it difficult, even for radiologists to identify MCs, but with current image processing capabilities, it is possible to design tools that can help radiologists with identifying MCs [19]. However, there does not exist a standard definition of what an MC is in terms of imaging techniques, or how the performances of these techniques can be evaluated [31]. The purposes for automated detection of MCs are to more efficiently analyze mammograms and to possibly reduce the number of errors that are made by radiologists. Although there is a need for a golden (or ideal) performance standard, no attempt will be made to create one.

Reproducibility is another major issue when it comes to MC detection algorithms; most papers do not mention it. Since there is no performance standard, most software developers use their own measure of performance and usually verify their findings with a radiologist. Due to this lack of performance standard and the fact that many details from the detection methods are left out, results reported are typically difficult to reproduce.

1.3 Purpose

The primary focus of this paper is on the detection of individual MCs. Although the indicators of breast cancer are clusters of MCs, it is trivial to cluster MCs once they are detected. Therefore, clustering is not explicitly discussed. Many papers discuss detection of individual MCs, yet provide the true positive detection rate of the clusters. This allows a better detection rate to be reported, since not all MCs have to be detected. Individual MCs are difficult to detect thus providing lower detection rates, however, when clustered the overall performance is much better. The results of the classifiers in this paper are reported in terms of the individual MCs.

Secondly, the features used to represent MCs are analyzed. These features provide better understanding of MC detection based on the physical characteristics of the MCs. The study helps to explain the reasons these detection methods work. This analysis discusses many aspects of the relationships between features, including: the more discriminative features, the linearity (or nonlinearity) of feature pairs, and the independent feature pairs.

However, this paper does not:

1. Use classifiers to detect very large regions containing clusters of MCs.

2. Use classifiers to detect large regions containing single MCs.

These type MC detection methods rely on very large feature vectors that can only be classified by an SVM, due to the "curse of dimensionality." Although, these methods work well, no physical meaning can be interpreted from the results.

1.4 Organization of Thesis

Chapter 1 provides an introduction to MC detection, the problems radiologists have with identifying the MCs, and the reasons for this study. The screening process that radiologists use for identifying MCs is provided. In addition, automated MC detection and its related issues are discussed.

Chapter 2 describes the typical components of MC detection and a brief literature review. The pre-processing and segmentation methods commonly used are given. A list of features commonly used for detecting MCs are provided, and various classifiers and their general performances are discussed.

Chapter 3 discusses the methodology of this study. The details of the morphological preprocessor are explained. Then, the "hill climbing" algorithm, used for segmenting the MCs, is discussed. The features, and their mathematical definitions, that are specifically chosen for this study are provided. It also reviews the classifiers studied. Chapter 4 discusses data sets used in this study, including the databases from which the collections of mammograms were retrieved. This chapter discusses the details of the data sets including size, resolution, and sampling. Chapter 5 provides the results from the study are discussed. The results from the MC detection method using the various classifiers are discussed and compared. These results are presented and the implications of the results are discussed.

Chapter 6 discusses the analysis of the the features. First, the means and variances of the individual features are provided. Then, the relationships between feature pairs are discussed in terms of their ability to discriminate between MCs and non-MCs, and in terms of linearity and nonlinearity. Also, the covariance matrices for the true positives and the true negatives are discussed.

Chapter 7 provides the conclusions drawn from the study. Future work in the area of MC detection techniques is discussed, as well as, proposed experiments extending to this study and its results.

Chapter 2

Previous Work

Most of the detection methods have similar components: a pre-processor, segmenter, feature extractor, and classifier [4]. The pre-processor eliminates regions of the mammogram that most likely are not associated with an MC. They are typically not very robust when solely used for classification. The segmenter separates the pixels of the MCs from the rest of the image. Segmenters are usually used so that features can be correctly extracted. Feature extraction is similiar among all methods, however, there are a wide variety of features used for different detection techniques. Lastly, the classifier is the final determining factor whether a region is considered to be an MC or not.

2.1 Pre-processing Methods

Three types of pre-processing techniques seen in the literature include thresholding, morphological, and filtering methods. Thresholding, in general, is a method of segmenting object pixels, in this case pixels belonging to MCs, from background pixels. Morphological methods typically change the image such that much of the background and noise are removed from the object image. Last, filtering methods tend to eliminate specific bands in the frequency domain so that information retained correspond more to MCs.

2.1.1 Thresholding

Thresholding can be local or global; global thresholding depends on the entire image and local thresholding depends on the local regions or neighborhoods. Global thresholding is performed in a manner such that

$$C = \{f(x) > T\} \tag{2.1}$$

for a given threshold, T, and a given image, f. Local thresholding is similar, except that the threshold, T is dynamically determined by regional features (e.g. average pixel brightness over 3×3 neighborhood). In general, local thresholding tends to be better than global since there is a high degree of variability in mammograms; a single threshold is not accurate enough to detect the MCs. Since the optical density is the sum of all absorption through the tissue, the MC's density appears to be different when the MC is seen against a high density compared to a low density background. Therefore, simple thresholding is not sufficient to accurately detect MCs.

There are a few methods that analyze thresholding in conjunction with MC detection, which include Kallergi et al. [16] and Davies and Dance [6]. Kallergi et al. compares local thresholding with local region growing, and the results show that local thresholding is more consistent comparatively, however, local thresholding "is [also] more dependent on parameter selection than the [region growing] procedure" [16]. Davies and Dance [6] also use local thresholding with overlapping sub-images, so that a pixel must be in a pre-determined number of sub-images in order to be in the final segmented image. This method, in conjunction with classifier, performed with 92% correct classification of clusters on a set of 50 mammograms. Similar thresholding techniques which are used for MC detection are discussed in the survey by Sahoo et al. [25].

2.1.2 Morphological Operators

"Mathematical morphology is the application of lattice theory to spatial structure" [28], or the analysis of shape, structure, and form. Then, a morphological operator, or morphological transformation, is defined by the infinite set of operators formed by combining any form (i.e union, product, or complement) of the hit-or-miss transformations (i.e. dilation or erosion) [28]. There are a variety of morphological operations that are used to enhance the mammograms and to eliminate areas of the mammogram so that detection is simplified and less computationally complex. Some of these operations include the top hat transform [26], h-dome transform, dilation [27], and opening. Mossi and Albiol [20] implement the top-hat transform, which is the difference between the original image and its grey-level opening (discussed in section 3.2.2). The top hat transform uses morphology to detect regions that include regional maxima. This transform enhances the detected maxima (or domes) that are smaller than the structuring element.

The h-dome transform is similar to the top-hat transform except instead of opening, the grayscale reconstruction (section 3.2.3) operation is used. This type of operation also enhances features that are smaller than the structuring element, but in addition, it also suppresses noise with magnitude less than the user defined constant "h" [10, 34]. Halkiotis et al. [10] also include opening (i.e. erosion followed by dilation) for noise removal in section 3.2.2. Details of

the morphological operations [36] used, and the h-dome transform are discussed in full detail in section 3.2.

2.1.3 Filters

Linear filtering [23] methods are also very common pre-processing techniques. The different types of filtering techniques include high pass filtering and wavelet analysis (or filter bank type of analysis). Since MCs are small (relative to the breast), their presence introduces sudden changes in brightness, which correspond to high spatial frequencies. Therefore, a high pass filter will be the most sensitive to the presence of MCs. This simple idea is extended to filter banks and wavelet analysis, in which processing is performed in the higher frequency sub-bands. Srivastava [31] uses both of these types of pre-processing for comparison of methods, but noted that they did not have the best performance. This may be due to the fact that MCs are not the only high frequency feature in mammograms.

Wavelet analysis is shown to be "useful in enhancing the visibility of [MCs]" [32]. Both Strickland and Hahn [32], and Nakayama et al. [21] use wavelet analysis to pre-process the mammograms. Strickland and Hahn model the MCs by "using a circularly-symmetric Guassian function" [32], which yields around 55% true positive detection rate with the wavelet analysis. Achieving 94.3% detection of true clusters of MCs, Nakayama et al. [21] use filter banks to enhance the mammograms such that MCs are better classified.

Laine et al. [18] use multiscale analysis, which is an extension from wavelet analysis where various scalings for the wavelets are used. The multiscale analysis is used for adaptive contrast enhancement specifically for mammography. Like the wavelet analysis and other filtering techniques, multiscale analysis allows for better MC classification. The disadvantages to filtering are that information is lost, and it is more difficult to understand and define MCs in terms of their features once the features have been filtered.

2.2 Segmentation

There are many types of segmentation algorithms. However, two common methods used for segmenting MCs are region growing and "hill climbing" [1]. Both techniques try to group pixels into a region from a starting point, or seed pixel. This starting point is typically a pixel that is suspected to be the center or near the center of the MC.

2.2.1 Region Growing

Region growing starts with a seed pixel and recursively finds all pixels which are "close" in brightness to their neighboring pixels, thus producing a single connected bright region. Kallergi

et al. [16] and Bankman et al. [1] both perform comparative studies with region growing. Bankman et al. [1] conclude that region growing typically provides regions larger than expected for segmenting the MCs.

2.2.2 "Hill Climbing"

The "hill climbing" algorithm on the other hand finds the largest slope between the center pixel and pixels down a particular radial line [1]. The maximum measure of slope from each radial line marks the boundary pixels of the MCs. This segmentation algorithm works well because the MC pixels are bright with respect to the surrounding pixels. Note, there exists a point where the brightness changes the most in a given distance from the center thus marking the boundary of the MC. This method seems to work better than region growing according to Bankman et al. [1]. The "hill climbing" algorithm is discussed more in depth in section 3.3.2.

2.3 Features

There are two approaches to classification of regions as containing an MC: 1) make some small number of measurements of the region and make decision based on those few measurements, or 2) use all the pixels in the region, themselves, as features. There are a variety of measurements or features used for detecting MCs some of which are listed in Table 2.1. The features that provided better results may differ between detection methods due to differences in pre-processing and segmentation of the MCs. Wroblewska et al. [35] provide a more complete list of features. There are statistical features like mean, variance, and background which are grey-level type of computations. Then, there are shape features like the moments, compactness, and perimeter. Textural features [17] include measurements like skew and kurtosis. Last, co-occurrence matrix features include measures like contrast, correlation, entropy and energy. These are the types of features commonly used, in a variety of ways, for detecting calcifications.

Instead of classifying multiple measurements, a region of pixels itself can be classified. Srivastava [31] classifies a 9×9 region of pixels, which corresponds to an 81×1 feature vector in which the pixel brightnesses are the features. Srivastava gets this idea of classifying regions of pixels from El-Naqa et al. [9]. Each region is centered around a single pixel, which is classified either as belonging to an MC or not. Although the results show that classifying the region performs well, the disadvantages include:

1. Computational complexity due to the large feature vector and having to classify every pixel in the image.

2. Difficulty of drawing conclusions from the classification, since the classifiers act like "black boxes."

Table 2.1: List of microcalcification features.

No.	Feature	No.	Feature
1	Area	8	Entropy
2	Mean	9	Energy
3	Standard deviation	10	Correlation
4	Background	11	Variance
5	Compactness	12	Skew
6	Hu's Invariant Moments [15]	13	Kurtosis
7	Contrast		

2.4 Classification Methods

The final component of MC detection is the classifier, which is a way to determine the class (i.e. "MC" or "not MC") a particular sample belongs. The set of possible classifiers which may be used for MC classification include:

- linear classifiers (perceptrons and Fisher's linear discriminant)
- multilayer perceptrons (MLPs)
- radial basis function (RBF) networks
- support vector machines (SVMs)

where linear and nonlinear here are defined in terms of the boundary that separates the classes (not the activation functions).

The classifiers listed above all use "supervised learning." That is, sample inputs with known expected outputs are used to train the classifier. This type of classifier is typically used independent of which pre-processing or segmentation algorithms are used.

2.4.1 Fisher's Linear Discriminant

Fisher's linear discriminant [8] is a classifier that determines the d-dimensional vector $\mathbf{w}$, where the d-dimensional samples from each class are projected onto $\mathbf{w}$, such that the projected data from both classes are separated with minimal error using a simple scalar threshold. The samples

from class i, $\mathbf{x}_i \in \mathbb{R}^d$, are transformed into the projected samples from class i, $y_i \in \mathbb{R}^1$, using $y_i = \mathbf{w}^T\mathbf{x}_i$. In order to calculate the optimum vector $\mathbf{w}$, the objective function

$$J(\mathbf{w}) = \frac{\mathbf{w}^T\mathbf{S}_B\mathbf{w}}{\mathbf{w}^T\mathbf{S}_W\mathbf{w}} \tag{2.2}$$

is maximized with respect to $\mathbf{w}$, where $\mathbf{S}_B$ is the between class scatter matrix and $\mathbf{S}_W$ is the within class scatter matrix.

$$\mathbf{S}_B = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \tag{2.3}$$

$$\mathbf{S}_W = \sum_{\mathbf{x}\in C_1} (\mathbf{x} - \boldsymbol{\mu}_1)(\mathbf{x} - \boldsymbol{\mu}_1)^T + \sum_{\mathbf{x}\in C_2} (\mathbf{x} - \boldsymbol{\mu}_2)(\mathbf{x} - \boldsymbol{\mu}_2)^T \tag{2.4}$$

Equation 2.3 defines the between class scatter matrix, and Equation 2.4 defines the within class scatter matrix, where

$$\boldsymbol{\mu}_i = \frac{1}{n_i} \sum_{\mathbf{x}\in C_i} \mathbf{x} \tag{2.5}$$

Fisher's linear discriminant, being a linear classifier, is limited in performance since only a linear boundary between the classes can be formed. This classifier works best when the data of the classes are distributed in Gaussian clusters and have minimal overlap. Typically, this is not the case with most features of MCs. However, Fisher's linear discriminant is used for MC detection in [24]. Rodriguez et al. [24] claim their classification performance using Fisher's linear discriminant is 90% correct classification of MCs.

2.4.2 Artificial Neural Networks

There are a variety of artificial neural network (ANN) classifiers, the simplest being the perceptron (section 3.5.2), which models the behavior of a single neuron. The multilayer perceptron (MLP) is probably the most popular of neural networks used today, largely because a simple algorithm (back-propagation) is available to determine a good choice of operating parameters. The MLP (section 3.5.3) is a nonlinear classifier and tends to perform better than the perceptron because the perceptron, being a linear machine, can only use a linear surface (line, plane, hyperplane) to divide the classes, whereas the MLP can develop arbitrarily complex dividing surfaces. Thus, the MLP is more commonly used than the perceptron for MC detection.

Halkiotis et al. [10] compare two different MLP topologies, one single hidden layer with 5 hidden neurons and another with 10 hidden neurons. The network with 10 hidden neurons achieved the best classification of MCs. Papadopoulos et al. [22] also tested various config-

urations of MLPs for MC detection, which include an architecture of 15 hidden nodes, the use of principal component anlaysis (PCA), and different training algorithms, like gradient descent, resilient back-propagation, conjugate gradient and quasi-Newton. For the MLP with 15 neurons in the hidden layer, the quasi-Newton training algorithm provided the most accurate classification of MC clusters.

2.4.3 Radial Basis Function Networks

Radial basis function (RBF) networks are also used for classifying MCs. This type of classifier takes advantage of Cover's theorem [5], which states that "a complex pattern-classification problem, cast in high dimensional space nonlinearly, is more likely to be linearly separable than in a low-dimensional space, provided that the space is not densely populated" [11]. The RBF classifier operates by nonlinearly casting the m_0-dimensional vector into the m_1-dimensional space, $\mathbb{R}^{m_1}$, where $m_1 > m_0$. This is accomplished using radial basis functions, which are real valued functions that depend only on the Euclidean distance from each of the m_1 cluster centers. This allows for a linear surface to divide the classes in $\mathbb{R}^{m_1}$, given that the classes are separable. When the linear surface in $\mathbb{R}^{m_1}$ is nonlinearly projected back into $\mathbb{R}^{m_0}$, the resulting dividing surface is nonlinear. The implementation of the RBF network for classifying MCs is discussed further in section 3.5.4.

Halkiotis et al. compare the MLP classifiers with an RBF network classifier, and reported that the MLP with ten hidden nodes achieved the best classification (although the best true positive rate was the same for both the RBF network and MLP network) [10].

2.4.4 Support Vector Machines

Last, support vector machines (SVMs) are another type of classifier used for MC detection. SVMs also take advantage of Cover's theorem [5] and in general provide good classification, given a representative dataset. SVMs, like RBF networks, map the m_0-dimensional input vector into $\mathbb{R}^{m_1}$. Instead of using radial basis functions, SVMs use kernel functions which depend on the Euclidean distance from each of the m_1 support vectors. Given the support vectors and the outputs from the kernel functions, a linear surface can be found that will divide the classes (provided that the classes are separable) in $\mathbb{R}^{m_1}$. Therefore, the surface separating the classes in $\mathbb{R}^{m_0}$ is nonlinear. The implementation of the SVM for classifying MCs, including how the support vectors are determined, is discussed in greater detail in section 3.5.5.

In classifying MCs, Papadopoulos et al. [22] reported a better performance with an SVM with a Gaussian kernel than the MLP, and Bazzani [2] claimed to have comparable results with other methods using SVMs. El-Naqa et al. show that SVM classification of MCs provided excellent performance, and claims that "the SVM classifier out performed all other methods

considered" [9].

Chapter 3

Methodology of Study

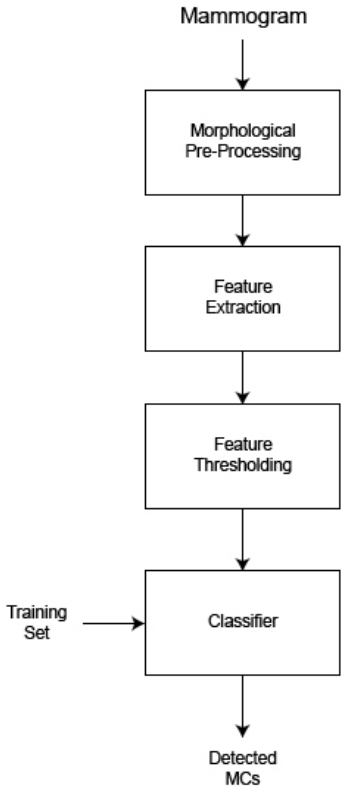


Figure 3.1: Overview of detection method.

First, an analysis of the film response is discussed to ensure that no correction is needed for the pixel brightness to correctly represent density. The components of the study include the h-dome transform (the pre-processor chosen to study), the "hill climbing" algorithm (segmentation method), the features vector, and the classifiers. Each component is discussed in full detail. The block diagram in Figure 3.1 provides the overall flow of the method used for autonomously identifying the MCs.

3.1 Film Response

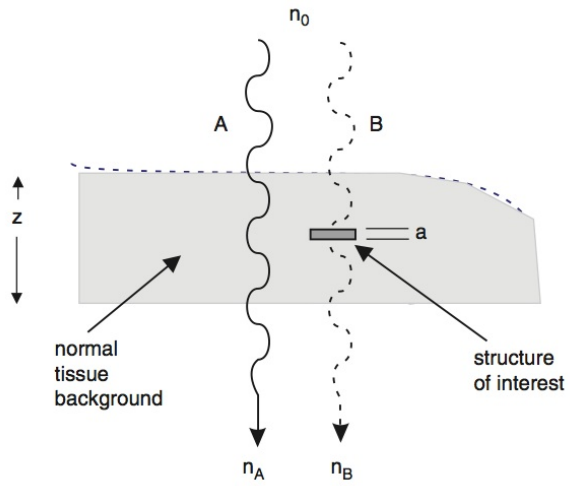


Figure 3.2: X-rays passing through breast tissue and structure of interest [3].

The basic concept of x-ray mammography is shown in Figure 3.2, where n_0 defines the mean number of x-rays incident on the breast. The thickness of the normal breast tissue, z, and the thickness of the structure of interest, a, are labeled in the diagram. The formula for calculating the mean number of x-rays, n_A and n_B, after passing through the breast along paths A and B are listed in Equations 3.1 and 3.2, where the x-ray attenuation coefficients of the tissue and structure are represented by μ and μ'.

$$n_A = n_0 \exp(-\mu z) \tag{3.1}$$

$$n_B = n_0 \exp(-\mu(z-a) - \mu' a) \tag{3.2}$$

n_A and n_B indicate a measure of density for the tissue. The relationship between the mean number of x-rays detected (after passing through the breast) and the density is inversely proportional. The mammogram itself represents a density mapping of the breast. Table 3.1 [12, 13] shows an example of the mapping from pixel value to optical density, D. This mapping is defined by the model: $D = (f(x) - 4096.99)/(-1009.01)$, where $f(x)$ is the pixel value. This model is the linear regression of the experimental data (measured optical density and mean pixel brightness). This mapping applies to film since silicon detectors for digital mammograms do not have the same type of response. The brighter regions, in terms of pixel brightness, correspond to more dense areas, which may correspond to tumors or calcifications. The relationship between optical density, which is the measure of light transmitted through the film, and the logarithm of the relative x-ray exposure is shown in Figure 3.3, where the maximum slope of the linear region is represented by γ. From Figure 3.3, the formula (Equation 3.3) for optical density, D, can be derived. I_0 represents the amount of light incident on the film and I_t is the amount of light transmitted through the film. The relationship between optical density and exposure reveals that the film response is nonlinear. Note, this nonlinear relationship does not hold for digital mammography, which is discussed in section 4.1.

$$D = \gamma \log\left(\frac{I_0}{I_t}\right) \tag{3.3}$$

Table 3.1: Experimental brightness to density mapping for Lumisys scanner [12, 13].

