A Computer Aided Diagnosis System for Breast Cancer Using Independent Component Analysis and Fuzzy Classifier

Fadi Abu-Amara
A COMPUTER AIDED DIAGNOSIS SYSTEM FOR BREAST CANCER USING INDEPENDENT COMPONENT ANALYSIS AND FUZZY CLASSIFIER

by

Fadi Abu-Amara

A Thesis
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Master of Science in Engineering (Computer)
Department of Electrical and Computer Engineering

Western Michigan University
Kalamazoo, Michigan
April 2007
ACKNOWLEDGMENTS

I would like to express my appreciation to my advisor Dr. Ikhlas Abdel-Qader for her supervision, support, and critique of this work. I have truly gained invaluable knowledge and enjoyed working with her.

Secondly, I would like to thank my committee members, Drs. Janos Grantner and Ala Al-Fuqaha, for their precious suggestions that enriched my work. I would like to express my appreciation to Western Michigan University for its generous support to the Information Technology and Image Analysis Center. Partial support of this work was provided by the National Science Foundation (MRI-0215356), FRACASF Award (WMU:2005-2007). Also, I would like to express my appreciation to Al Hussein Bin Talal University for funding my study.

Also, I would like to thank my fiancé for her love and patience during the last semester of my work. Finally, I would like to express my deepest gratitude to my parents for their immense love, dedication, and encouragement throughout my life and this work.

Fadi Abu-Amara
A COMPUTER AIDED DIAGNOSIS SYSTEM FOR BREAST CANCER USING INDEPENDENT COMPONENT ANALYSIS AND FUZZY CLASSIFIER

Fadi Abu-Amara, M.S.E.
Western Michigan University, 2007

Screening mammograms by a radiologist is a repetitive task that causes fatigue and eye strain. For every thousand cases analyzed by a radiologist, only 3 to 4 are cancerous and thus an abnormality may be overlooked. Computer-Aided Detection (CAD) algorithms have been developed to assist radiologists in detecting mammographic lesions. CAD algorithms have improved total radiologists' detection accuracy of cancerous tissues. In this thesis, a computer aided detection and diagnosis (CADD) system for breast cancer is developed. The algorithm framework is based on combining principal component analysis (PCA), independent component analysis (ICA), and fuzzy classifier to identify and label suspicious regions from digitized mammograms. This proposed algorithm is novel in utilizing a fuzzy classifier integrated with the ICA model. This system is implemented and tested by using images from the MIAS database and results in labeling the tested image as either normal or abnormal. If abnormal, CADD differentiates it into a benign or a malignant tissue. Experimental results show that the proposed algorithm has 84.03% accuracy in detecting all kinds of abnormalities and 78% diagnosis accuracy.
# TABLE OF CONTENTS

ACKNOWLEDGMENTS ........................................................................................................ ii

LIST OF TABLES .............................................................................................................. vii

LIST OF FIGURES............................................................................................................. viii

CHAPTER

1. INTRODUCTION........................................................................................................ 1