Optical Density	**Mean Grey Level**
0.04	4069
0.21	3897
0.34	3759
0.50	3601
0.69	3443
0.82	3269
0.96	3113
1.13	2948
1.28	2794
1.44	2635
1.59	2478
1.74	2324
1.89	2167
2.07	2004
2.21	1855
2.37	1690
2.52	1542
2.68	1386
2.84	1235
2.99	1099
3.13	957
3.60	495

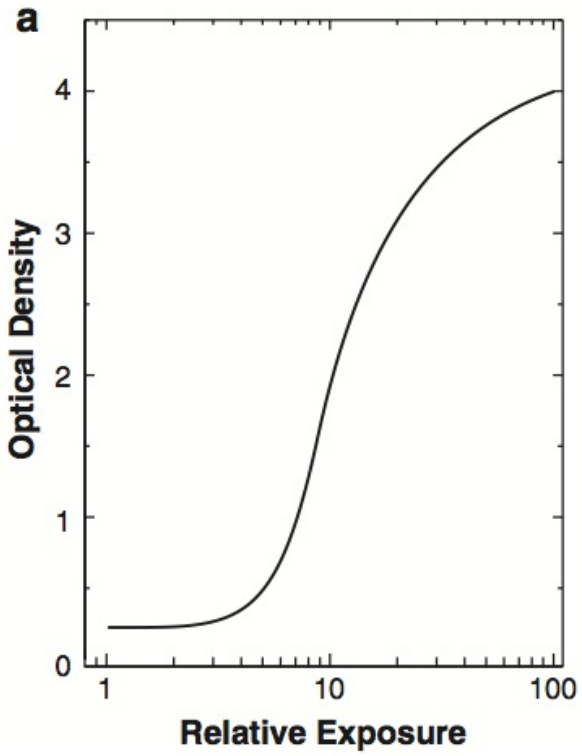


Figure 3.3: Relationship between optical density and exposure [3].

As mentioned before, there is a nonlinear relationship between the optical density (approximated by a linear relationship with pixel brightness) and the logarithm of the exposure (Figure 3.3). In order to verify that brightness correction is not necessary due to this nonlinearity, the contrast between optical densities, in terms of exposure, is analyzed. Let d_{B_1} be the optical density associated with one background (one part of the breast), and let d_{μ_1} be the optical density of an MC in the breast tissue associated with the first background. Equation 3.4 shows the definition of the the optical density of the background alone where α_1 is a factor related to the absorption of x-rays in the first background, and Equation 3.5 shows the optical density of the background and the MC where β is a factor related to the absorption in the MC only.

$$d_{B_1} = \gamma \log_{10}(\alpha_1 \eta_0) \tag{3.4}$$

$$d_{\mu_1} = \gamma \log_{10}(\alpha_1 \beta \eta_0) \tag{3.5}$$

Looking at the difference of these two expressions, we get a measure of contrast between the optical density of the MC and the optical density of its backgrounds (Equation 3.6). Using the same MC with the same absorption parameter, β, and placing it in another background with density, d_{B_2}, and absorption parameter, α_2, a measure of contrast (Equation 3.7) with respect to the second background (a different part of the breast) can be calculated.

$$\begin{aligned} C_{B_1} &= d_{\mu_1} - d_{B_1} \\ &= \gamma \log_{10}(\alpha_1 \beta \eta_0) - \gamma \log_{10}(\alpha_1 \eta_0) \\ &= \gamma \log_{10}(\beta) \end{aligned} \tag{3.6}$$

$$\begin{aligned} C_{B_2} &= d_{\mu_2} - d_{B_2} \\ &= \gamma \log_{10}(\alpha_2 \beta \eta_0) - \gamma \log_{10}(\alpha_2 \eta_0) \\ &= \gamma \log_{10}(\beta) \end{aligned} \tag{3.7}$$

Note that the contrasts, despite what background is present, are independent of the background. This conclusion is verified by plotting the contrast (feature described below) with respect to background. See Figure 3.4 for the plot. The plot shows that, for all intents and purposes, the brightness contrast is independent of background brightness. Also, the MCs tend to have a contrast measuring larger than the non-MCs, and the non-MCs may have a negative contrast which means that the background brightness is greater than the brightness of the MC (see definition in section 3.4.2).

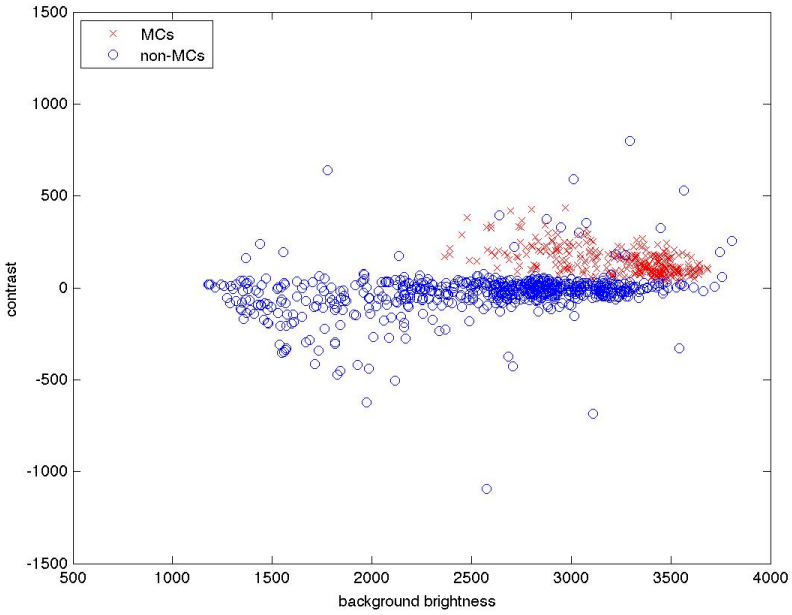


Figure 3.4: Contrast independent with respect to background.

3.2 Pre-processing

For the pre-processor, the morphological methods by Halkiotis et al. [10] are followed because their method explicitly locates regional maxima while suppressing noise. Although, they do use a series of morphological transformations and operations that extract regional maxima, they do not use an explicit segmentation method. Their method assumes that the domes correctly segment the MCs by using the binary image as a mask. The different morphological components include the h-dome transform, grayscale opening [30], grayscale reconstruction [34], extended maxima transform [30], and connected components labeling [29].

3.2.1 H-dome transform

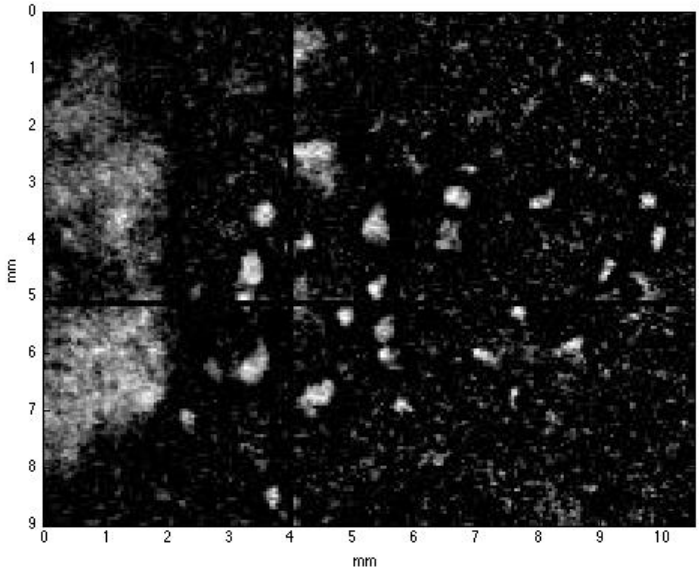
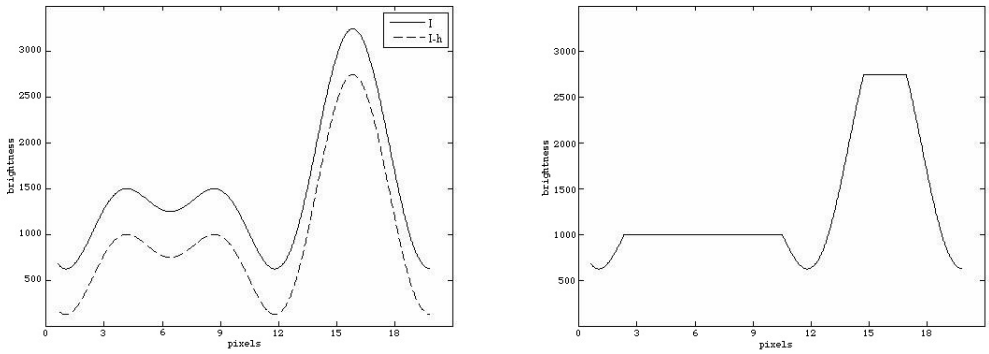


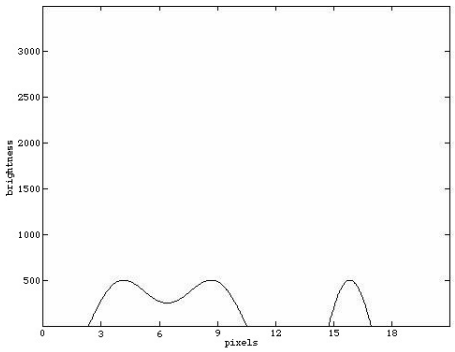
Figure 3.5: Output from h-dome.

The h-dome transform is used because it is a morphological operation that locates regional maxima (essentially what an MC is in mammograms) in addition to limiting the effects from noise. The h-dome transform starts by subtracting a constant "h" from every pixel in the image (Figure 3.6a). The constant "h" is defined to be a small, user defined constant which may be determined by the maximum possible brightness in the mammogram. Halkiotis et al. [10] note that a broad spectrum of the constant "h" yield maxima due to MCs. However, "h" should be chosen to be "small" compared to the maximum possible brightness. Experimentally, "h" is chosen to be 100 with respect to 4095 (12 bits per pixel). By reconstructing the original image (I) from the difference image ($I - h$), the newly reconstructed image will match the original image except at regional maxima (Figure 3.6b). For details on reconstruction, refer to grayscale reconstruction section 3.2.3. Then, the difference between the original image and the reconstructed image yield the regional maxima or domes (Figure 3.6c). The h-dome transform does suppress some noise, however noise is still present (Figure 3.5). Therefore, the noise removal operation of opening is employed.



(a) original

(b) reconstructed



(c) h-dome transform

Figure 3.6: H-dome transform example.

3.2.2 Opening

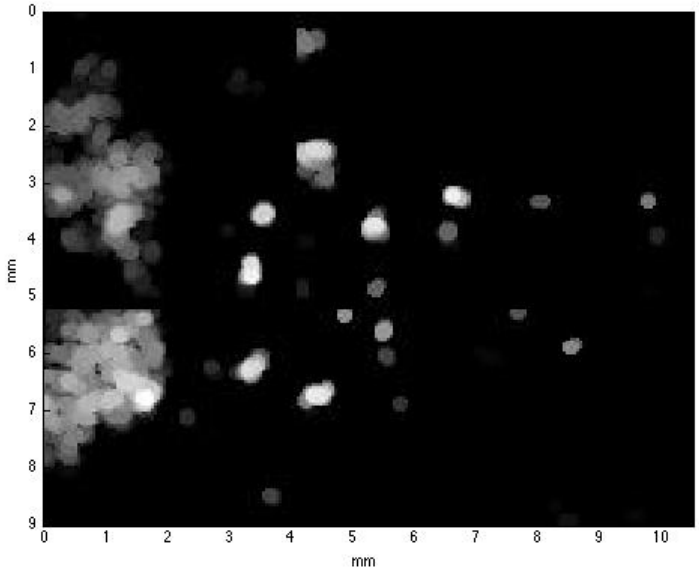


Figure 3.7: Output after opening.

Opening, which is erosion followed by dilation (see definition of erosion in Equation 3.8 and dilation in Equation 3.9), is simply a way of reducing the amount of noise output from the h-dome transform. The choice of constant structuring element, B, is a circle with a radius of two pixels [10]. The method used by Halkiotis et al. [10] is followed here as well since it involves morphological noise removal. The output after opening significantly reduces the amount of noise (Figure 3.7). Opening accomplishes this by eliminating the pixels in Figure 3.5 that are smaller than the structuring element B.

$$\varepsilon_B(f(x)) = \min_{b \in B} f(x + b) \tag{3.8}$$

$$\delta_B(f(x)) = \max_{b \in B} f(x + b) \tag{3.9}$$

3.2.3 Grayscale Reconstruction

Grayscale reconstruction, which is shown in Figure 3.6a and Figure 3.6b, comes from a geodesic dilation operation (Equation 3.10). Dilation itself is a maximum type operation of over the

structuring element, which allows the pixels to increase toward the value in the original image. The structuring element is the size and shape which is used to change the pixels in the images. However, geodesic dilation also includes a pointwise minimum operator ($\wedge$). So, the geodesic dilation is the dilation of image $J \leq I$ "under" I, which is defined as the pointwise minimum of the dilation of image J by a structuring element B and the original image I [34].

$$\delta_I^{(1)}(J) = \delta_B(J) \wedge I \tag{3.10}$$

Now, the grayscale reconstruction can formally be defined as the iterative geodesic dilation of J under I until it reaches stability [34]. The definition of stability in [34] is vague, so stability is more rigorously defined to be no changes from the reconstructed image at iteration n to iteration $n+1$. This means that the grayscale reconstruction relies on global convergence over the entire image, which can be very slow depending on the size of the image and choice of "h". The grayscale reconstruction equation is shown in Equation 3.11. The $\vee$ operator denotes the pointwise maximum operation. Since, $J \leq I$ then $\delta_I^{(1)}(J) \leq I$. Despite how many times the geodesic dilations are performed, this property holds and thus never can exceed I. The maximum brightness over the reconstructed image is $\max(J)$, assuming $J \leq I$. Remember, from above, that the reconstruction in the h-dome transform J is defined to be $I-h$.

$$\rho_I(J) = \bigvee_{n \geq 1} \delta_I^{(n)}(J) \tag{3.11}$$

The combination of opening followed by grayscale reconstruction allows for the removal of noise and domes with low dynamics [10]. This reduces the number of possible MCs that will be segmented and classified.

3.2.4 Extended Maxima Transform

The extended maxima transform is a morphological form of thresholding. "The extended maxima tranformation is the regional maxima computation of the corresponding h-maxima transformation" [10], or

$$EMAX_h(f) = RMAX(HMAX_h(f)) \tag{3.12}$$

The h-maxima transform is the reconstruction by dilation (same as grayscale reconstruction definition) of image f by $f-h$, where "h" again is a small constant (Equation 3.13). Note, the constant "h" here may be different from the one used for the h-dome transform. Next, the regional maxima computation takes the difference of the input image and the reconstruction of the input image by the input image minus one (Equation 3.14).

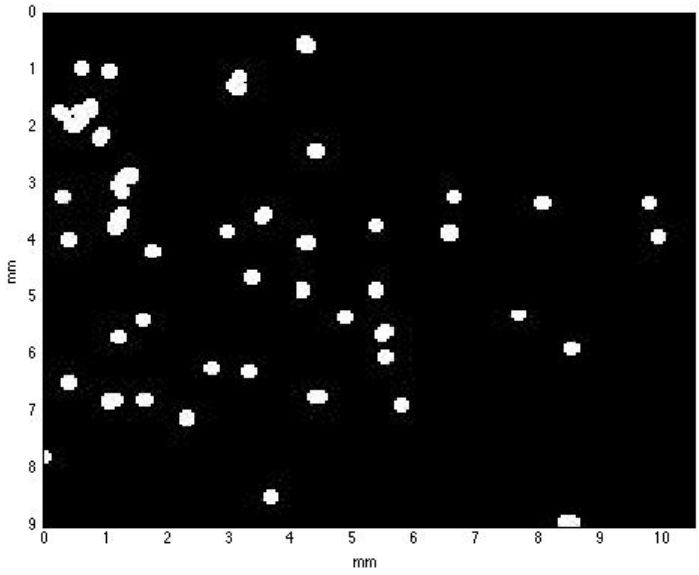


Figure 3.8: Output after extended maxima transfrom.

$$HMAX_h(f) = \rho_f(f - h) \tag{3.13}$$

$$RMAX(f) = f - \rho_f(f - 1) \tag{3.14}$$

Therefore, after all the reconstruction thresholding, much of the noise is removed and a mask for the domes with certain dynamics are left. An example of the output from the extended maxima stage is shown in Figure 3.8.

3.2.5 Connected Components

The connected component provides the final binary masking of the all the morphological operators discussed above. Diverging slightly from Halkiotis et al. [10], the center of gravity is computed for each connected component. The seed for segmenting the MCs is the center of gravity.

3.3 Segmentation

3.3.1 Region Growing

Region growing is one method used for segmenting MCs. As mentioned before region growing depends on the neighbor and local neighborhood thresholds set. There exist dynamic methods of setting thresholds, since MCs may somewhat differ in neighborhood contrast as well as in rate of change between pixel brightness. Bankman et al. [1] show that the "hill climbing" algorithm outperforms region growing and so, "hill climbing" is used in this work.

3.3.2 "Hill Climbing"

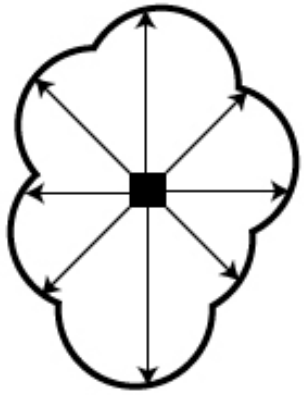


Figure 3.9: Radial lines for the "hill climbing" algorithm.

The "hill climbing" algorithm is based on finding the maximum rate of change of brightness as one travels from the center pixel along a radial line outward (Figure 3.9). Bankman et al. [1] use sixteen equally spaced radial lines, however, connectivity of the region is better ensured by calculating the number of radial lines needed. The number of radial lines is given by $\lceil 2\pi r_{max} \rceil$, where r_{max} is the maximum radius considered to be the boundary. In our work, $r_{max} = 15$ which corresponds to a maximum physical radius of $0.75mm$. This method is indeed superior since the region considered to be an MC is better defined.

Consider a seed pixel that is not part of an MC, then the maximum change in brightness from the center tends to be farther away from the center (which is the maximum brightness

of the region). Therefore, the maximum slope marking the boundary of regions corresponding to non-MCs are closer to the center of the regions, providing smaller areas to extract features. Conversely, MCs tend to correspond to larger areas than non-MCs.

The definition of slope or rate of change in brightness is shown in Equation 3.15, where $f(x_0, y_0)$ corresponds to the maximum brightness in the region (i.e brightness of center) and $d(x_0, y_0, x, y)$ is the Euclidean distance, in the plane, between the center pixel, (x_0, y_0), and the pixel, (x, y), considered down the radial line. The maximum slope for each radial line, by this definition of slope, marks a boundary pixel of the MC for that particular radial line.

$$s(x, y) = \frac{f(x_0, y_0) - f(x, y)}{d(x_0, y_0, x, y)} \tag{3.15}$$

3.4 Extracted Features

In the case of MCs, the features are selected such that they provide a statistical representation of a suspicious area in the mammogram. These suspicious areas are found by pre-processing (discussed in section 3.2) the mammograms. The features of MCs calculated include: area, contrast, average brightness, variance, second central moments, and entropy.

3.4.1 Area

Typically, mammogram film scanners digitize film using resolution, σ_r, of either $\sigma_r = 100\mu m$ or $\sigma_r = 50\mu m$. After segmentation, the area of an MC is estimated by $A = \sigma_r^2 n$, where n is the number of pixels in the MC. The calcification is segmented during pre-processing, and the total number of pixels is easily calculated. Area is used to account for any other feature that may have a dependence on the size of the object. Secondly, the size of the calcification determines whether there is an indication of cancer or not. Typically larger calcifications are not associated with malignant lesions.

3.4.2 Contrast

Literally, the contrast is the difference between the MC pixels and the background, which is the same definition used in Equations 3.6 and 3.7. In this case, it is more rigorously defined as:

$$C_{mc} = \frac{1}{9}\sum_{x=1}^{9} f(x) - \frac{1}{N_B}\sum_{y=1}^{N_B} f(y) \tag{3.16}$$

where the first term is the average brightness over a 3×3 neighborhood about the center pixel, and the second term is the average brightness over the N_B number of background pixels. A background pixel is defined to be 2 pixels out from the boundary of the MC in the direction

of the vector formed between the center pixel and the boundary pixel. This provides a general relationship between the center of the suspicious region and the background. MCs are known to have larger contrast compared to the contrast of soft tissue on average, however, due to noise and other variability in mammograms, contrast cannot be used as the only determining feature. For example, the crossing ducts in Figure 3.10 show up as bright relative to the background. More features are needed to detect the MCs without increasing the number of false positives.

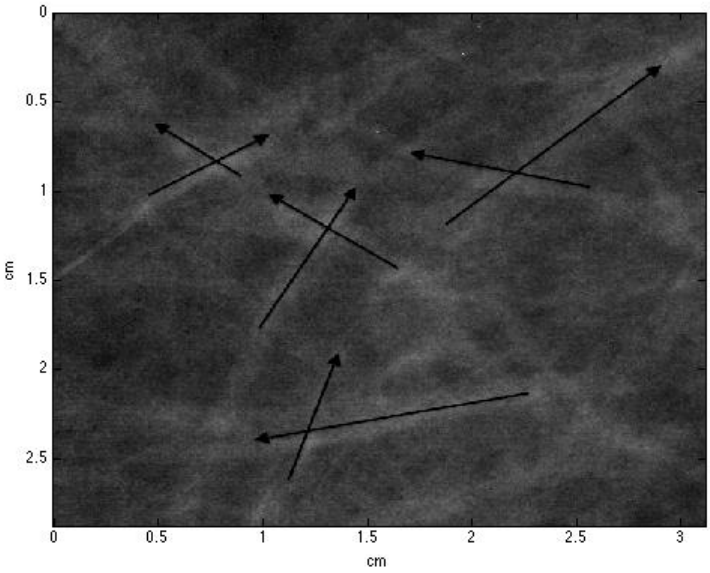


Figure 3.10: Crossing ducts in mammogram.