1.1 Detection of Masses ..................................................................................... 5

1.2 Detection of Microcalcifications ............................................................. 5

2. LITERATURE REVIEW..................................................................................... 9

2.1 MIAS Database ................................................................................. 9

2.2 Breast Border Detection ................................................................. 10

2.3 Active Contours-Snake Algorithm ................................................... 11

2.4 Spatial Filtering Techniques ............................................................... 12

2.4.1 Directional Filtering .................................................................. 12

2.4.2 Non-Linear Filters ..................................................................... 13

2.4.3 Iris Filter ................................................................................. 16

2.5 Enhancement Techniques ........................................................................ 16

2.5.1 Preprocessing Methods ............................................................... 17

2.5.2 Conventional Techniques ............................................................ 17

2.5.3 Adaptive Enhancement ............................................................... 20

2.5.4 Density-Weighted Contrast Enhancement ...................................... 21

2.5.5 Modal-Based Image Enhancement ................................................ 21
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.6</td>
<td>Noise Equalization</td>
<td>22</td>
</tr>
<tr>
<td>2.7</td>
<td>Fractal Modeling</td>
<td>23</td>
</tr>
<tr>
<td>2.8</td>
<td>Image Segmentation</td>
<td>24</td>
</tr>
<tr>
<td>2.8.1</td>
<td>Segmentation and Thresholding</td>
<td>25</td>
</tr>
<tr>
<td>2.8.2</td>
<td>Markov Random Field</td>
<td>26</td>
</tr>
<tr>
<td>2.8.3</td>
<td>Deformable Models</td>
<td>26</td>
</tr>
<tr>
<td>2.8.4</td>
<td>Pectoral Muscle Segmentation</td>
<td>27</td>
</tr>
<tr>
<td>2.9</td>
<td>Wavelet Transform</td>
<td>27</td>
</tr>
<tr>
<td>2.9.1</td>
<td>Wavelet Expansion</td>
<td>29</td>
</tr>
<tr>
<td>2.9.2</td>
<td>Wavelet Sub-Band Decomposition</td>
<td>30</td>
</tr>
<tr>
<td>2.10</td>
<td>Fuzzy Logic Techniques</td>
<td>31</td>
</tr>
<tr>
<td>2.10.1</td>
<td>Normalization</td>
<td>31</td>
</tr>
<tr>
<td>2.10.2</td>
<td>Fuzzy-Genetic</td>
<td>32</td>
</tr>
<tr>
<td>2.11</td>
<td>Statistical Analysis</td>
<td>32</td>
</tr>
<tr>
<td>2.12</td>
<td>Classifiers</td>
<td>33</td>
</tr>
<tr>
<td>2.12.1</td>
<td>Artificial Neural Networks</td>
<td>33</td>
</tr>
<tr>
<td>2.12.2</td>
<td>Pattern Recognition Algorithms</td>
<td>35</td>
</tr>
<tr>
<td>2.12.3</td>
<td>Hybrid Neural Networks</td>
<td>35</td>
</tr>
<tr>
<td>2.12.4</td>
<td>Modular Neural Networks</td>
<td>36</td>
</tr>
<tr>
<td>2.12.5</td>
<td>Neuro-Fuzzy</td>
<td>36</td>
</tr>
<tr>
<td>2.12.6</td>
<td>Fuzzy K-Nearest Classifier</td>
<td>37</td>
</tr>
</tbody>
</table>
## Table of Contents—Continued

### CHAPTER

2.12.7 Convolution Neural Networks .................................................. 37
2.12.8 Self-Organizing Map .......................................................... 38
2.12.9 Enhanced Rough Set Approach .............................................. 38
2.12.10 Fuzzy Classifier with ICA ................................................. 39

2.13 Linear Transformation Methods .................................................. 39
   2.13.1 Principal Component Analysis-PCA .................................. 39
   2.13.2 Independent Component Analysis ..................................... 40
   2.13.3 Feature Extraction Methods ............................................ 41

2.14 Discussion of Literature .......................................................... 42
   2.14.1 ROC and FROC Curve Analysis ...................................... 42
   2.14.2 Sensitivity versus Specificity ........................................ 42

3. LINEAR TRANSFORMATION METHODS .......................................... 44
   3.1 Principal Component Analysis ............................................. 44
   3.2 Independent Component Analysis ........................................ 46

4. BACKGROUND ON FUZZY CLASSIFIERS ......................................... 50
   4.1 Approximate Fuzzy Rule Base ............................................ 51
   4.2 Descriptive Fuzzy Rule Base .............................................. 53
   4.3 Fuzzy Classifier Structure ................................................ 53
   4.4 Fuzzy Model Initialization ................................................ 54

5. FRAMEWORK OF PROPOSED CADD ALGORITHM .................................. 56
   5.1 Sub-Images Generation ...................................................... 57
Table of Contents—Continued

CHAPTER

5.2 Training Procedure................................................................. 57
5.3 Unsupervised Learning Algorithm............................................. 58
5.4 Testing Procedure.................................................................... 60
5.5 Fuzzy Classifier Modeling......................................................... 61

6. EXPERIMENTAL RESULTS.......................................................... 63
6.1 Parameters Selection .............................................................. 63
  6.1.1 Number of Selected Principal Components........................ 63
  6.1.2 Learning Rate................................................................. 64
  6.1.3 Moment Algorithm.......................................................... 67
  6.1.4 Mapping Range.............................................................. 67
  6.1.5 Cropping and Scaling Criterion ........................................ 70
6.2 Discussion of Experimental Results......................................... 70

7. CONCLUSIONS AND FUTURE DIRECTIONS................................. 74

APPENDIX
  Simulation Results for the Proposed CADD Algorithm Set #5........ 75

BIBLIOGRAPHY............................................................................... 80
# LIST OF TABLES

1. A Comparison between Fuzzy Set and Crisp Set Theory ........................................... 50
2. Different Sets Used To Evaluate the Detection Algorithm Performance ...... 57
3. Number of Selected Principal Components Impact on Algorithm Accuracy Where Learning Rate and Mapping Range of Each Set are Kept Fixed .................................................................................................................... 64
4. FP and FN and Total PCA, CADD Algorithms Accuracy ........................................... 71
5. FP and FN and Total ICA, CADD Algorithms Accuracy ........................................... 71
6. Computer Aided-Diagnosis Using CADD Algorithm ................................................ 73
LIST OF FIGURES

1. (a) Midolateral Oblique View and (b) Craniocaudal View......................... 1
2. Abnormal Regions Where (a) and (b) an Architectural Distortion, (c) and (d) Microcalcifications and Speculated Mass ............................................... 3
3. Main Steps Involved in Most CADD Systems ............................................. 4
4. Morphology of Calcifications ..................................................................... 7
5. (a) and (b) Microcalcifications With Higher Probability of Malignancy....... 8
6. (a), (b), and (c) Typically Benign Microcalcifications.............................. 8
7. A Triangular Membership Function of a Particular Fuzzy Set...................... 52
8. Fuzzy Space of a Particular Object of Figure 7 ........................................ 52
9. Trapezoidal Fuzzy Set Example ............................................................... 54
10. (a) General Block Diagram of the Proposed Algorithm and (b) Detailed Block Diagram of Proposed Algorithm ......................................................... 56
11. Learning Rate Impact on Algorithm Accuracy for Test Set #1 Where Other Parameters Kept Constant................................................................. 65
12. Learning Rate Impact on Algorithm Accuracy for Test Set #2 Where Other Parameters Kept Constant................................................................. 65
13. Learning Rate Impact on Algorithm Accuracy for Test Set #3 Where Other Parameters Kept Constant................................................................. 66
14. Learning Rate Impact on Algorithm Accuracy for Test Set #4 Where Other Parameters Kept Constant................................................................. 66
15. Learning Rate Impact on Algorithm Accuracy for Test Set #5 Where Other Parameters Kept Constant................................................................. 67
16. Mapping Range Impact on Algorithm Accuracy for Test Set #1.............. 68
17. Mapping Range Impact on Algorithm Accuracy for Test Set #2.............. 68
List of Figures—Continued

18. Mapping Range Impact on Algorithm Accuracy for Test Set #3................... 69
19. Mapping Range Impact on Algorithm Accuracy for Test Set #4................... 69
20. Mapping Range Impact on Algorithm Accuracy for Test Set #5................... 70
1. INTRODUCTION

Breast cancer is considered one of the most common and fatal cancers among women in the USA [1]. The American Cancer Society estimated that about 40,970 women died in 2006 due to breast cancer, and on average, every three minutes one woman is diagnosed with breast cancer. Right now there are over two million women in the US have been treated for this disease [1]. Therefore, an early detection of this disease will improve the treatment options.

Screening mammography, low-dose X-rays 2D imaging of the human breast, is the most successful tool for early detection of non-palpable and potentially curable breast cancer that was first reported in 1913. Screening mammograms can be done in two ways, the craniocaudal view (CC) and the mediolateral oblique view (MLO). In CC view, the breast image is taken from top to bottom, while MLO view is a side view taken from an angle. Both methods are shown in Figure 1.

Figure 1: (a) Mediolateral Oblique View and (b) Craniocaudal View.
Radiologists visually examine mammograms to search for signs of abnormal regions. They usually look for clusters of microcalcifications, architectural distortions, or masses. A mass is described by its marginal characteristics and shape. It can be defined as a region-occupying lesion seen in at least two different projections [2].