3.4.3 Average Brightness

The average brightness of all pixels in the potential calcification is used as another feature. This feature is useful because MCs tend to more commonly be in malignant lesions, although they can be present throughout the soft breast tissue too. Since MCs can vary across many brightness levels, this feature is used to account for any brightness dependency. Equation 3.17 shows the formula for calculating the average brightness, where M is defined to be the set of all pixels in a particular MC and N is the number of pixels within the MC.

$$\mu = \frac{1}{N} \sum_{x \in M} f(x) \tag{3.17}$$

3.4.4 Variance

Variance of all pixels in the suspicious area is defined in Equation 3.18. The variance provides insight into how MCs vary compared to the soft tissue. This is a widely used feature for detection and classification of MCs. For example, Halkiotis et al. [10] use variance as one of the features for their classifier. Since MCs correspond to higher spatial frequency, they tend to have higher variance than soft tissue. However, noise and area also become a factor with this feature and therefore more features are needed to better describe MCs.

$$\sigma^2 = \frac{1}{N} \sum_{x \in M} [f(x) - \mu]^2 \tag{3.18}$$

3.4.5 Second Central Moments

Before jumping to the central moments, the moments m_{pq}, themselves, are calculated by integrating over the area containing the MC, represented by M, as in Equation 3.19. Values p and q are substituted into the equation and determine which moments are calculated (i.e. first, second, etc.). In order to calculate the central moments, the center of gravity is needed. The center of gravity is computed by calculating m_x and m_y (Equation 3.20), which are the centers of gravity with respect to each axis. m_x is computed using Equation 3.19 with $p = 1$ and $q = 0$ in the numerator, and $p = 0$ and $q = 0$ in the denominator. Then, m_y is computed using Equation 3.19 with $p = 0$ and $q = 1$ in the numerator, and $p = 0$ and $q = 0$ in the denominator.

The central moments, μ_{pq}, are defined in Equation 3.21 [29], where m_x and m_y are computed as described above. The difference between the moments and the central moments is that the central moments are invariant to both translation and rotation. The second central moments are computed by letting $p = 2$ and $q = 0$ for one feature, and $p = 0$ and $q = 2$ for the other. As mentioned before, calcifications vary in size and shape, and these features help to describe the shape of the MCs.

$$m_{pq} = \iint_M x^p y^q f(x, y) \, \mathrm{d}x \, \mathrm{d}y \tag{3.19}$$

$$m_x = \frac{m_{10}}{m_{00}} \qquad\qquad m_y = \frac{m_{01}}{m_{00}} \tag{3.20}$$

$$\mu_{pq} = \iint_M (x - m_x)^p (y - m_y)^q f(x, y) \, \mathrm{d}x \, \mathrm{d}y \tag{3.21}$$

3.4.6 Entropy

Last, entropy is defined in Equation 3.22, where the summation is over the number of brightness values, or bins. The probability function of brightness, $v = P[f(x, y)]$, is estimated by a histogram. Entropy is a measure of randomness, and is used as a characteristic of texture for the suspicious region. The feature is chosen because the texture is different between soft tissue and MCs.

$$S = -\sum v \cdot \log_2 v \tag{3.22}$$

Many different features have been studied in the application for MC detection. During this study, there exist features that were not used in this paper. Some of those feature include:

- Average brightness of the background
- Kurtosis
- Skew

The average brightness of the background, as discussed in section 3.4.2, is not directly useful for detecting MCs. This feature does not provide any information about the calcification, rather, it provides information about the surrounding tissue. This background is merely used as reference, and therefore is not used directly in detecting MCs.

The features describing texture, kurtosis and skew, are also not used our method. Some other methods choose to use these features, such as Halkiotis et. al. [10]. However, in this study the experimental results did not show any improvement in detection by using these features. Since features describing texture are already being used to describe MCs, it is sufficient not to use kurtosis and skew.

All of the features listed make up the input space for the classifiers tested. These features help describe several aspects of MCs and soft tissue in mammograms, such as size, statistical measures of brightness (i.e. mean, variance, and etc.), shape, and texture. The combination of these features will help classify suspicious regions into MCs and soft tissue.

3.5 Classifiers

In this section, we review the classifiers, including the Mahalanobis distance, perceptron, multilayer perceptron (MLP), radial basis function (RBF) networks, and support vector machines (SVMs), so we can compare the results of the classifiers for the MC detection method discussed.

3.5.1 Mahalanobis Distance

The Mahalanobis distance is a simple pattern recognition technique that is used to identify which class an input vector belongs, where the feature vector, $\mathbf{x}$, is the set of measurements or features extracted from a segmented region in the mammogram. The Mahalanobis distance is computed with the following formula, using vector notation

$$r^2 = (\mathbf{x} - E[\mathbf{x}])^T \mathbf{\Sigma}^{-1} (\mathbf{x} - E[\mathbf{x}]) \tag{3.23}$$

where $\mathbf{\Sigma}^{-1}$ is the matrix inverse of the covariance matrix $\mathbf{\Sigma}$. The covariance matrix is then defined in Equation 3.24.

$$\mathbf{\Sigma} = E[(\mathbf{x} - E[\mathbf{x}])(\mathbf{x} - E[\mathbf{x}])^T] \tag{3.24}$$

However, since the expectation of all possible samples is not feasible. An estimation of the covariance is required. The estimator of the covariance is given by the sample covariance matrix, where

$$E[\mathbf{x}] = \boldsymbol{\mu} = [\mu_1 \ \mu_2 \cdots \ \mu_N]^T \tag{3.25}$$

The sample covariance matrix generally is computed as

$$\mathbf{\Sigma} \approx \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1j} \\ \Sigma_{21} & \Sigma_{22} & \cdots & \Sigma_{2j} \\ \vdots & \vdots & \cdots & \vdots \\ \Sigma_{i1} & \Sigma_{i2} & \cdots & \Sigma_{ij} \end{bmatrix} \tag{3.26}$$

where the sample covariance of element i, j is computed as

$$\Sigma_{ij} = \frac{1}{N} \sum_{k=1}^{N} (F_i(k) - \mu_i)(F_j(k) - \mu_j) \tag{3.27}$$

where N is defined to be the number of samples for the features F_i, and $\mu_i = \sum_{k=1}^{N} F_i(k)$.

By computing the sample covariance matrices for both the true positive training samples and the true negative training samples, the Mahalanobis distance can be computed for both sets too. After calculating these distance measures, the class to which the input $\mathbf{x}$ belongs is determined by the class with the smallest Mahalanobis distance. Every input is then classified using the smallest Mahalanobis distances.

3.5.2 Perceptron with Least Mean Square weight update

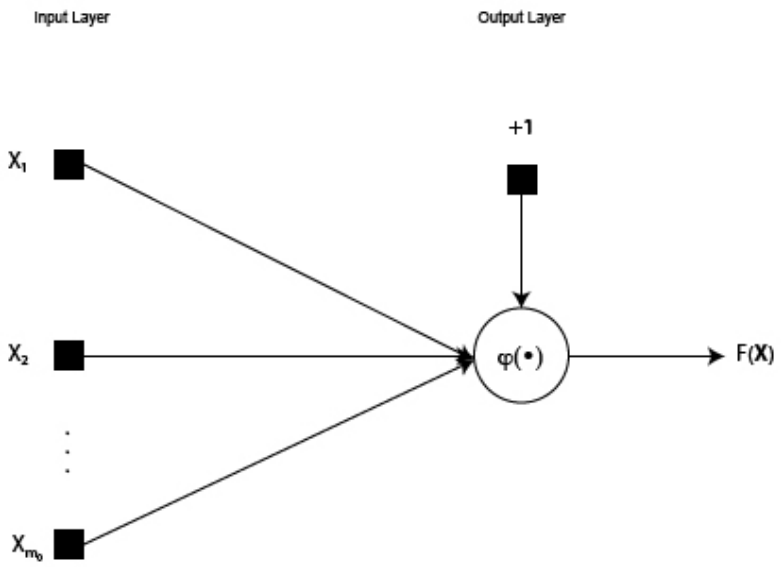


Figure 3.11: Perceptron model for one neuron.

The perceptron uses a weight vector, $\mathbf{w}$, and an input vector, $\mathbf{x}$, both of dimension $(m_0+1)\times 1$, where:

$$\mathbf{x} = [1 \; x_1 \; x_2 \; \cdots \; x_{m_0}]^T \tag{3.28}$$

$$\mathbf{w} = [w_0 \; w_1 \; w_2 \; \cdots \; w_{m_0}]^T \tag{3.29}$$

such that the classification boundary, $\mathbf{w}^T\mathbf{x} = 0$, separates the classes.

The output of the perceptron, given the input vector, can be calculated using Equation 3.30 which comes from the perceptron diagram in Figure 3.11. The activation function, $\varphi(v)$, chosen is the "tansig" function (Equation 3.31) because this function has both positive and negative outputs, and it provides a simple way to determine the classification using a threshold at 0. A negative output corresponds to a non-MC and a positive output corresponds to an MC. The choice of activation function is not crucial for this study since training methods are essentially the same for different activation functions.

$$F(\mathbf{x}) = \varphi(\mathbf{w}^T\mathbf{x}) \tag{3.30}$$

$$\varphi(v) = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)} \tag{3.31}$$

The perceptron, being a supervised learning method, requires training (or in other words the weight vector has to be determined). The weight vector is determined using the least means square (LMS) weight update. This method is stochastic, which means the converged weights are non-deterministic, and requires the continual presentation of training samples in order to determine the weight vector. The LMS weight update equation is specifically chosen for simplicity. This weight update equation is

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \eta\mathbf{x}(n)(d(n) - y(n)) \tag{3.32}$$

where $d(n)$ is the desired or expected output, and η is the learning rate parameter. Therefore, the weight vector is updated based on the error, or how far the current output is from the expect output. The desired output takes on a value of "+1", indicating a true positive; or "-1" indicating a true negative. The learning parameter determines how fast the weights converge. If the weights converge too quickly, then the weight vector may not be optimal.

3.5.3 Multilayer Perceptron

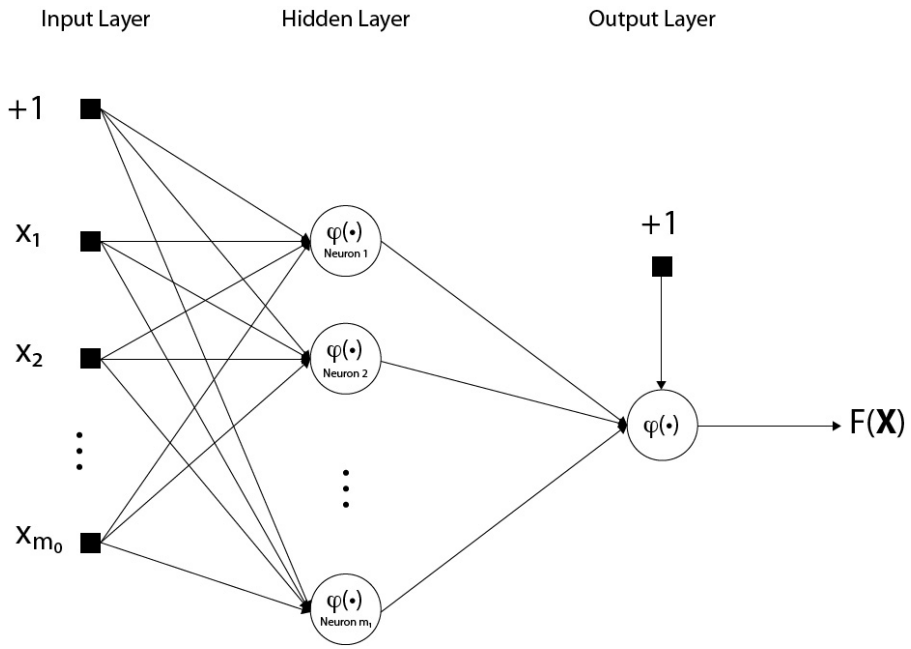


Figure 3.12: MLP model with one hidden layer and one output neuron.

The MLP is an extension from the the perceptron, but the difference is that the MLP allows for a nonlinear classification boundary. This type of ANN still provides a single output like the perceptron, however, this architecture has more neurons. The MLP typically contains layers of neurons between the inputs and outputs, called "hidden layers" [11]. A single hidden layer is considered because according to the universal approximation theorem, a given training set can be sufficiently approximated by a MLP with a single hidden layer [11].

Therefore, one hidden layer is sufficient for the purposes of this paper. The activation functions for every neuron (in both layers) are also chosen to be the "tansig" function (Equation 3.31), even though these activation functions maybe different.

The output of the MLP is computed using a network of neurons (Figure 3.12), where m_1 is the number of neurons in the hidden layer. Like the perceptron, the MLP has an input vector, $\mathbf{x}$ of dimension $(m_0 + 1) \times 1$. The MLP also has weights, however, now there is a weight matrix,

$\mathbf{W}$, with dimensions $(m_0 + 1) \times m_1$ where

$$\mathbf{W} = \begin{bmatrix} w_{10}^{(1)} & w_{20}^{(1)} & \cdots & w_{m_1 0}^{(1)} \\ w_{11}^{(1)} & w_{21}^{(1)} & \cdots & w_{m_1 1}^{(1)} \\ \vdots & \vdots & & \vdots \\ w_{1m_0}^{(1)} & w_{2m_0}^{(1)} & \cdots & w_{m_1 m_0}^{(1)} \end{bmatrix} \tag{3.33}$$

Each column of $\mathbf{W}$ represents a weight vector for a single neuron (or perceptron). The output of the hidden layer is

$$\mathbf{x}_{h_0} = \boldsymbol{\varphi}(\mathbf{W}^T \mathbf{x}) \tag{3.34}$$

where the vector notation for the activation function is $\boldsymbol{\varphi}(\mathbf{v}) = [\varphi(v_1)\ \varphi(v_2)\ \cdots\ \varphi(v_{m_1})]^T$. The vector output from the hidden layer, $\mathbf{x}_{h_0}$, is now used as part of the "input" to the output layer. The "input" to the output layer is the $m_1 + 1$ dimensional vector

$$\mathbf{x}_{h_1} = [1\ \mathbf{x}_{h_0}]^T \tag{3.35}$$

The output of the network can now be computed as

$$F(\mathbf{x}) = \varphi(\mathbf{w}_o^T \mathbf{x}_{h_1}) \tag{3.36}$$

where

$$\mathbf{w}_o = [w_{10}^{(2)}\ w_{11}^{(2)}\ \cdots\ w_{1m_1}^{(2)}]^T \tag{3.37}$$

The updating of the weights in this type of network is not as straight forward as that of the single neuron. In order to perform the weight update, back-propagation is used. Back-propagation, like the LMS weight update, is also an iterative method. The back-propagation algorithm is defined in [11].

3.5.4 Radial Basis Function Network

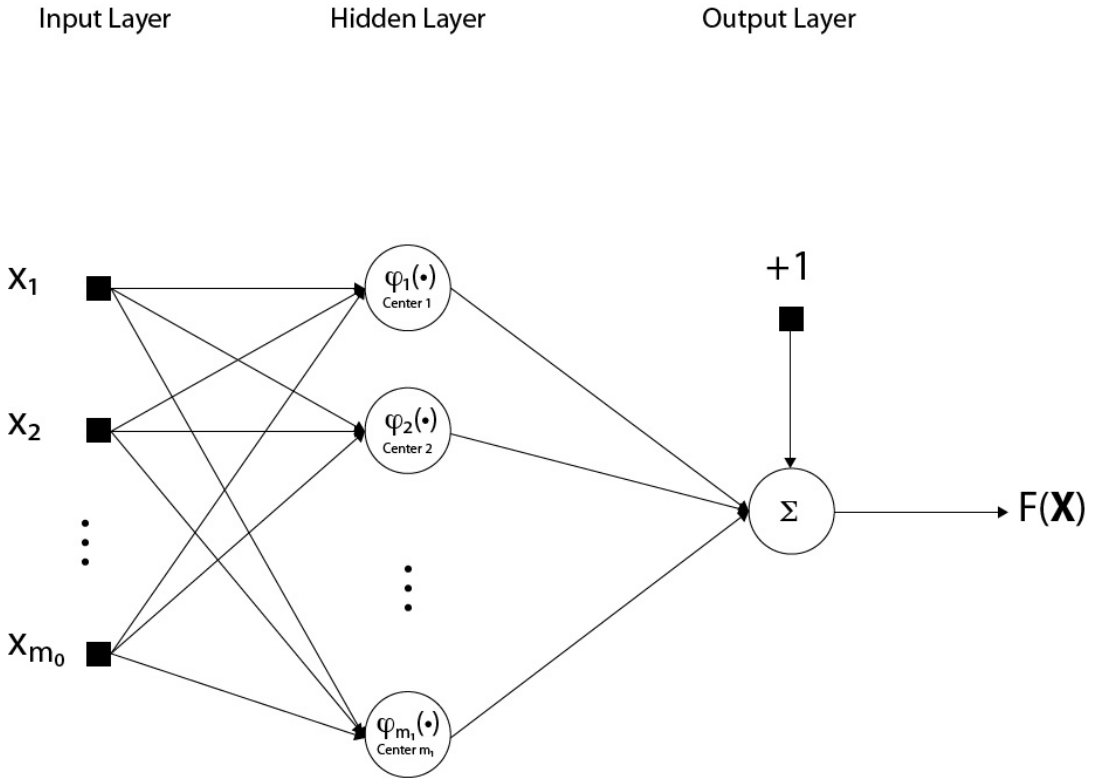


Figure 3.13: RBF network model with a single hidden layer and one output.

The RBF network is a nonlinear classification technique like the MLP, however, it relies on finding linear separability in a higher dimensional space. Based on Cover's theorem [5], the data is more likely to be linearly separable when cast nonlinearly into a higher dimensional space [11]. The RBF classifier has an input vector, $\mathbf{x}$, of dimensions $m_0 \times 1$. The inputs do not have any weights associated with them (i.e. weight are all "+1"). The inputs are just fed forward to the hidden layer. The hidden layer is composed of m_1 nodes, where $m_1 > m_0$ (i.e. the number nodes in the hidden must be larger than the dimension of the input vector). These nodes in the hidden layer are distinguish from the neurons of the MLP because these nodes do not function in the same way. These nodes are the collection of radial-basis functions for the network.

In order to determine the radial-basis functions, k-means clustering is first used to find the cluster centers. K-means simply finds "k" number of cluster centers for the data (both classes), it does not perform any type of classification. After k-means clustering, the cluster

centers are used to determine the radial-basis functions for the network. There are different forms of radial-basis functions, such as multiquadradics, inverse multiquadradics, and Gaussian functions. For the purposes of this paper, the Gaussian radial-basis function is used. The Gaussian radial-basis function for the jth center, $\mathbf{x}_j$, is

$$\varphi_j(\mathbf{x}) = \exp(-\frac{1}{2\sigma_j^2} \|\mathbf{x} - \mathbf{x}_j\|^2) \tag{3.38}$$

The output of the network $F(\mathbf{x}) = \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x})$ where

$$\boldsymbol{\phi}(\mathbf{x}) = [1 \ \varphi_1(\mathbf{x}) \ \varphi_2(\mathbf{x}) \ \cdots \ \varphi_{m_1}(\mathbf{x})]^T \tag{3.39}$$

$$\mathbf{w} = [w_0 \ w_1 \ \cdots \ w_{m_1}] \tag{3.40}$$

The weights for the RBF network are also non-deterministic, therefore the network has to be trained using samples. The weights are determined very much in the same way as the perceptron, since the only part of the network with weights is essentially a perceptron. Refer back to section 3.5.2 for details on updating the weights.

3.5.5 Support Vector Machine

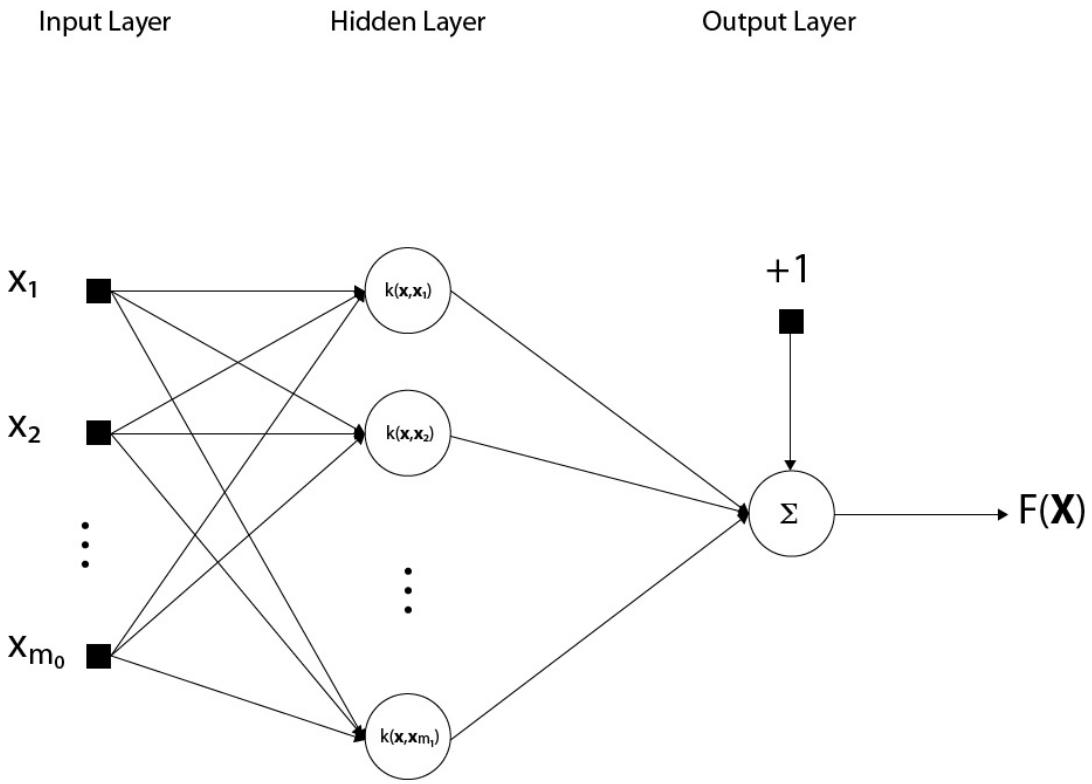


Figure 3.14: SVM model with one hidden layer and one output neuron.

SVMs are based around a simple objective function $Q(\boldsymbol{\alpha})$ shown in Equation 3.41 [11].

$$Q(\boldsymbol{\alpha}) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j d_i d_j k(\mathbf{x}_i, \mathbf{x}_j) \tag{3.41}$$

where d_i is the desired output for the ith sample. Given the training samples, find the Lagrange multipliers (α_i's) that maximize the objective function, subject to the constraints:

$$(1) \quad \sum_{i=1}^{N} \alpha_i d_i = 0$$

$$(2) \quad 0 \leq \alpha_i \leq C \quad \text{for } i = 1, 2, ..., N$$

This is the dual form for the constrained optimization of a support vector machine. All training samples with non-zero α_i's are the support vectors. There are fewer support vectors than the number of training samples.

The SVM uses a Gaussian kernel which takes the same form as the radial-basis function above in Equation 3.38, however, $\mathbf{x}_j$ represents the support vectors rather than the cluster centers. Equation 3.42 defines the Gaussian kernel [11].

$$k(\mathbf{x}, \mathbf{x}_i) = \exp(-\frac{1}{2\sigma^2} \|\mathbf{x} - \mathbf{x}_i\|^2) \tag{3.42}$$

The SVM output of the network is computed by feeding an input vector, $\mathbf{x}$, foraward through the network (Figure 3.14). The input vector, which is of dimensions $m_0 \times 1$, does not have any weights applied between the input and hidden layers. Next, the input vector is applied to the m_1 kernel, which come substituting the m_1 support vectors into Equation 3.42. The output layer is a single neuron (or perceptron), which means that weight are iteratively computed in the same manner. Thus, the output of the SVM is

$$F(\mathbf{x}) = \sum_{i=1}^{m_1} \alpha_i d_i k(\mathbf{x}, \mathbf{x}_i) \tag{3.43}$$

where d_i is the expected output for the ith support vector.

SVMs work well because they also apply Cover's theorem [5]. In order for the SVMs to work well the number of support vectors must be greater than the dimension of the input vector, or

$$m_1 > m_0 \tag{3.44}$$

Typically, for the training samples to be linearly separable in the m_1 dimensional space

$$m_1 \geq N/2 \tag{3.45}$$

where N is the number of samples used to train the SVM.

Chapter 4

Training Networks

In this chapter, the extraction of the training data from collections of images is discussed. The details for each training set is discussed, including the number of samples in the set, resolution of the mammograms, etc. Also, the selection methods for the samples corresponding MCs and those corresponding to non-MCs are discussed.

4.1 Training Sets

There are 2 different collections of images used to extract the training sets for MC detection:

1. DDSM (Digital Database for Screening Mammography) provided by the University of South Florida (USF): 40 images [12, 13]

2. Baptist Hospital at Wake Forest University (WFU): 80 images

The DDSM is a collection of digitized film mammograms. This collection of mammograms, in particular, are generated by exposing radiographic film to the x-rays (after passing through the breast tissue). Then, the mammograms are digitized using scanners. The film response, as noted in section 3.1, is nonlinear. Conversely, the WFU mammogram collection for are not digitized through a scanner. These mammogram, when created during screening process, first use a a scintillator to convert the x-rays to light. Then, the silicon detectors respond to the light emitted from the scintillator. This process directly captures the x-rays that are not absorbed in the tissue, and a computer generates the digital mammogram from the amount of x-rays detected. Thus, eliminating the effects of noise from the scanners (noise from detectors is still present). It is important to note that the response for digital mammography is very close to a linear response, which differs from the nonlinear response for film mammography.

4.1.1 DDSM

The training set for the DDSM database consists of 662 samples, 332 of which correspond to MCs and 330 that correspond to non-MCs, which come from mammograms that have a resolution of $50\mu m$ per pixel. The samples are split in half into training and test sets for each of the experiments, such that performance (see section 5.1) can be measured for both sets. The training set is composed of 331 randomly permuted samples, and the test set contains the remaining 331 samples. Each sample has every feature from Section 3.2 calculated for the corresponding region, and the expected output of the classifier when given the sample. An output corresponding to an MC is considered to be "+1", and one not corresponding to an MC is "-1".

4.1.2 WFU

The training set for the WFU database consists of 800 samples, 400 of which correspond to MCs and 400 that correspond to non-MCs. These samples come from a collection of mammograms which have a resolution of $100\mu m$ per pixel. These samples are also split into training and testing sets, so that performances can be measured (section 5.1). The same features: area, contrast, average brightness, brightness variance, second central moments, and entropy are given as part of each sample, as well as, the expected output.

4.2 Sampling

4.2.1 True Positive Sampling

The true positive samples used (sections 4.1.1 and 4.1.2) for the data sets are chosen manually because the classifiers must be trained to detect MCs and random sampling will train false positives to be classified as MCs. Each feature described above is calculated for the manually selected points in images that contain MCs from the DDSM and WFU mammogram sets. Note that many of the MCs selected are located within clusters already identified through the databases, such that the samples entered into the training set are less likely to be false. The features are then computed from the regions surrounding the corresponding points.

4.2.2 False Positive Sampling

The false positive samples are obtained in different manner than the true positive samples. Uniformly distributed random points using multiple images from the databases that are categorized as "normal" (meaning not malignant), are selected. Non-malignant mammograms may contain MCs that do not belong to clusters, however, by sampling a much smaller number

of points compared to the number of pixels in the mammogram, the probability of entering features corresponding to an MC in not likely. However, it is possible and thus the true negative training set must be accurately checked. Each random point has each feature computed so that the classifiers can determine how to classify regions not corresponding to MCs. Since the resolutions for each mammogram set are different, note that the two training sets are kept separate (i.e. the sampling is performed individually on the DDSM and WFU databases).

Chapter 5

Results

In this chapter, the performances of the MC detection methods are discussed. The overall performances of the MC detection methods depend on both the performance of the pre-processor and the performance of the classifier. The measures of performance are the true positive rates and false positive rates. First, the performance of the pre-processor (the hdome transform, and thresholding) alone is analyzed. Then, the overall performances through the various classifiers are measured.

5.1 Measuring Performance

The two measures of performance used are the true positive rate and the false positive rate. The true positive rate is a measure used to show percentage of MCs correctly classified, and the false positive rate is a measure used to show the percentage of non-MCs that are classified as MCs. Ideally, the true positive rate should be 100% and the false positive rate should be 0%. Realistically, the classifiers should be as close as possible to this ideal performance.

The true positive rate (TPR) is computed as

$$TPR = \frac{TP}{TP + FN} \tag{5.1}$$

where TP is the number of true positive samples (i.e. number MCs classified as MCs) and FN is the number of false negative samples (i.e. number of MCs classified as non-MCs). Then, the false positive rate (FPR) is computed as

$$FPR = \frac{FP}{FP + TN} \tag{5.2}$$

where FP is the number of false positive samples (i.e number of non-MCs classified as MCs) and TN is the number of true negative samples (i.e. number of non-MCs classified as MCs).

5.2 Pre-Processor Performance

In order to generalize the method, the performance of the pre-processor is measured on both sets of mammograms (DDSM and WFU). The true positive rates achieved for the pre-processor include:

- **DDSM** detects 95% of MCs
- **WFU** detects 93% of MCs

This shows only the percentage of the MCs that are correctly detected after the pre-processor. The classifier is needed in order to improve performance by reducing the large number of false positives remaining after the pre-processor.

5.3 Classifier Performance

5.3.1 Mahalanobis Distance

The Mahalanobis distance considers "closeness" to the means of the true positive samples and the true negatives samples, and the covariance matrices of both classes. The Mahalanobis distance classifier performances are tabulated in Table 5.1, which shows the minimum, maximum, and average (over 30 trials) true positive rate (TPR) and false positive rate (FPR) for both training and testing sets. Table 5.1 shows that that Mahalanobis distance classifier is not well suited for detecting MCs because many MCs are being missed and many maxima that are not MCs are classified as MCs.

Table 5.1: Performances of the Mahalanobis distance.

		Training	Testing
TPR	min	0%	0%
	max	100%	100%
	avg.	82.7%	38.7%
FPR	min	0%	0%
	max	33.33%	75%
	avg.	1.59%	38.86%

5.3.2 Perceptron

The perceptron classifier performance is not the best, which is expected since the classification limited to a linear boundary. Even though the perceptron classification on the test set shows decent performance, the results when classifying all the points output from the h-dome transform are not sufficient enough to detect all MC clusters. The perceptron classifier outputs too many false positive MCs. The perceptron classification performance is shown in Table 5.2, which shows the minimum, maximum, and average (over 30 trials) TPR and FPR for both training and testing sets.

Table 5.2: Performances of the perceptron.

		Training	Testing
TPR	min	97.7%	95.6%
	max	100%	100%
	avg.	99.6%	98.7%
FPR	min	0%	0.58%
	max	3.45%	6.02%
	avg.	1%	2.21%

5.3.3 Multilayer Perceptron

The MLP classifier provides varying performance, which is dependent upon the number of neurons in the hidden layer. Given sufficient number of neurons in the hidden layer (and enough training samples that are representative of MCs), the MLP provides a good classification. The MLP is able to correctly classify MCs with few false positives. The performances for the MLP classifier are shown in Table 5.3, which shows the minimum, maximum, and average (over 30 trials) TPR and FPR for training and testing sets. This table shows the performance for MLP classifiers with 5, 10, and 20 hidden neurons (HNs).

Table 5.3: Performances of the MLPs.

		5 HNs		10 HNs		20 HNs	
		Training	Testing	Training	Testing	Training	Testing
TPR	min	95.1%	94.6%	94.9%	88.61%	0%	0%
	max	100%	100%	100%	98.7%	100%	100%
	avg.	92.3%	91.1%	93.3%	91.1%	95.8%	93.3%
FPR	min	0%	0%	0%	1.15%	0%	0.61%
	max	6.62%	8.86%	5.92%	11.18%	49.0%	40.0%
	avg.	3.00%	4.56%	2.87%	5.74%	1.97%	5.43%

The MLP with 20 hidden neurons provides the best performance in training for the MLPs tested. However, the MLP with 5 hidden neurons performs the best on the test set (i.e. generalizes the best). The results in Table 5.3 show that as the number of neurons increases, the ability to generalize degrades.

5.3.4 Radial Basis Function Network

The RBF classifier provides varying performance as well, which is dependent on the number of cluster centers. The RBF provides really good results, given a sufficient number cluster centers and training samples. The RBF network is able to correctly classify MCs with minimal false positives and false negatives. The performances of the RBF networks are tabulated in Table 5.4, which show the minimum, maximum, and average (over 30 trials) TPR and FPR for both training and testing sets. This table shows these performance measures for RBF networks with 50, 100, and 150 cluster centers.

Table 5.4: Performances of the RBF networks.

		50 centers		100 centers		150 centers	
		Training	Testing	Training	Testing	Training	Testing
TPR	min	94.4%	88.7%	99.4%	91.1%	99.3%	82.5%
	max	99.4%	97.0%	100%	98.0%	100%	96.9%
	avg.	97.7%	92.5%	99.9%	94.0%	99.9%	89.7%
FPR	min	1.14%	3.11%	0%	4.73%	0%	7.32%
	max	7.19%	10.86%	2.61%	13.94%	1.30%	17.37%
	avg.	3.52%	7.68%	1.26%	9.14%	0.80%	11.08%

The RBF network with 150 cluster centers provides the best training performance for the RBF networks tested. However, the RBF network with 50 cluster centers generalizes the best. As the number of cluster centers increase, the ability to generalize degrades. The larger number of cluster centers provides near perfect classification on the training set. However, if a samples does not belong to the training set there is a higher probability that it will be misclassified.

5.3.5 Support Vector Machine

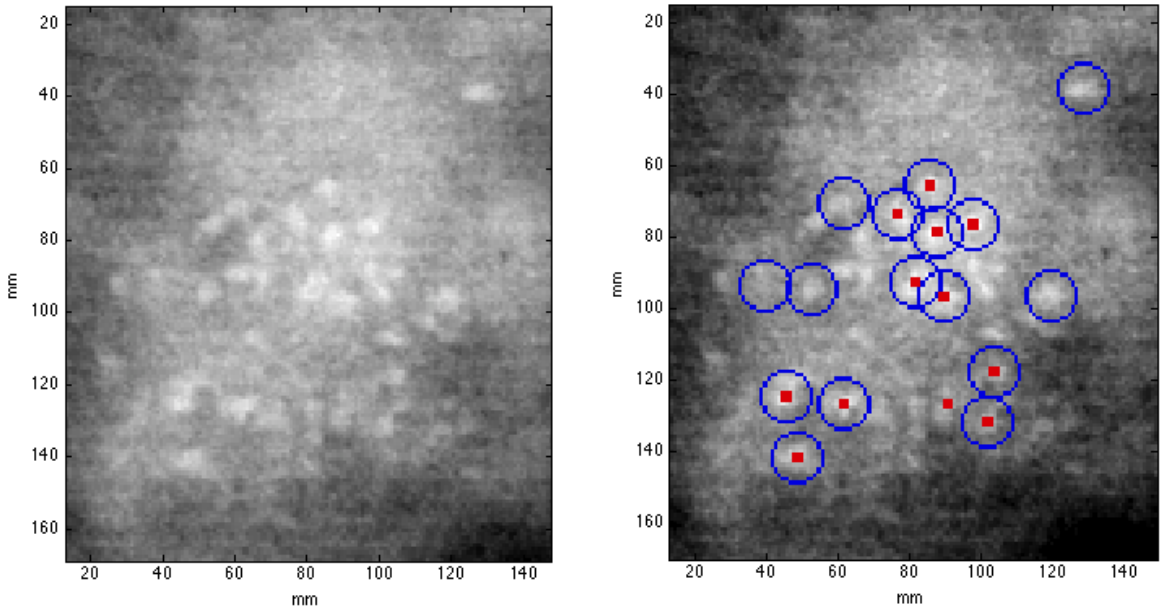
The SVM, in general, classifies well given a sufficient number of representative training samples. The performance of the SVM is dependent on the number support vectors and which support vectors are chosen. The performance measures are tabulate in Table 5.5, which shows the minimum, maximum, and average (over 30 trials) TPR and FPR for both training and testing sets. The SVM machine proves to classify very well on a set of given MCs (training set), however performance on the test set is not the highest out of the methods tested. Therefore, the SVM does not generalize the best.

Table 5.5: Performances of the SVM.

		Training	Testing
TPR	min	97.1%	90.2%
	max	100%	100%
	avg.	99.3%	94.0%
FPR	min	0%	3.2%
	max	9.21%	12.14%
	avg.	1.8%	7.90%

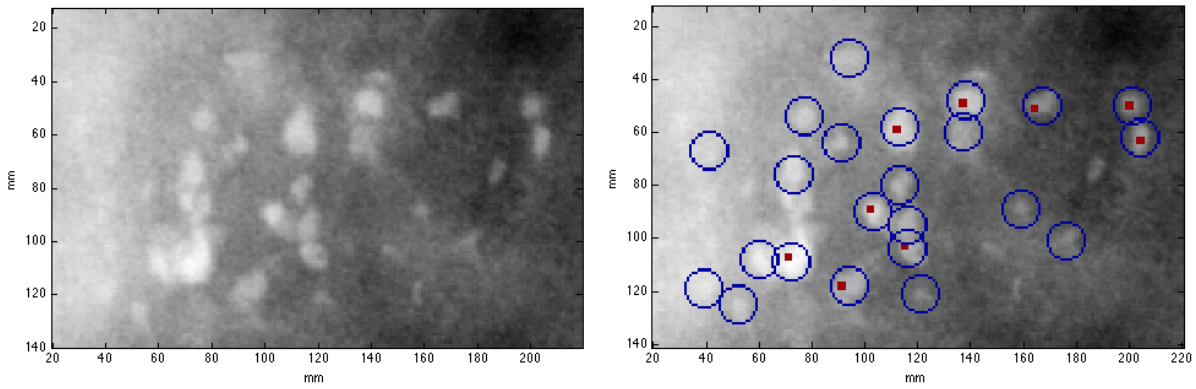
5.4 Detection Output

The major goal is to show that it is possible to detect MCs using the methods described in chapter 3. Figures 5.1 and 5.2 show two different MC clusters and the individual MCs that are detected within those clusters. The left image in each figure shows the cluster without any markings so that the MCs are clearly seen. The images on the right in each figure show the detected MCs with a blue circle and the MCs that are used in the training set with a small red square. These figures show that the method (pre-processor, segmenter, feature extractor, and classifier) is capable of detecting a majority of MCs in training set and those not in the training set.



(a) MC cluster. (b) Marked MC cluster.

Figure 5.1: Example detection of MCs.



(a) Another MC cluster. (b) Marked MC cluster.

Figure 5.2: Another example of detecting MCs

Figure 5.3 shows the output of a mammogram that was processed using the detection method with the SVM. The figure shows a correctly detected cluster of MCs. Clustering is

not explicitly discussed because it non-trivial to cluster a set of detected MC locations using k-means clustering or some other simple clustering algorithm. The detected MCs are better shown by using clustering, since a cluster is easier to spot than individual MCs. Therefore, the detection method can be shown to correctly identify a cluster of MCs correctly, although it may have missed some individual MCs.

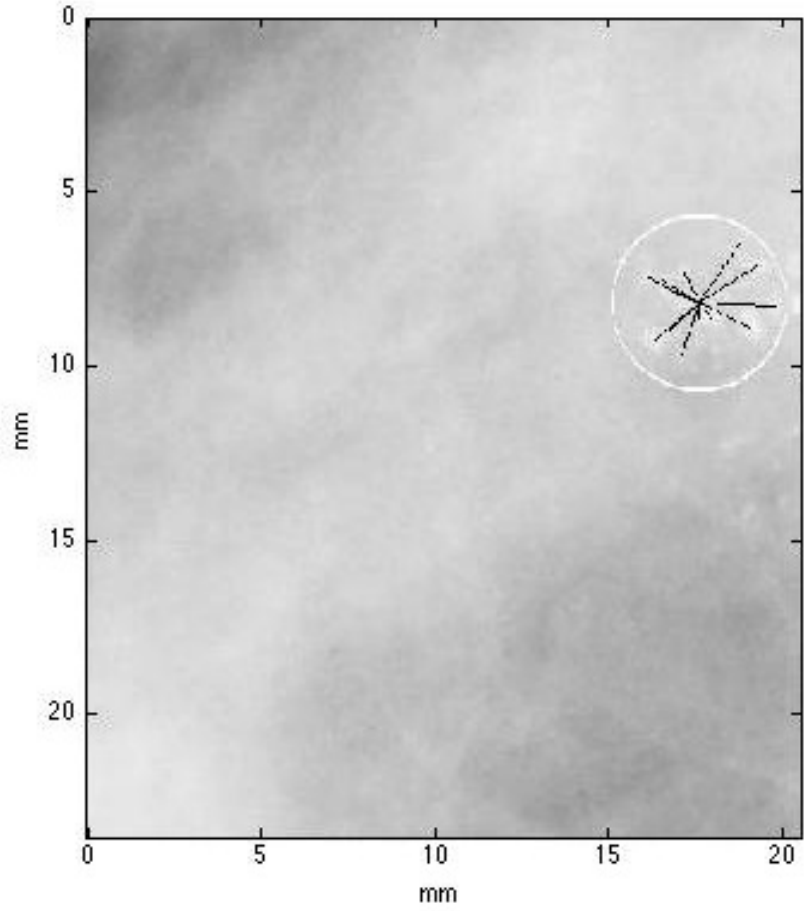


Figure 5.3: Detection using SVM (with clustering).

5.5 Missed Microcalcifications

The reality is that detecting MCs is very complex, and achieving a 100% detection rate with zero false positives is unrealistic. Understanding the reasons why the method does not detect certain MCs is key for understanding MC detection. There are at least two different explanations for

why some MCs were not detected:

1. The MCs are too subtle.

2. The training set is too small (or not representative of MCs)

The MCs are considered too subtle if the pre-process does not identify a peak at the location. If the pre-processor does not identify the peak, then the classifier never considers that point. In this case, this is a fault with the pre-processing technique not the classifier. Figure 5.4 shows an example cluster of subtle MCs, and Figure 5.5 shows the corresponding pre-processing output.