Calcifications are small deposits of calcium within the breast that appear as white spots. There are two types of calcifications. First, macrocalcifications are large calcium deposits appear as big white dots or dashes. Second, microcalcifications are small calcium deposits appear as fine white specks. Of the two, microcalcifications are of most concern. The shape and distribution of microcalcifications assist the radiologist in determining the likelihood of breast cancer. An architectural distortion is not a mass but a desmoplastic reaction in which there is focal disruption of the normal breast tissue pattern. It appears as a distortion in which surrounding breast tissues seem to be pulled inward into a focal point. Typical example of each type is shown in Figure 2.

Early detection of breast cancer via mammography can improve treatment chances and survival rates [3]. Unfortunately, mammography is not perfect. False positive (FP) rates are 15-30% due to the overlap in the appearance of malignant and benign abnormalities. False negative (FN) rates are 10-30%. An FP is when a radiologist reports a suspicious change in the breast but no cancer is found after further examinations. FP leads to unnecessary biopsies and anxiety. An FN means failure to detect or correctly characterize breast cancer in a case of which later tests
conclude that cancer is present. Nonetheless, mammography has an overall accuracy rate of 90% [4].

Figure 2: Abnormal Regions Where (a) and (b) an Architectural Distortion, (c) and (d) Microcalcifications and Speculated Mass.

Source: MIAS Database. (a) Mdb155, (b) Mdb124, (c) Mdb219, and (d) Mdb206.

Screening mammograms by a radiologist is a repetitive task that causes fatigue and eye strain. For every thousand cases analyzed by a radiologist, only 3 to 4 are cancerous and thus an abnormality may be overlooked. Computer-Aided Detection and Diagnosis (CADD) algorithms have been developed to assist radiologists in
detecting mammographic lesions. These systems are regarded as a second reader, and the final decision is left to the radiologist. CADD algorithms have improved total radiologist accuracy of detection of cancerous tissues [5]. CADD algorithms are considered as an extremely challenging task for various reasons. First, the imaging system may have serious imperfections. Second, the image analysis task is compounded by the large variability in the appearance of abnormal regions. Finally, abnormal regions are often hidden in dense breast tissue. The goal of the detection stage is to assist radiologists in locating abnormal tissue. A flowchart showing different steps involved in most detection algorithm is shown in Figure 3.

![Flowchart of CADD Systems]

Figure 3: Main Steps Involved in Most CADD Systems.
1.1 Detection of Masses

There are many mass detection algorithms reported in the literature. Masses with speculated margins have a high likelihood of malignancy. In mammograms, speculated masses that have a stellate appearance are the most difficult type of tumors to detect for several reasons. First, their central masses are irregular with ill-defined borders. Second, their diameter size varies from a few millimeters to many centimeters. A speculated mass is characterized by lines radiating from its margins [6]. However, not all malignant masses are speculated. Therefore, any mass detection algorithm should also detect non-speculated masses. Most mass detection algorithms consist of two stages: suspicious regions detection (feature extraction) and suspicious regions classification as abnormal mass or normal tissue. Algorithms of first stage can generally be pixel- or region-based [7].

1.2 Detection of Microcalcifications

Microcalcifications are small calcium deposits that appear as fine white specks. They can either be benign or malignant. In mammography, they are characterized by their size, morphology, shape, and number. They are usually clustered within one area of the breast. Their size is very small, less than 1 mm, and the average diameter is 0.3 mm [6]. Malignant microcalcifications vary extremely in number, shape, size, and density. They are usually clustered within one area of the breast. Figure 4 shows calcifications morphology. Two basic types of malignant microcalcifications are shown in Figure 5. Benign microcalcifications are
characterized by sharp outline, homogeneous shape, and uniform density. Figure 6 shows benign microcalcifications.

Many factors affect microcalcification detection algorithms. First, the detection algorithm may miss small microcalcifications due to the overlapping breast regions. Second, it is difficult to localize the microcalcifications in the dense regions. Finally, since microcalcifications may have a low contrast in comparison to the breast background, they can be regarded as a noise [8]. Fortunately, the high degree of their localization makes their modeling easier by regarding them as impulse-like noise. A number of detection methods have been developed utilizing this model [9], [15], and [30].

The remainder of this thesis is organized as follows: Chapter 2 presents a literature review of the most recent detection algorithms. In Chapter 3, linear transformation methods are explored while in Chapter 4, fuzzy logic as a classifier is presented. In Chapter 5, the proposed CADD algorithm is developed and in Chapter 6, the experimental results are discussed while conclusions and recommendations for future work are presented in Chapter 7.
Figure 4: Morphology of Calcifications.