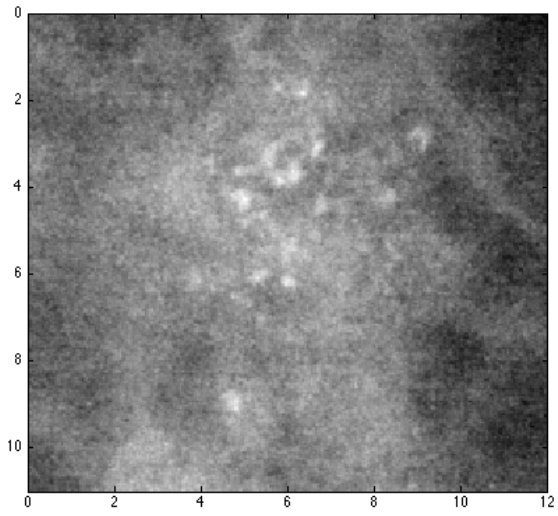


Figure 5.4: A cluster of subtle MCs.

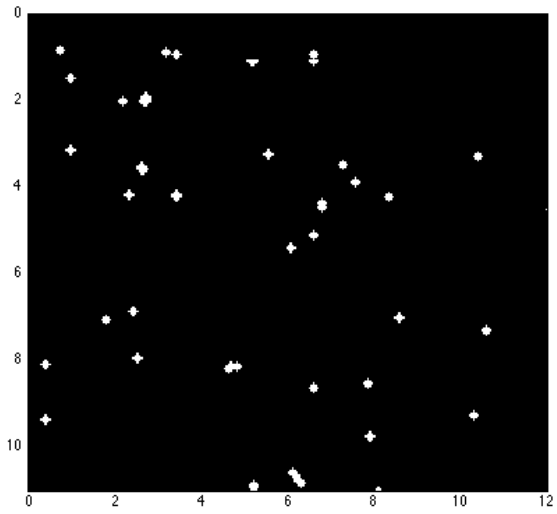


Figure 5.5: Pre-processing output.

Notice how the majority of MCs in the cluster are not found by the preprocessor; this cluster shows a few example of subtle MCs. One way to prevent this type of occurrence is to make the pre-processor more sensitive to maxima. However, the trade off is that more false positives may be detected with the classifier.

For the classifiers to generalize well, a sufficiently large training set is needed. More specifically, a training set that is representative of MCs and normal tissue. In either case, the classifier may not generalize well (i.e. samples that are not in the training set may be misclassified.

Chapter 6

Features

The MC features (area, contrast, average brightness, brightness variance, second central moments, and entropy) are analyzed in two ways, individually and pairwise, to better understand the intricacies of MC detection. The individual analysis includes measuring the mean and variance of each feature, and the pairwise analysis includes determining the separability between the classes (MCs and non-MCs) in the 2-dimensional spaces given by the projections of the 7-dimensional samples. The individual analysis provides the expected values of the features and insight into how each feature varies. Then, the pairwise analysis shows features that are approximately linearly separable and those that are not separable. This shows the more discriminatory features and feature pairs. Lastly, the covariance matrix is analyzed so that the feature dependencies can be seen more clearly.

6.1 Feature Statistics

6.1.1 True Positives

The statistics of the features (i.e., sample mean and sample standard deviation) provide the overall expectation of each MC feature. The sample mean

$$\mu_t(i) = \frac{1}{N_t} \sum_{k=1}^{N_t} F_i(k) \tag{6.1}$$

where N_t is the number of MCs and F_i is the value for the ith feature. Then, the sample standard deviation

$$\sigma_t(i) = \sqrt{\frac{1}{N_t} \sum_{k=1}^{N_t} (F_i(k) - \mu_t(i))^2} \tag{6.2}$$

is computed for each feature. Table 6.1 shows the sample mean and standard deviation of each feature from the set of MCs.

Table 6.1: Feature means and standard deviation for true positives.

Feature	μ_t	σ_t
Area	79.0	37.3
Contrast	142.5	73.3
Avg. Brightness	3313.7	274.2
Brightness Variance	4345.1	3971.3
μ_{20}	10.5	4.7
μ_{02}	5.0	2.7
Entropy	-31.3	29.3

Table 6.1 provides an overall sense of what an MC is in terms of the features defined in 3.4. The values in this table can be compared with the same statistics, mean and standard deviation, of the the non-MCs (Table 6.2). By comparing the two tables, we see the general differences between MCs and non-MCs. Notice the large overlap for many of the features, which means any one feature, alone, may not distinguish between MCs and non-MCs very well.

6.1.2 True Negatives

In this section, the same feature statistics, mean and standard deviation, are measured for the true negatives (non-MCs) so that the non-MCs can be contrasted with the MCs. The sample mean and standard deviation equations for the true negatives are very similar to those for the true positives. The equation for computing the sample mean is given by

$$\mu_f(i) = \frac{1}{N_f} \sum_{k=1}^{N_f} F_i(k) \tag{6.3}$$

where, conversely, N_f is the number of non-MCs. Then, the sample standard deviation of the true negatives is

$$\sigma_f(i) = \sqrt{\frac{1}{N_f} \sum_{k=1}^{N_f} (F_i(k) - \mu_f(i))^2} \tag{6.4}$$

The only difference between these equations and those in the previous section is the set of samples used for the computation. Table 6.2 shows the sample mean and standard deviation

of each feature from the set of true negatives.

Next, the MCs and non-MCs are contrasted, and conclusions are drawn from the features themselves. Some generalities about the features are seen when comparing Table 6.1 and Table 6.2, including: real MCs tend to have a larger size, contrast, and average brightness; also a different shape and texture. However, due to the large sample variances and small differences between the sample means for each feature, it is seen that simple feature thresholding does not yield the best classification. There exists a significant overlap between the MC samples and the non-MC samples for most features, thus making MC detection more difficult.

Table 6.2: Feature means and standard deviation for true negatives.

Feature	μ_f	σ_f
Area	37.7	25.3
Contrast	-29.6	155.2
Avg. Brightness	2556.0	643.4
Brightness Variance	4782.2	25596.6
μ_{20}	7.7	4.7
μ_{02}	2.5	2.0
Entropy	-20.3	57.8

6.1.3 Normalized Statistics

Tables 6.1 and 6.2 show that the average brightness and brightness variance features are dominating the other features, since the values of average brightness and brightness variance are significantly larger than the values of the other features. A dominating feature in a feature vector causes difficulty in determining the relationships between features. Therefore, standard practice using feature vectors that have one or more dominating features is to normalize the features by their standard deviation (or variance).

In this case, the samples are normalized by the sample standard deviation of the entire data set (i.e both MCs and non-MCs). Table 6.3 shows the feature means and variances of the normalized MC samples, and Table 6.4 shows the same statistics for the normalized non-MC samples.

Table 6.3: Normalized feature means and standard deviation for true positives.

Feature	μ_t	σ_t
Area	2.081	0.984
Contrast	0.956	0.492
Avg. Brightness	5.309	0.439
Brightness Variance	0.236	0.216
μ_{20}	2.130	0.954
μ_{02}	1.848	1.004
Entropy	-0.676	0.632

Table 6.4: Normalized feature means and standard deviation for true negatives.

Feature	μ_f	σ_f
Area	0.993	0.666
Contrast	-0.199	1.041
Avg. Brightness	4.095	1.031
Brightness Variance	0.260	1.391
μ_{20}	1.557	0.962
μ_{02}	0.914	0.749
Entropy	-0.438	1.250

These tables still provide the same general information between MCs and non-MCs. MCs typically have larger area, contrast, and average brightness, which are the same relationships observed from Tables 6.1 and 6.2. However, notice that the features are now roughly on the same scale.

6.2 Feature Pair Analysis

Next, the features are analyzed pairwise, meaning each sample is projected from 6-dimensions to the 2-dimensional feature space formed by two different features. A scatter plot is generated for every pair of features, thus there are 21 plots generated. When the samples are plotted (both MCs and non-MCs), the separability, or non-separability, of the true positives and true negatives is observed. Thus, the plots show whether the classes are approximately linearly

separable or non-separable.

The definition of linear and nonlinear used here is empirical; if the two classes are clearly discriminated by a single straight line in the 2-dimensional projection, we say that they are linearly separable. If more than 92% of samples are classified correctly (equivalent to less than 8% misclassified) by a single straight line, then we say they are approximately linearly. separable. Approximately linearly separable implies that the data does not have to be completely linearly separable (i.e 100% correct classification), therefore the classification rate constraint is relaxed slightly.

6.2.1 Approximately Linearly Separable

First, the approximately linearly separable MC feature pairs are discussed. The following feature pairs display approximate linear separation:

- Contrast vs. area (Figure 6.1).
- Entropy vs. contrast (Figure 6.2).
- Mean brightness vs. contrast (Figure 6.3).
- μ_{02} vs. contrast (Figure 6.4).
- μ_{20} vs. contrast (Figure 6.5).
- Brightness variance vs. contrast (Figure 6.6).
- Brightness variance vs. mean brightness (Figure 6.7).

Analysis of these plots reveals much about the usefulness of specific feature pairs. For example, consider contrast and any other feature. Any plot showing contrast (Figures 6.1–6.6), reveals that the true positive and true negative samples are "close" to being linearly separable. More importantly, these plots show that contrast is an important feature for classifying MCs.

The contrast, alone, is a very discriminating feature, since a simple scalar contrast threshold almost classifies all the MCs correctly. However, contrast is not the only example of an approximately linearly separable feature, but most are not as discriminating as contrast.

Although most feature pairs are not as discriminatory as contrast, there is only one other feature pair the displays an approximate linear separation between the classes; that is brightness variance and average brightness.

All of the MC feature pairs listed above, when plotted, reveal an approximate linear boundary between the classes. These relationships help explain the fundamental defining features of MCs (e.g. contrast), thus providing simple models to help detect MCs. The analysis shows

that average brightness, brightness variance, and contrast all are useful features due to their ability to discriminate between MCs and non-MCs well. It is seen that these plots discriminate better than many other feature pairs. For example, comparing Figure 6.7 and Figure 6.18, we observe that the classes overlap much worse for μ_{02} vs. area than for brightness variance vs. average brightness.

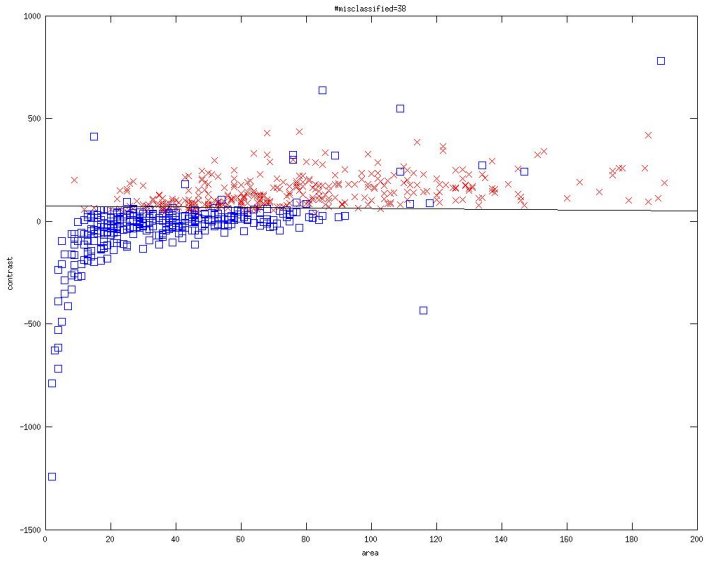


Figure 6.1: Contrast vs. area: 5.7% misclassification rate.

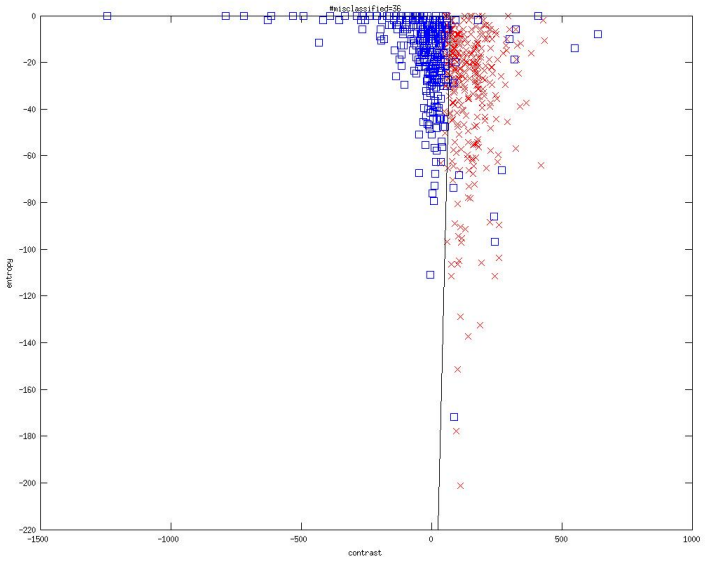


Figure 6.2: Entropy vs. contrast: 5.4% misclassification rate.

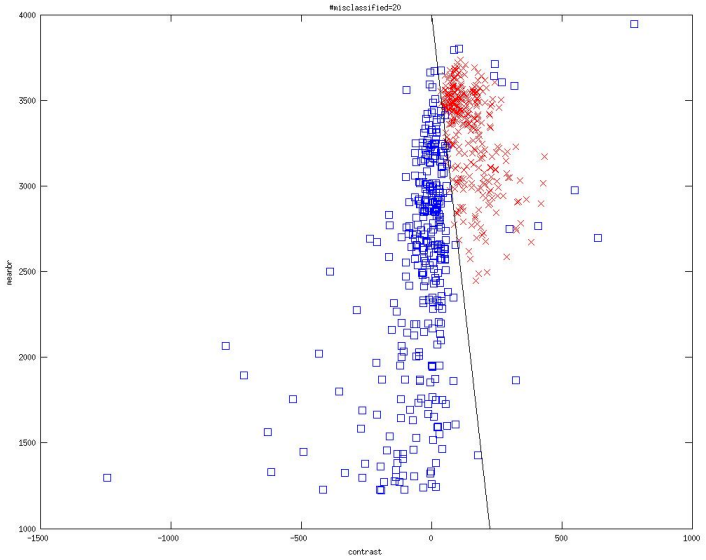


Figure 6.3: Mean Brightness vs. contrast: 3% misclassification rate.

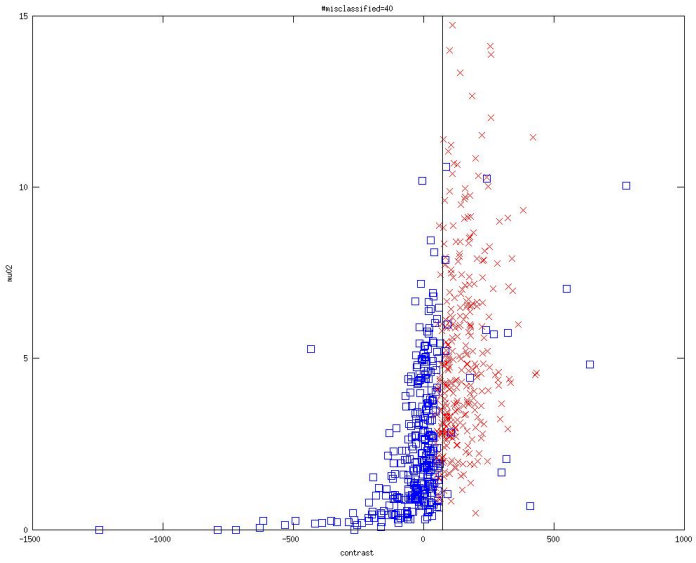


Figure 6.4: μ_{02} vs. contrast: 6% misclassification rate.

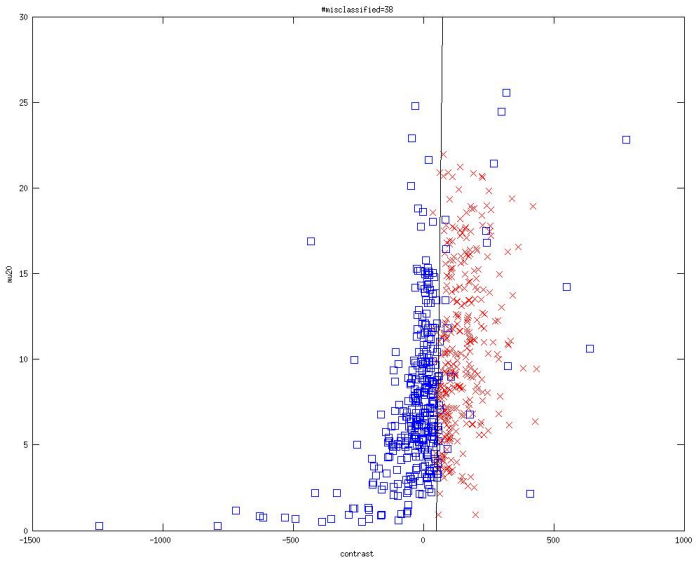


Figure 6.5: μ_{20} vs. contrast: 5.7% misclassification rate.

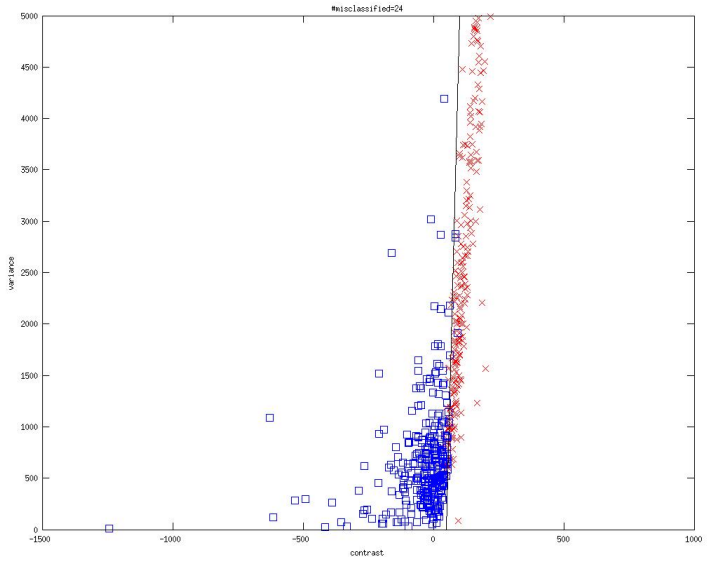


Figure 6.6: Brightness variance vs. contrast: 3.6% misclassification rate.

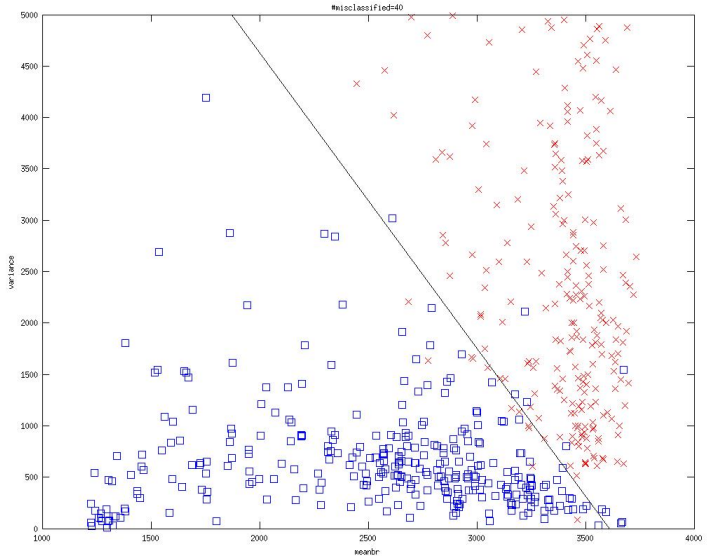


Figure 6.7: Brightness variance vs. mean brightness: 6% misclassification rate.

6.2.2 Non-separable

Next, the nonlinearly separable features pairs are discussed. The following feature pairs display no separation between the classes:

- Brightness variance vs. area (Figure 6.8).
- Entropy vs. brightness variance (Figure 6.9).
- μ_{02} vs. brightness variance (Figure 6.10).
- μ_{20} vs. brightness variance (Figure 6.11).
- Mean brightness vs. area (Figure 6.12).
- Entropy vs. mean brightness (Figure 6.13).
- μ_{02} vs. mean brightness (Figure 6.14).
- μ_{20} vs. mean brightness (Figure 6.15).
- μ_{02} vs. μ_{20} (Figure 6.16).
- Entropy vs. area (Figure 6.17).
- μ_{02} vs. area (Figure 6.18).
- μ_{20} vs. area (Figure 6.19).
- Entropy vs. μ_{02} (Figure 6.20).
- Entropy vs. μ_{20} (Figure 6.21).

This analysis identifies the pairs of features that do have any separation between the classes, and it identifies the feature pairs that are more discriminating than others. The scatter plots in Figures 6.8–6.21 all show the feature pairs with no separation or boundary between MCs and non-MCs. Consider the remaining plots comparing features with brightness variance (Figures 6.8–Figure 6.11). These plots reveal that brightness variance is a useful feature because the MCs and non-MCs are, to some degree, distinct. There is still a significant overlap between MCs and non-MCs, implying that the classes cannot be separated by a line, but the distributions of MCs and non-MCs are farther apart comparatively.

Again, compare Figures 6.8–Figure 6.11 with the rest of the nonlinear plots. Notice, most of the plots are not as discriminating as the plots with brightness variance. However, this does not mean that the other features, or feature pairs, are not important. Remember, these plots show the projections from a 6-dimensional space. In the 6-dimensional space, the classes may

be separable (nonlinearly or linearly). In order to better understand the interrelations among these feature, the covariance matrices may be analyzed for both the MCs and non-MCs.

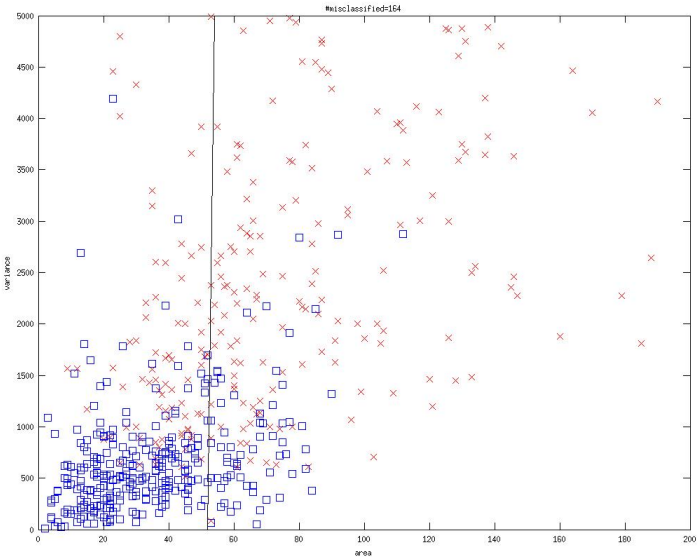


Figure 6.8: Brightness variance vs. area: 24.8% misclassification rate.

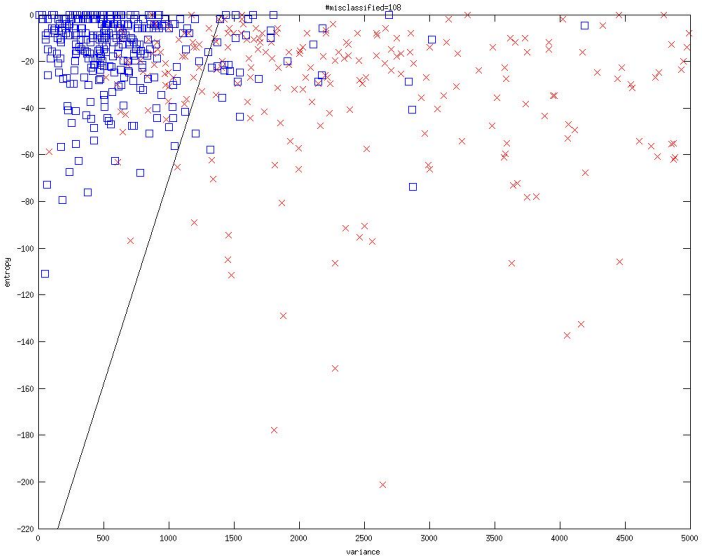


Figure 6.9: Entropy vs. brightness variance: 16.3% misclassification rate.

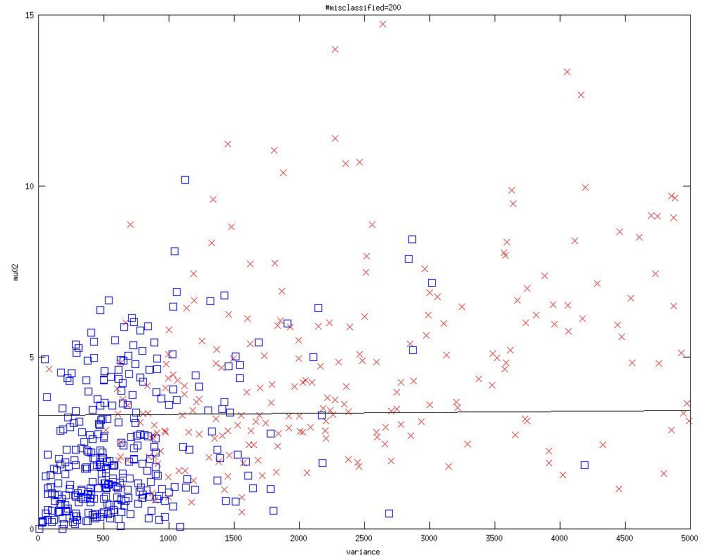


Figure 6.10: μ_{02} vs. brightness variance: 30.2% misclassification rate.

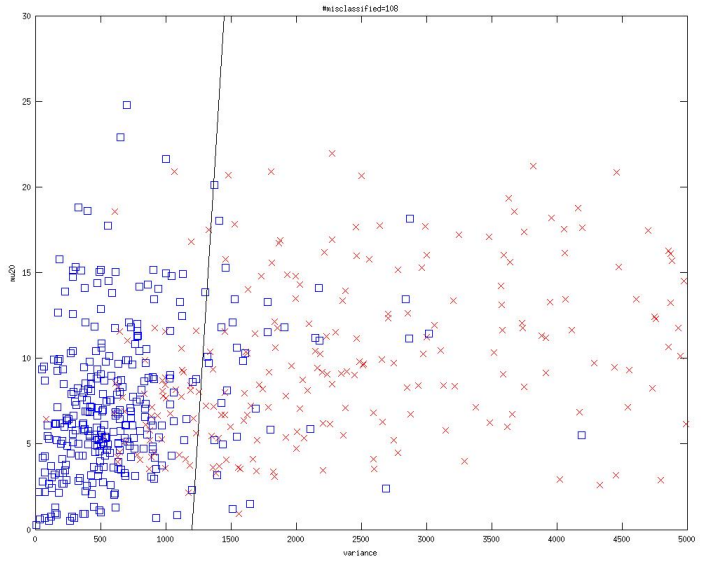


Figure 6.11: μ_{20} vs. brightness variance: 16.3% misclassification rate.

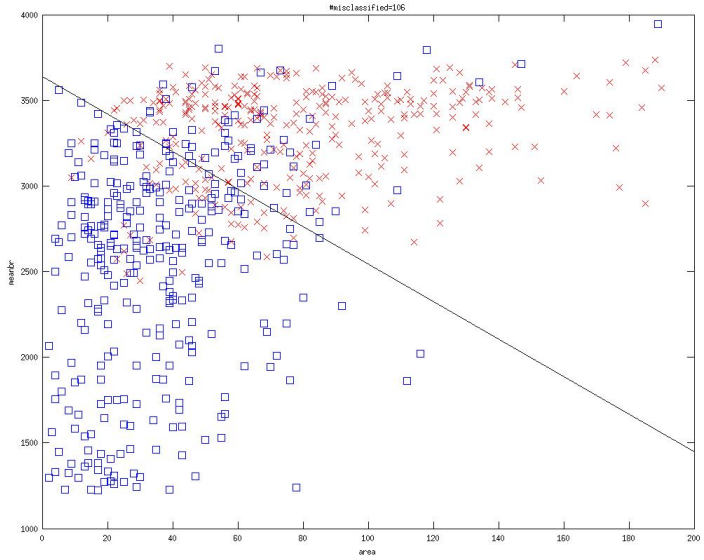


Figure 6.12: Mean brightness vs. area: 16% misclassification rate.

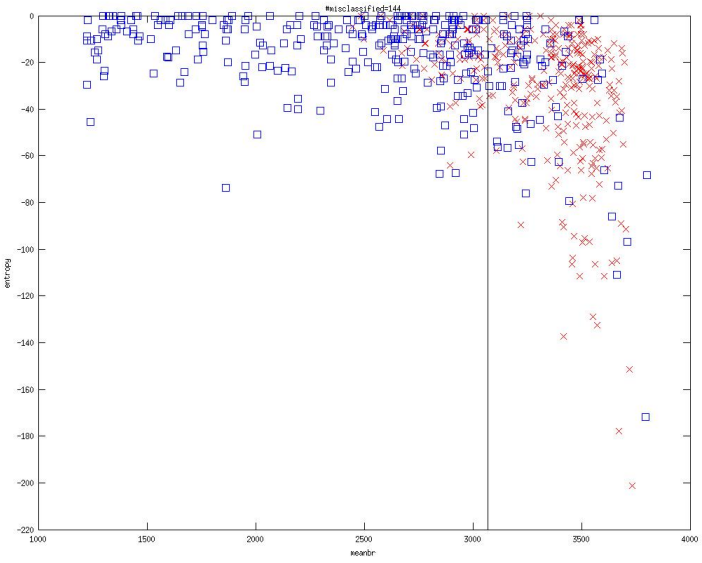


Figure 6.13: Entropy vs. mean brightness: 21.7% misclassification rate.

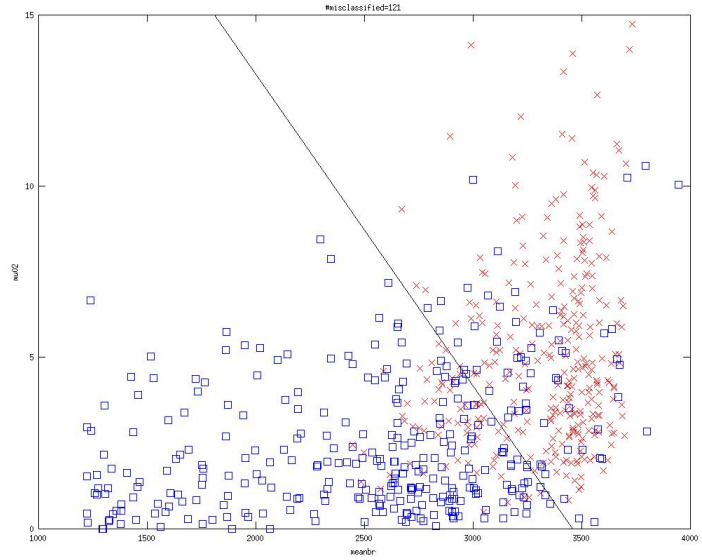


Figure 6.14: μ_{02} vs. mean brightness: 18.3% misclassification rate.

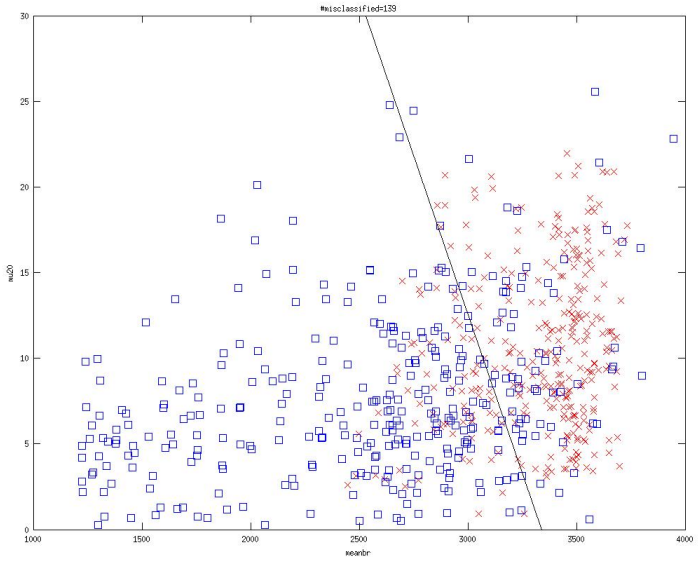


Figure 6.15: μ_{20} vs. mean brightness: 21% misclassification rate.

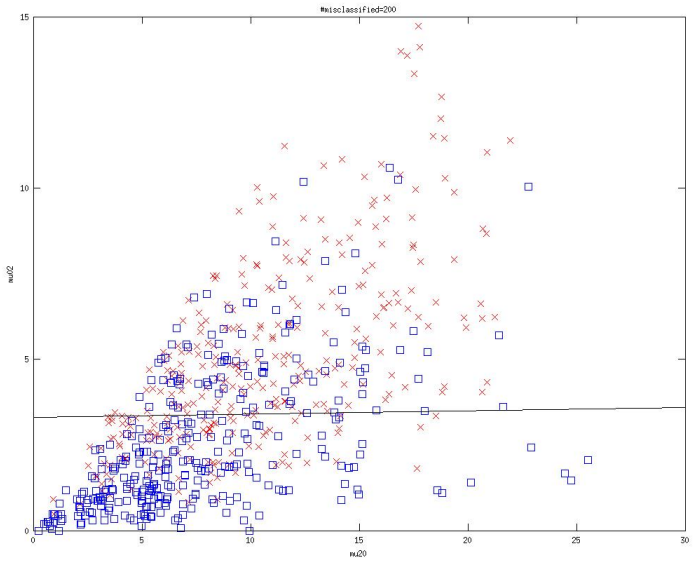


Figure 6.16: μ_{02} vs. μ_{20}: 30.2% misclassification rate.

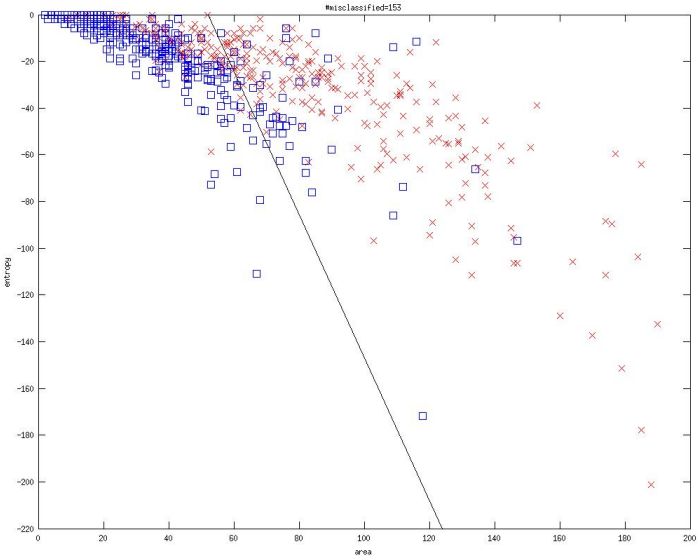


Figure 6.17: Entropy vs. area: 23.1% misclassification rate.

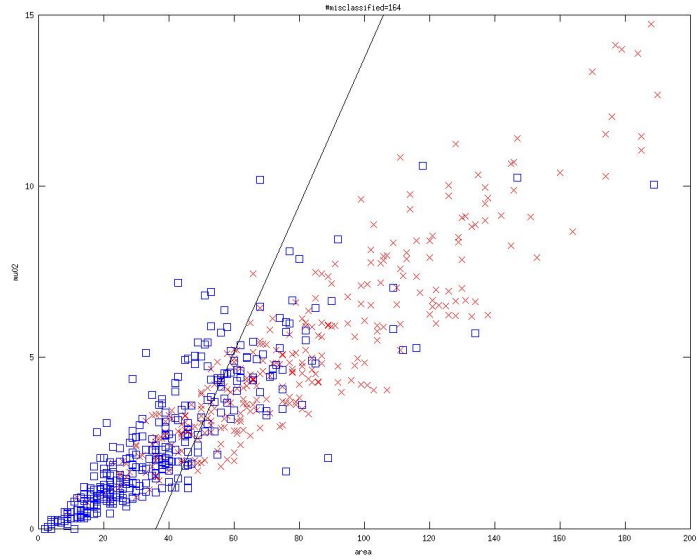


Figure 6.18: μ_{02} vs. area: 24.8% misclassification rate.

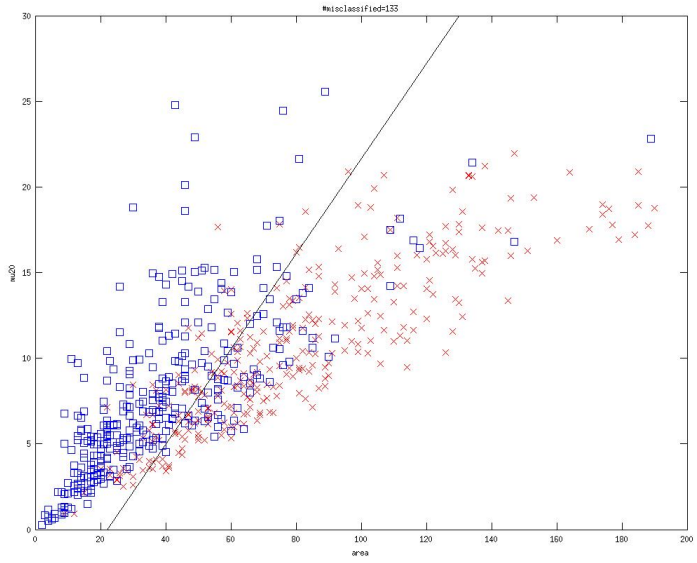


Figure 6.19: μ_{20} vs. area: 20.1% misclassification rate.

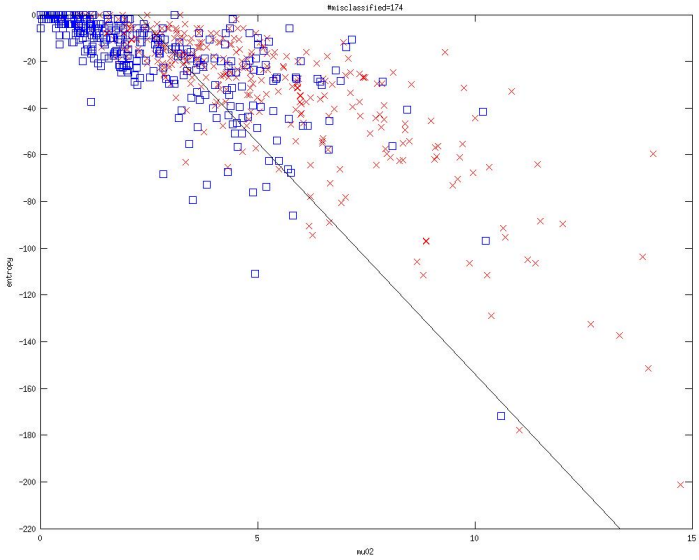


Figure 6.20: Entropy vs. μ_{02}: 26.3% misclassification rate.

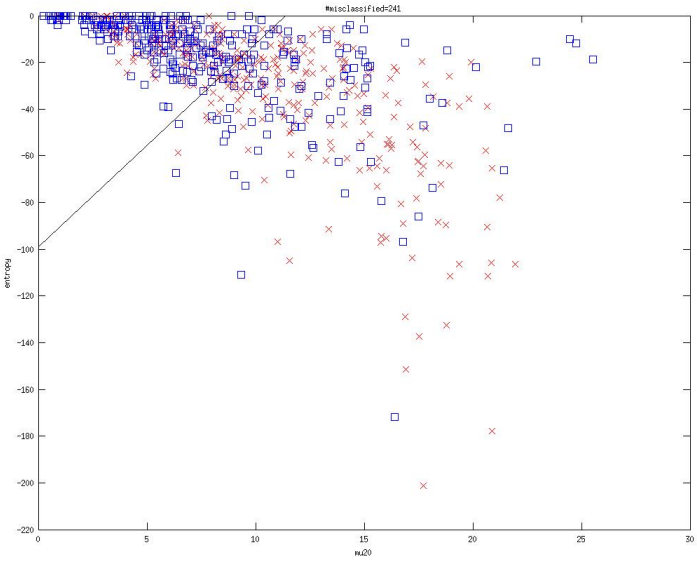


Figure 6.21: Entropy vs. μ_{20}: 36.4% misclassification rate.

6.2.3 Normalized Feature Plots

Since the features were normalized in section 6.1.3, it seems necessary to also analyze the normalized feature pairs. This analysis simply reveals the feature pairs that approximately have a linear boundary, and the features (or feature pairs) that are more discriminating than others.

The plots that display approximately linear separation include the following feature pairs:

- Contrast vs. area (Figure 6.22).
- Entropy vs. contrast (Figure 6.23).
- Mean brightness vs. contrast (Figure 6.24).
- μ_{02} vs. contrast (Figure 6.25).
- μ_{20} vs. contrast (Figure 6.26).
- Brightness variance vs. contrast (Figure 6.27).
- Brightness variance vs. mean brightness (Figure 6.28).

Note, these feature pairs have not changed from section 6.2.1, which means that no additional information is gained. The only difference between the normalized plots and the unnormalized

plots is the scaling of the axes. The plots reiterate the fact that these feature pairs almost have a linear boundary separating the MCs and non-MCs, and that contrast, average brightness, and brightness variance are descriptive features for MC detection.

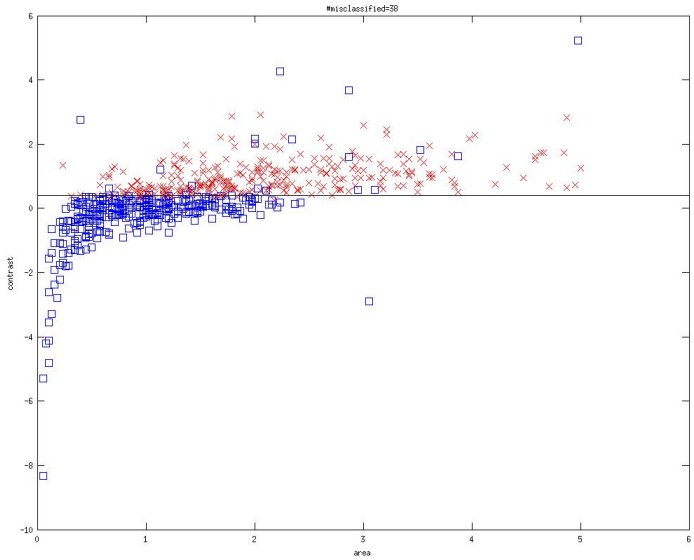


Figure 6.22: Contrast vs. area: 5.7% misclassification rate.

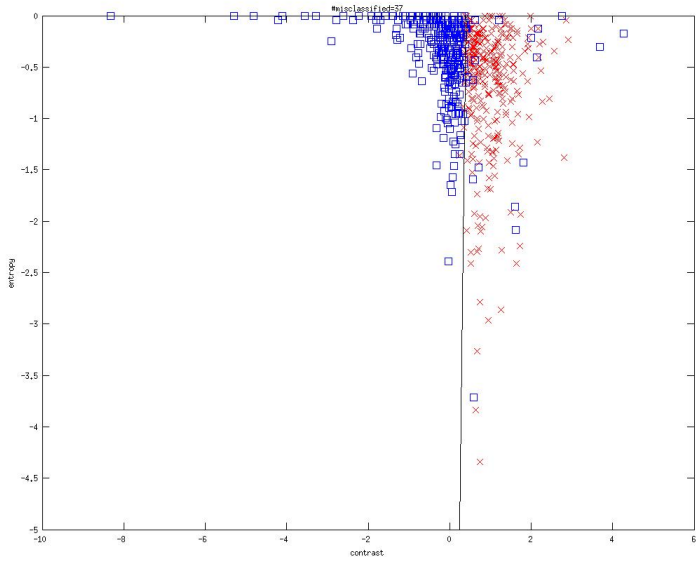


Figure 6.23: Entropy vs. contrast: 5.6% misclassification rate.

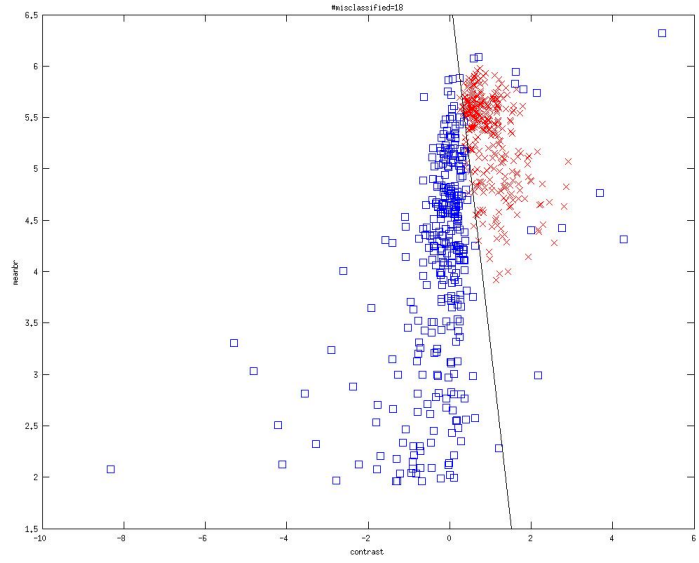


Figure 6.24: Mean Brightness vs. contrast: 2.7% misclassification rate.

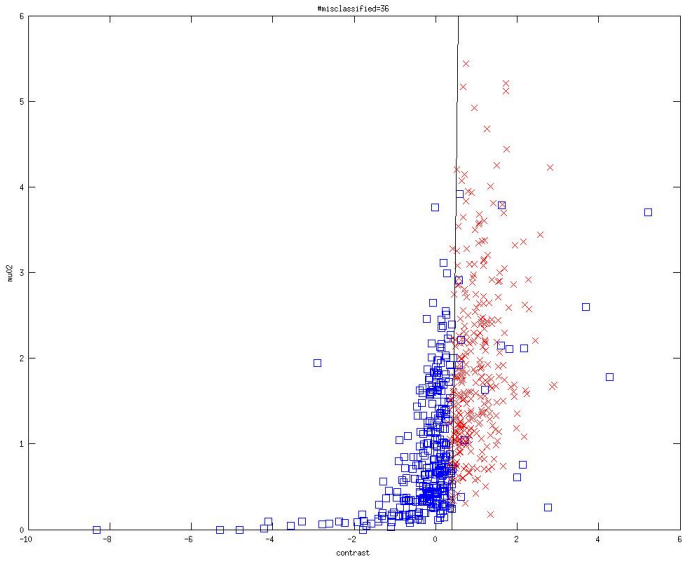


Figure 6.25: μ_{02} vs. contrast: 5.4% misclassification rate.

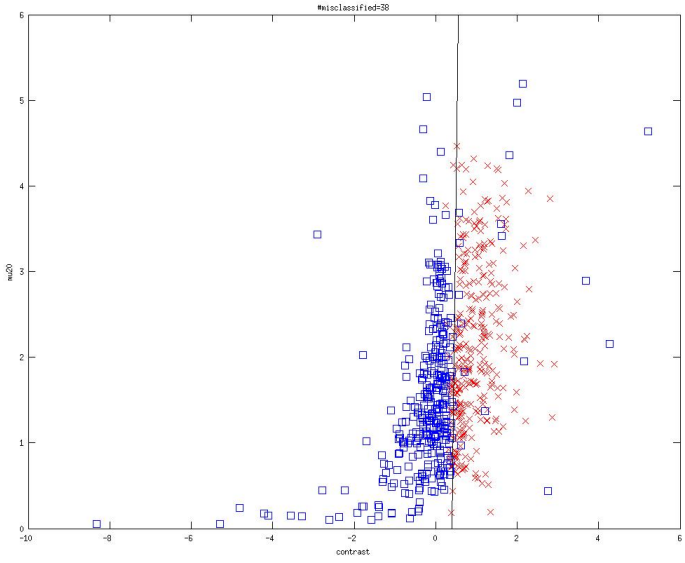


Figure 6.26: μ_{20} vs. contrast: 5.7% misclassification rate.

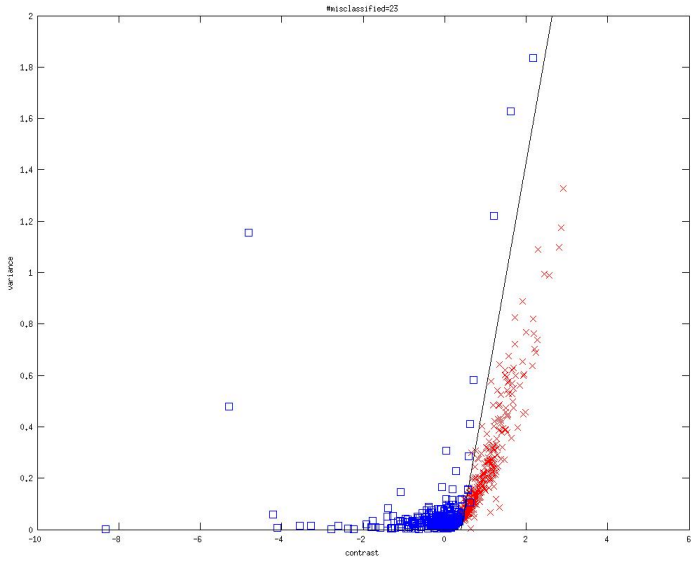


Figure 6.27: Brightness variance vs. contrast: 3.5% misclassification rate.

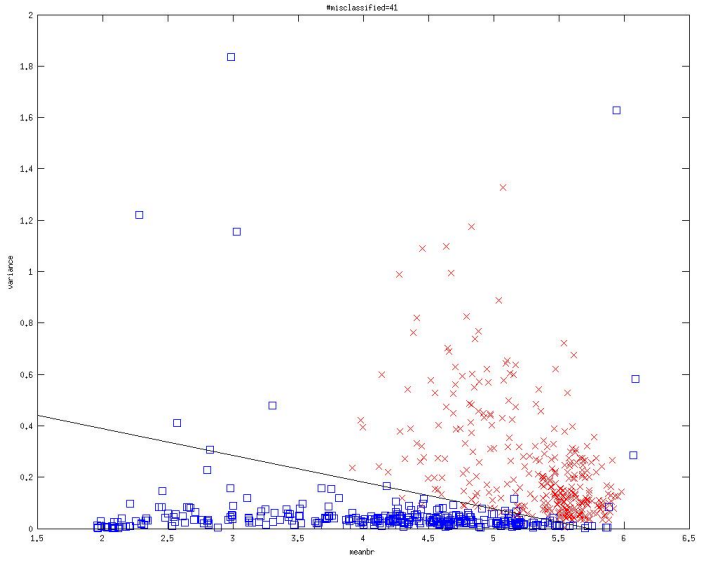


Figure 6.28: Brightness variance vs. mean brightness: 6.2% misclassification rate.

The feature pairs that are non-separable also have not changed from section 6.2.2. The non-separable feature pairs include:

- Brightness variance vs. area (Figure 6.29).
- Entropy vs. brightness variance (Figure 6.30).
- μ_{02} vs. brightness variance (Figure 6.31).
- μ_{20} vs. brightness variance (Figure 6.32).
- Mean brightness vs. area (Figure 6.33).
- Entropy vs. mean brightness (Figure 6.34).
- μ_{02} vs. mean brightness (Figure 6.35).
- μ_{20} vs. mean brightness (Figure 6.36).
- μ_{02} vs. μ_{20} (Figure 6.37).
- Entropy vs. area (Figure 6.38).
- μ_{02} vs. area (Figure 6.39).
- μ_{20} vs. area (Figure 6.40).
- Entropy vs. μ_{02} (Figure 6.41).
- Entropy vs. μ_{20} (Figure 6.42).

Again, no additional information is gained from these plots. The plots reiterate that variance is a more discriminating feature comparatively, and that very little can be determined from the remaining feature pairs. Although many of the features are not separable, this does not exclude them from being useful for MC detection because we are projecting the 6-dimensional data into a 2-dimensional space.

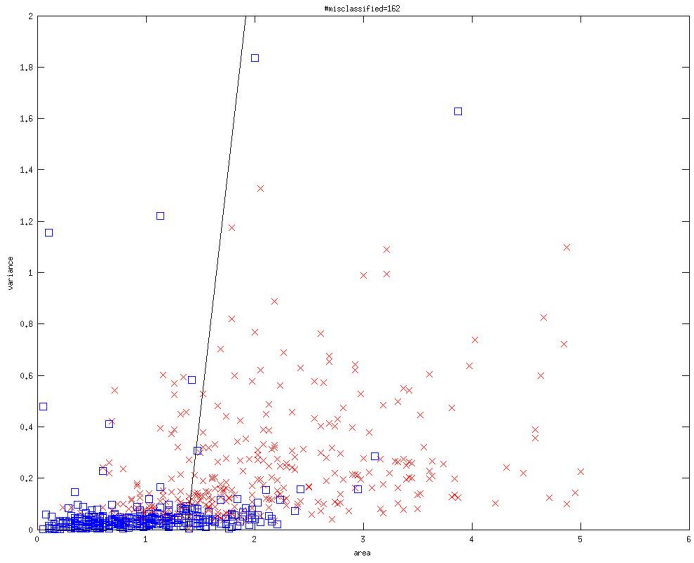


Figure 6.29: Brightness variance vs. area: 24.5% misclassification rate.

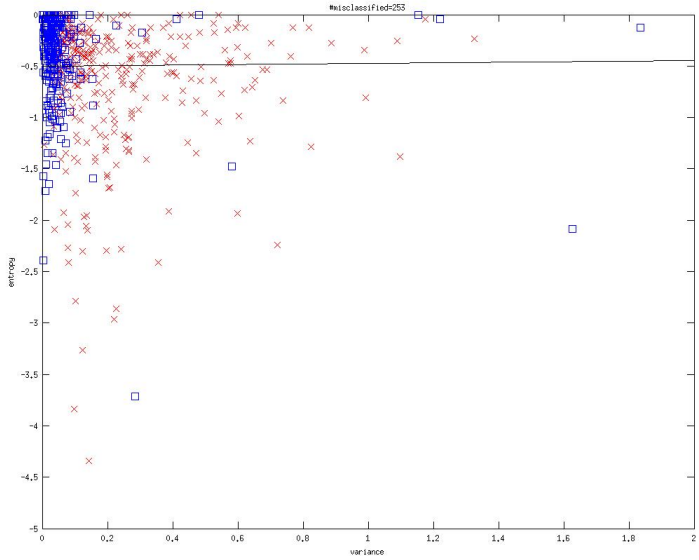


Figure 6.30: Entropy vs. brightness variance: 38.2% misclassification rate.

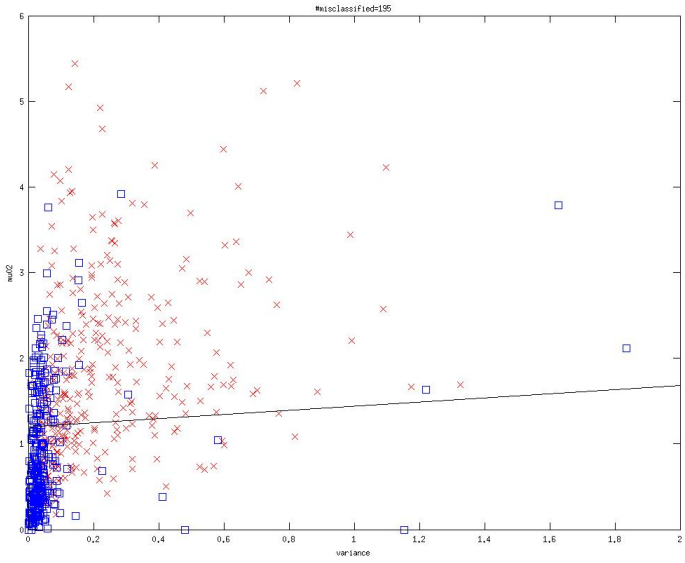


Figure 6.31: μ_{02} vs. brightness variance: 29.5% misclassification rate.

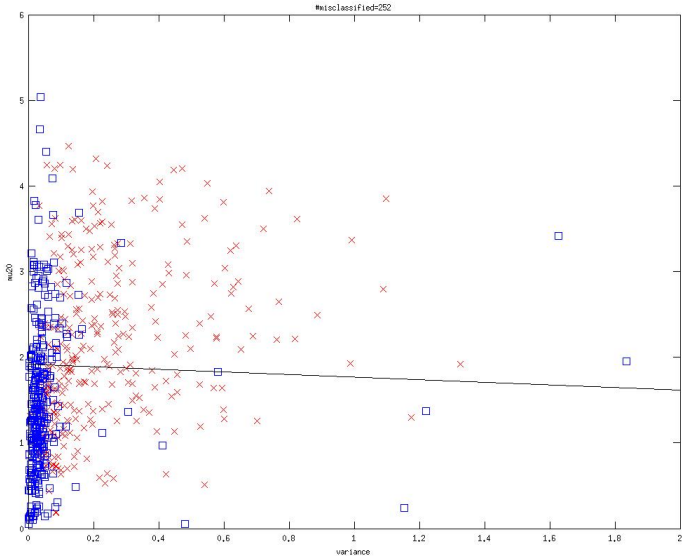


Figure 6.32: μ_{20} vs. brightness variance: 38.1% misclassification rate.

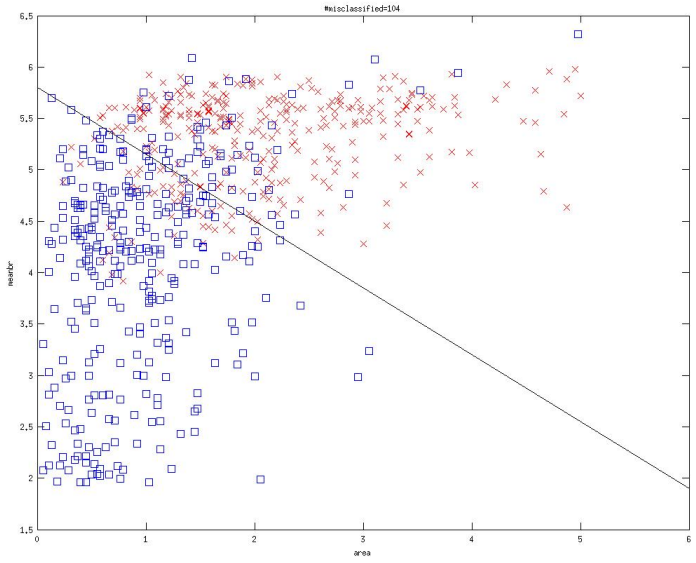


Figure 6.33: Mean brightness vs. area: 15.7% misclassification rate.

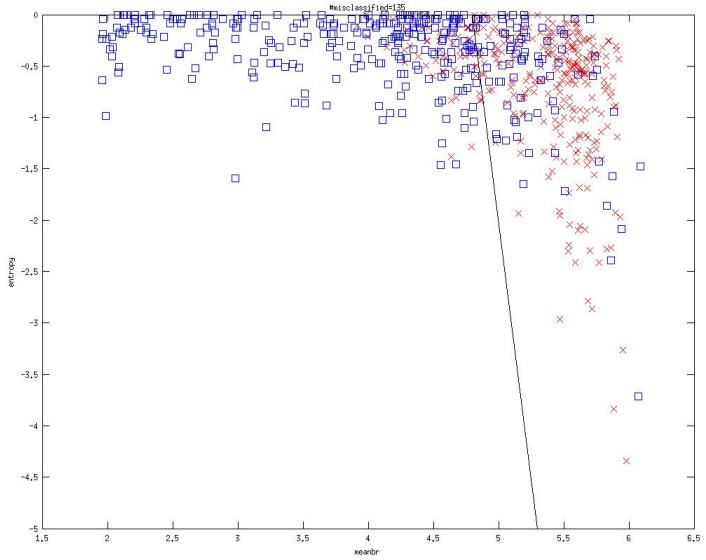


Figure 6.34: Entropy vs. mean brightness: 20.4% misclassification rate.

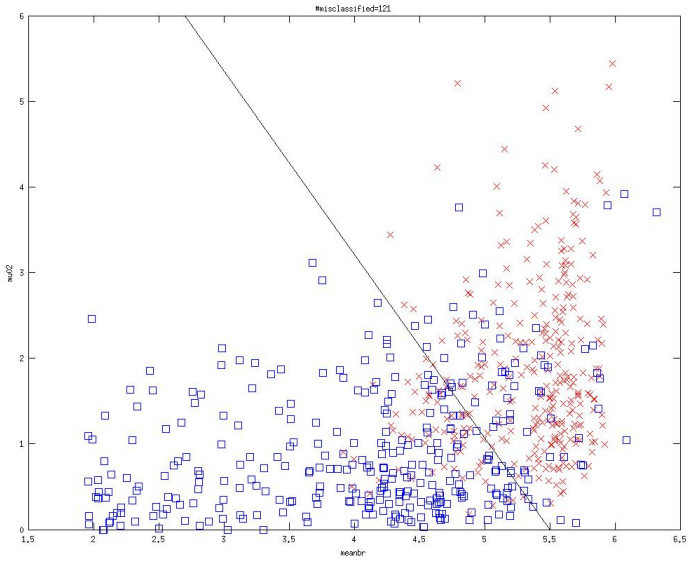


Figure 6.35: μ_{02} vs. mean brightness: 18.3% misclassification rate.

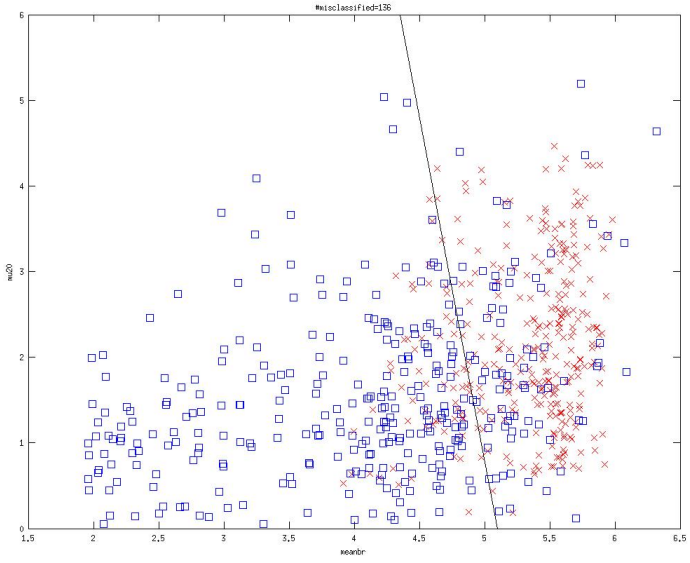


Figure 6.36: μ_{20} vs. mean brightness: 20.5% misclassification rate.

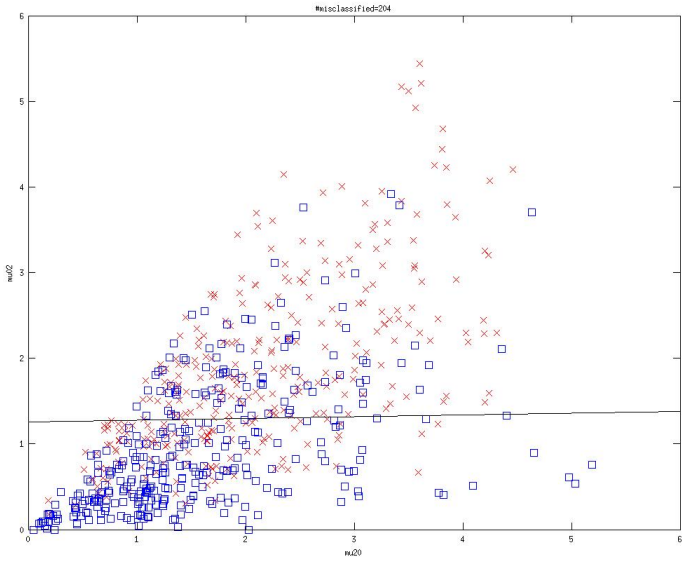


Figure 6.37: μ_{02} vs. μ_{20}: 30.8% misclassification rate.

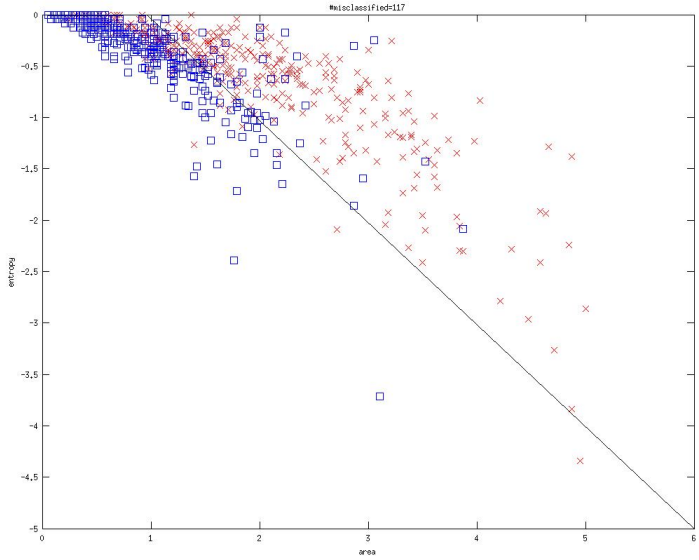


Figure 6.38: Entropy vs. area: 17.7% misclassification rate.

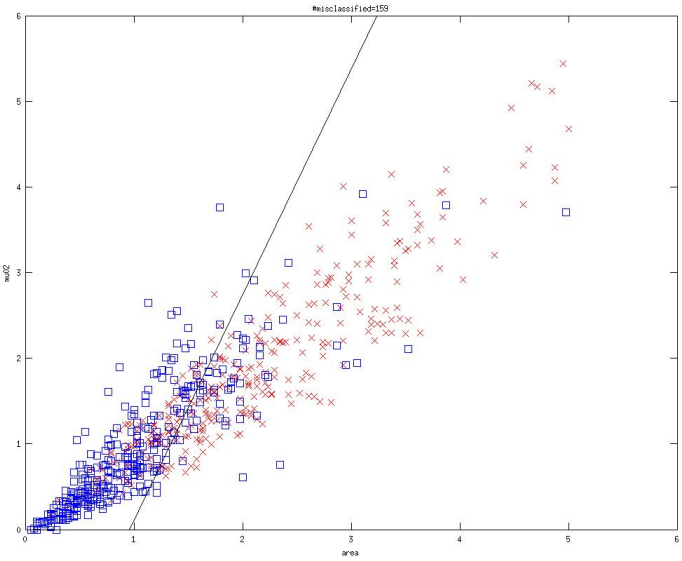


Figure 6.39: μ_{02} vs. area: 24% misclassification rate.

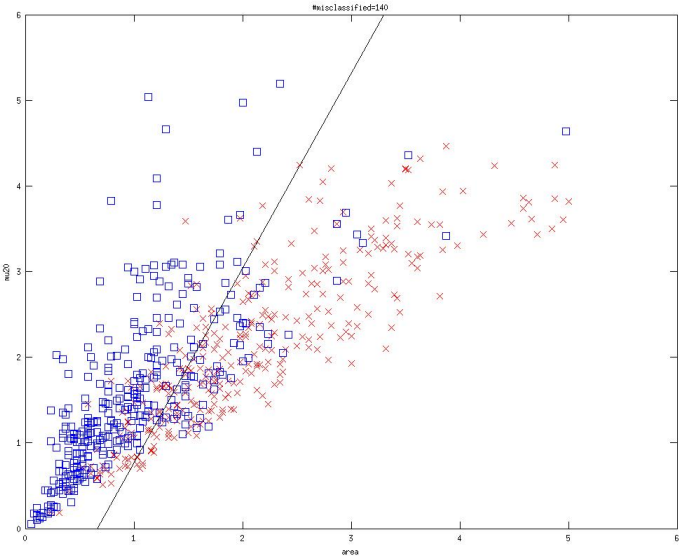


Figure 6.40: μ_{20} vs. area: 21.1% misclassification rate.

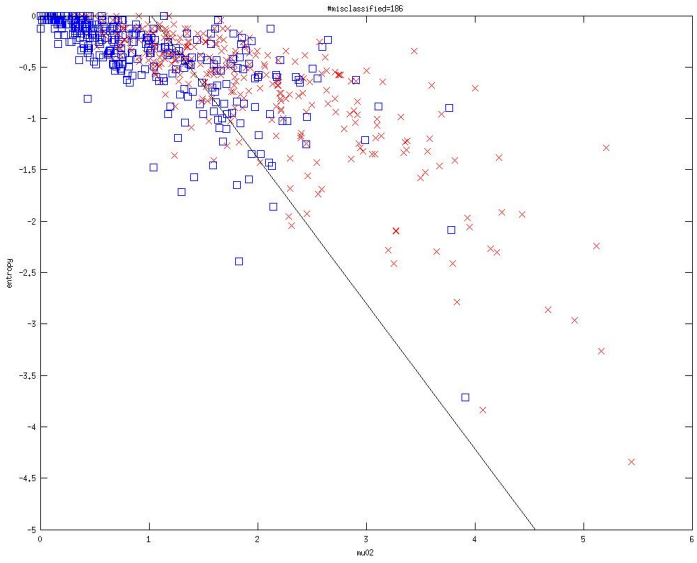


Figure 6.41: Entropy vs. μ_{02}: 28.1% misclassification rate.

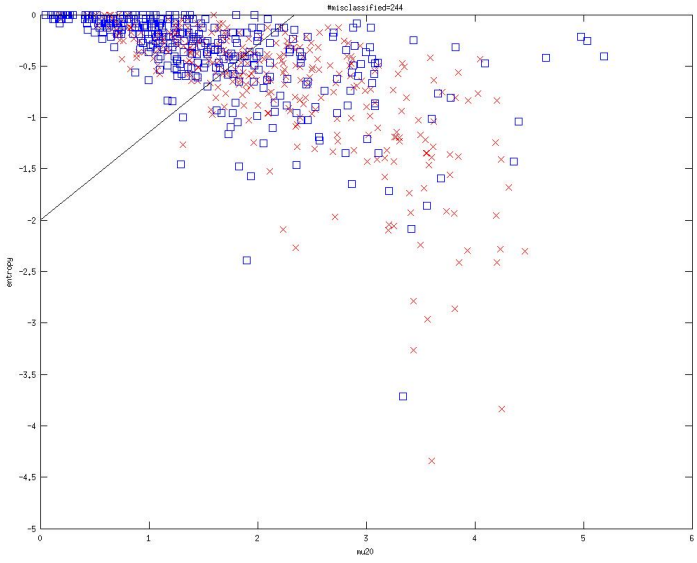


Figure 6.42: Entropy vs. μ_{20}: 36.9% misclassifications.

6.3 Covariance Matrix

The covariance matrices show how features generally vary with one another. The independent features are shown by covariances of zero. Features vary in the same direction (e.g. the second central moments tend to increase as area increases) if the covariances are positive, and they vary in opposing directions if negative. The covariance matrix is computed for both the MCs and non-MC so that the relationships between the features can be compared. The covariance matrices provide insight and understanding of MC detection based on how features affect one another.

The covariance matrix, more specifically the sample covariance matrix, is defined previously in section 3.5.1. Note, the matrix is symmetric and the diagonals of the covariance matrix are equal to the variances of the features, by definition. From here forth, the sample covariance matrix is just referred to as the covariance matrix, since the sample covariance matrix is the estimation of the covariances over all samples.

6.3.1 True Positives

First, the covariance matrix for the true positive samples is computed:

$$\begin{bmatrix} 0.969 & 0.195 & 0.090 & 0.074 & 0.797 & 0.905 & -0.521 \\ 0.195 & 0.242 & -0.102 & 0.099 & 0.137 & 0.150 & -0.002 \\ 0.090 & -0.102 & 0.193 & -0.048 & 0.071 & 0.096 & -0.108 \\ 0.074 & 0.099 & -0.048 & 0.047 & 0.055 & 0.059 & 0.007 \\ 0.797 & 0.137 & 0.071 & 0.055 & 0.910 & 0.609 & -0.416 \\ 0.905 & 0.150 & 0.096 & 0.059 & 0.609 & 1.008 & -0.510 \\ -0.521 & -0.002 & -0.108 & 0.007 & -0.416 & -0.510 & 0.400 \end{bmatrix} \tag{6.5}$$

where the ordering of the feature across columns (left to right), or down the rows (top to bottom), is area, contrast, average brightness, brightness variance, μ_{20}, μ_{02}, and entropy.

The covariance matrix for the MCs reveals a few facts about the MC features:

1. Contrast varies nearly independent of entropy.
2. Most features, except variance, vary inversely with entropy.
3. Contrast and average brightness vary inversely.
4. Average brightness and brightness variance vary inversely.

The covariance of contrast and entropy is relatively small compared to the variances of contrast and entropy. Therefore, the features, for all intents and purposes are independent of one

another. This suggests that these two features are good features for classification.

The covariance matrix shows that most features vary inversely with entropy, except brightness variance. Also, average brightness varies inversely the brightness variance. The rest of the off diagonal covariance elements simply show that the features vary in the same direction. This provides nothing more than a general relationship between those features, however, understanding the fundamental relationship between features is key in determining the best features used to model MCs. However, comparing the covariance matrix of the MCs with the covariance matrix of the non-MCs provides the necessary information in order to better understand MC detection.

6.3.2 True Negatives

The covariance matrix for the true negative samples is computed:

$$\begin{bmatrix} 0.444 & 0.377 & 0.216 & 0.379 & 0.481 & 0.428 & -0.471 \\ 0.377 & 1.084 & 0.447 & 0.523 & 0.458 & 0.344 & -0.471 \\ 0.216 & 0.447 & 1.062 & 0.125 & 0.288 & 0.209 & -0.302 \\ 0.379 & 0.523 & 0.125 & 1.936 & 0351 & 0.249 & -0.496 \\ 0.481 & 0.458 & 0.288 & 0.351 & 0.926 & 0.384 & -0.424 \\ 0.428 & 0.344 & 0.209 & 0.249 & 0.384 & 0.560 & -0.407 \\ -0.471 & -0.471 & -0.302 & -0.496 & -0.424 & -0.407 & 1.561 \end{bmatrix} \tag{6.6}$$

where the ordering of the features is the same as the covariance matrix for the MCs (Equation 6.5).

This covariance matrix reveals a few differences from the matrix for the MCs. The first being that entropy varies inversely with every other feature (no exclusions). All off diagonal elements of the matrix, except those varying with entropy, are positive which indicates that the features vary in the same direction. Also, entropy and contrast are not independent of one another for the non-MCs. Notice, the three elements which have opposite relations (i.e. changes sign) between the MCs and the non-MCs:

1. Entropy and variance.
2. Contrast and average brightness.
3. Average Brightness and brightness variance.

This shows that entropy and variance vary in the same direction for MCs, and then that they are inversely related for the non-MCs. The opposite is true for both contrast vs. average brightness and average brightness vs. brightness variance. For the MCs, these features are inversely

related, and then they vary in the same direction with the non-MCs.

6.4 Performance Using Two Features

In section 6.2, contrast is shown to be a very discriminating feature between MCs and non-MCs. Therefore, we show that it is plausible to use only two features, contrast being one, to detect MCs. This is shown by using preprocessing and segmentation methods discussed in chapter 3, classifying using a perceptron, and computing the performance (chapter 5). The features and their corresponding performances are shown in Tables 6.5 and 6.6. These tables reveal that contrast and variance appear to be the best feature pair of the ones tested, using the perceptron for classification. Although it is possible to use only two features, the performance decreases slightly when reducing the number of features used to classify MCs.

Table 6.5: TPR and FPR using contrast and one other feature.

		Area & Contrast		Contrast & Avg. Brightness		Contrast & Variance	
		Training	Testing	Training	Testing	Training	Testing
TPR	min	79.2%	76.8%	98.9%	97.0%	95.8%	92.8%
	max	99.4%	100%	100%	100%	99.4%	100%
	avg.	96.9%	96.7%	99.6%	99.1%	97.8%	97.2%
FPR	min	5.23%	6.29%	3.01%	2.87%	1.26%	1.92%
	max	10.00%	13.69%	7.14%	7.69%	6.33%	11.54%
	avg.	8.00%	9.01%	4.77%	5.28%	3.72%	4.54%

Table 6.6: TPR and FPR using contrast and another feature.

		Contrast & μ_{20}		Contrast & μ_{02}		Contrast & Entropy	
		Training	Testing	Training	Testing	Training	Testing
TPR	min	96.0%	94.7%	95.7%	92.2%	93.1%	89.3%
	max	100%	100%	100%	100%	99.4%	100%
	avg.	97.8%	97.6%	98.0%	97.6%	97.0%	96.3%
FPR	min	4.82%	6.54%	5.10%	5.59%	4.38%	3.87%
	max	11.18%	15.48%	10.13%	12.43%	13.38%	13.29%
	avg.	8.21%	9.03%	8.20%	8.83%	8.38%	8.39%

Overall, this analysis provides insight into feature based MC detection. The feature means and standard deviations highlight the differences and the spreads between the MCs and non-MCs. The 2-dimensional feature plots help identify the usefulness of feature pairs by determining separability of the classes. The covariance matrices reveal important relations, such as features that are independent and features that have opposite relations depending on whether analyzing MCs or non-MCs. Lastly, it shows that MCs can be detected using only two features. This analysis proves useful to the physical understanding behind the autonomous identification of MCs.

Chapter 7

Conclusion

MC detection is performed with several key components: pre-processor, segmenter, feature vector, and a classifier. The performance of the detection technique highly depends on the features and classifier selected. The performances of the classifiers, which include the Mahalanobis distance, perceptron, MLP, RBF, and SVM, were tested using the h-dome transform, "hill climbing" algorithm, and the descriptive features, including: area, contrast, average brightness, brightness variance, second central moments, and entropy.

First, the h-dome transform finds regional maxima, which correspond to MCs as well as peaks due to noise. Since the transform itself is not accurate enough to detect MCs, the potential MCs are segmented using the "hill climbing" algorithm and the feature vectors are extracted. Then, feature vectors are then classified using one of the methods described. The classifier further eliminates points not corresponding to MCs and provides the final list of MC locations in the image.

The experiments tested on both the DDSM and WFU databases show that the RBF network with 150 cluster centers provides the best classification performance, with 99.9% TPR and 0.8 FPR. However, the perceptron shows the best ability to generalize, with 98.7% TPR and 2.21% FPR. These classifiers show the best performances overall, however, optimal performance may be achieved using a different feature vector, network architecture (i.e. number of neurons in the hidden layer), classifier, or number of training samples. The performances using different combinations of these elements may provide the best classification and detection of MCs. More research is needed in order to explore the many variations of MC detection methods.

The feature study reveals the most descriptive and useful features (of those studied in this work) for classifying MCs by analyzing individual features, feature pairs, and the covariance matrices for both MCs and non-MCs. The individual feature analysis shows that most features have an area of overlapping MC samples and non-MC samples. The pairwise analysis shows that the most discriminative features or feature pairs include: contrast, average brightness vs.

brightness variance, and brightness variance. Then, the covariance matrices reveal the general dependencies between the features for both MCs and non-MCs. The covariance analysis shows that entropy and contrast are independent for MCs, which implies that contrast and entropy are useful features for modeling MCs. Lastly, the feature study shows that it is possible to use only two features to classify MCs.

7.1 Future Work

The golden standard of performance has yet to be determined. A key in defining this type of ideal standard is understanding what MCs are in terms of their defining features. The defining features may still need to be determined. More research is needed to be able to truly understand MCs and how to detect them. More experiments need to be performed with a larger training and testing sets in order to show better resolution of both features and classifiers. By testing with a larger data set, the classifiers should more conclusively show which classifier performs the best for MC detection.

A more extensive feature study should be performed in order to fully understand the defining features of MCs. This work only studied a small number of features; there are infinitely many measures that could possibly help describe MCs. However, there are only a small number that will best describe MCs.

REFERENCES

[1] I. N. Bankman, T. Nizialek, I. Simon, and O. B. Gatewood. Segmentation Algorithms for Detecting MIcrocalcifications. *IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE*, 1(2):141–149, 1977.

[2] A. Bazzani, A. Bevilacqua, D. Bollini, R. Brancaccio, R. Campanini, N. Lanconelli, A. Riccardi, D. Romani, and G. Zamboni. Automatic detection of clustered microcalcifications in digital mammograms using an SVM classifier. *Proceeding of the 8th Eur Symp on Artificial Neural Networks*, pages 195–200, 2000.

[3] U. Bick and F. Diekmann, editors. *Digital Mammography*. Springer, 2010.

[4] H. D. Cheng, X. Cai, X. Chen, and L. Hu. Computer-aided detection and classification of microcalcifications in mammograms: A survey. *Pattern Recognition*, 36(12):1467–1477, 2003.

[5] T. M. Cover. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Transactions on Electronic Computers*, EC-14(3):326–334, 1965.

[6] D. H. Davies and D. R. Dance. Automatic computer detection of clustered calcifications in digital mammograms. *Phys. Med. Biol.*, 35(8):1111–1118, 1990.

[7] J. Dengler, S. Behrens, and J. F. Desaga. Segmentation of Microcalcifications in Mammograms. *IEEE Trans Med Image*, 12(4):634–642, 1993.

[8] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley & Sons, 2nd edition, 2001.

[9] I. El-Naqa, Y. Yang, M. N. Wernick, N. P. Galatsanos, and R. Nishikawa. Support Vector Machine Learning for the Detection of Microcalcifications in Mammograms. *IEEE Trans Med Imag*, pages 201–204, 2002.

[10] S. Halkiotis, T. Botsis, and M. Rangoussi. Automatic detection of clustered microcalcifications in digital mammograms using mathematical morphology and neural networks. *Signal Processing*, 87(7):1559–1568, 2007.

[11] S. Haykin. *Neural Networks and Learning Machines*. Pearson Education, 3rd edition, 2009.

[12] M. Heath, K. Bowyer, D. Kopans, R. Moore, and W. P. Kegelmeyer. The Digital Database for Screening Mammography. In M. J Yaffe, editor, *Proceedings of the Fifth International Workshop on Digital Mammography*, pages 212–218. Medical Physics Publishing, 2001.

[13] M. Heath, K. Bowyer, D. Kopans, R. Moore, W. P. Kegelmeyer, K. Chang, and S. Munishkumaran. Current status of the Digital Database for Screening Mammography. In *Proceedings of the Fourth International Workshop on Digital Mammography*, pages 457–560. Kluwer Academic Publishers, 1998.

[14] W. R. Hensyl, editor. *Stedman's Medical Dictionary*. Williams & Wilkins, 25th edition, 1990.

[15] M. Hu. Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*, 8(2):179–187, 1962.

[16] M. Kallergi, K. Woods, L. P. Clarke, and W. Qian. Image segmentation in digital mammography: Comparison of local thresholding and region growing algorithms. *Computerized Medical Imaging and Graphics*, 16(5):323–331, 1992.

[17] J. K. Kim and H. W. Park. Statistical Textural Features for Detection of Microcalcifications in Digitized Mammograms. *IEEE TRANSACTIONS ON MEDICAL IMAGING*, 18(3):231–238, 1999.

[18] A. F. Laine, S. Schuler, J. Fan, and W. Huda. Mammographic Feature Enhancement by Multiscale Analysis. *IEEE TRANSACTIONS ON MEDICAL IMAGING*, 13(4):725–740, 1994.

[19] T. K. Lau and W. F. Bischof. Automated Detection of Breast Tumors Using the Asymmetry Approach. *Computers and Biomedical Research*, 24:273–295, 1991.

[20] J. M. Mossi and A. Albiol. Improving detection of clustered microcalcifications using morphological connected operators. *IEEE Image Processing and Its Applications*, pages 498–501, 1999.

[21] R. Nakayama, Y. Uchiyama, K. Yamamoto, and R. Watanabe. Detection of Clustered Microcalcifications on Mammograms Using New Filter Bank. *Systems and Computers in Japan*, 36(2):141–150, 2005.

[22] D. I. Papadopoulos, A. Fotiadis and A. Likas. Characterization of clustered microcalcifications in digitized mammograms using neural networks and support vector machines. *Artificial Intelligence in Medicine*, 34(4):725–740, 1994.

[23] L. Prasad and S. S. Iyengar. *Wavelet Analysis with Applications to Image Processing*. CRC Press, 1997.

[24] G. A. Rodriguez, J. A. Gonzalez, L. Altamirano, J. S. Guichard, and R. Diaz. Microcalcification Detection Using Fisher's Linear Discriminant and Breast Density. *Advances in Experimental Medicine and Biology*, 696(6), 2011.

[25] P. K. Sahoo, S. Soltani, and A. K. Wong. A Survey of Thresholding Techniques. *Computer Vision Graphics Image Processing*, 41:233–260, 1988.

[26] J. Serra. *Image Analysis and Mathematical Morphology*, volume 1. Academic Press, 1982.

[27] J. Serra. *Image Analysis and Mathematical Morphology*, volume 2. Academic Press, 1988.

[28] Jean Serra. Introduction to mathematical morphology. *Computer Vision, Graphics, and Image Processing*, 35:283–305, 1986.

[29] W. E. Snyder and H. Qi. *Machine Vision*. Cambridge University Press, 2004.

[30] P. Soille. *Morphological image analysis: Principles and applications*. Springer, 2nd edition, 2003.

[31] V. Srivastava. Performance of Microcalcification Detection Algorithms. Master's thesis, North Carolina State University, 2005.

[32] R. N. Strickland and H. I. Hahn. Wavelet Transforms for Detecting Microcalcifications in Mammograms. *IEEE TRANSACTIONS ON MEDICAL IMAGING*, 15(2):218–229, 1996.

[33] T. Suzuki, M. Toi, S. Saji, K. Horiguchi, T. Aruga, E. Suzuki, S. Horiguchi, N. Funata, K. Karasawa, and N. Kamata. Early Breast Cancer. *Int. J. Clin. Oncol*, 11:108–119, 2006.

[34] L. Vincent. Morphological Grayscale Reconstruction in Image Analysis: Applications and Efficient Algorithms. *IEEE TRANSACTIONS ON IMAGE PROCESSING*, 2(2):176–201, 1993.

[35] A. Wroblewska, P. Boninski, A. Przelaskowski, and M. Kazubek. Segmentation and feature extraction for reliable classification of microcalcifications in digital mammograms. *OPTO-ELECTRONIC REVIEW*, 1(3):227–235, 2003.

[36] D. Zhao, M. Shridhar, and D. G. Daut. Morphology on detection of calcifications in mammograms. *IEEE Acoustics, Speech, and Signal Processing*, 3(3):129–132, 1992.