Source: Cancer Medicine 6.
Figure 5: (a) and (b) Microcalcifications With Higher Probability of Malignancy. Source: Cancer Medicine 6.

Figure 6: (a), (b), and (c) Typically Benign Microcalcifications. Source: Cancer Medicine 6.
2. LITERATURE REVIEW

Most microcalcification detection methods are summarized in this chapter, and the accuracy of any of these methods depends on the spatial resolution and gray levels range of the digitized mammographic images. Some microcalcifications can be smaller than 0.1mm [9] which means that these small microcalcifications occupy more than one pixel in the image. Therefore, it is easier for detection algorithms to detect and distinguish them from noise.

2.1 MIAS Database

Mammograms used in this thesis are extracted from MIAS Database; the Mammography Image Analysis Society (MIAS) is a UK organization interested in the mammograms analysis that has produced a free digital mammography database [10]. All the images have been reduced to 200 micron pixels and clipped or padded so that every image is 1024 pixels by 1024 pixels. The database contains left and right breast images for 161 patients. It consists of 322 images, where 208 images are normal, 63 images are benign, and 51 images are malignant. It also contains 119 ROS that cover all types of abnormalities. The smallest abnormality extends over 3 pixels in radius, while the largest one covers 197 pixels in radius.
2.2 Breast Border Detection

The algorithms that fall under this category are used as pre-processing steps for any automatic detection system of breast cancer.

Mendez et al. [11] developed an algorithm for breast border and nipple detection as a preprocessing step for further image analysis. An algorithm that computes the gray levels gradient is used to detect the breast border. First, the image is smoothed by using 11x11 pixels filter mask. Next, the breast area is divided into three regions (I, II and III). This is done by automatically selecting five reference points, (x₁,y₁), (x₂,y₂), (x₃,y₃), (x₄,y₄), and (x₅,y₅). Finally, a tracking algorithm is used to detect the breast border which assumes that point (x,y) belongs to the breast border if the gray level value of its nine previous pixels satisfies the condition:

\[ f(x_1,y_1) < f(x_2,y_2) < \ldots < f(x_7,y_7) \leq f(x_8,y_8) \leq f(x_9,y_9) \leq (x,y) \]  

(1)

In region I, the breast border is searched by the algorithm from left to right. In region II, the breast border is searched from top to bottom. And in region III, the breast border is searched from right to left.

In order to detect the nipple, three algorithms are used and compared to each other. They are maximum height of the breast border, maximum gradient, and maximum second derivative of the gray levels across the median-top section of the breast. Also, a combined method is developed. The algorithms' performance is tested on 156 mammograms. In most of the tested mammograms, the estimated border is very close to the radiologist's estimated border.
2.3 Active Contours - Snake Algorithm

A snake is an active contour model that looks for lines and close edges and localizes them. To expand the detected regions, scale-space continuum algorithm is used.

Kass et al. [13] proposed a snake’s method that uses local minimum to look for contours. The contour is placed near desired image features, and then the snake takes a minimum energy measure for all possible points in a neighborhood surrounding each point to slither to the desired points.

Wirth and Stapinski [12] used active contours to detect the breast contour in mammograms. To make the snake technique feasible for segmentation, several mammograms' properties are used. First, the breast-air border itself is a very low gradient and may be obscured by noise. Second, the uncompressed fat close to the breast-air border is a gradient, growing as the fat closes to the center of the breast. From first property, a noise removal step is included in the algorithm to allow the snake to differentiate between the breast contour and the noise. In order to deal with lost contour details during the noise removal process, snakes are used to fill in gaps that occur in contours. From second property, right-to-left edge detection methods classify the gradient of the breast as an edge, especially as the breast approaching towards left. In contrast, left-to-right edge detection methods pick up noise. Then, a dual threshold creates a difference in terms of the detected area. This difference is used to approximate the breast contour location.
Chang et al. [14] developed a three-dimensional snake algorithm to find the tumor's contour. The estimated margin helps the physician in estimating the effect of the surgery. First, the noise is reduced by using an anisotropic diffusion filter. Then, the edges are enhanced by using the stick detection technique. Finally, the tumor's contour is estimated by using the gradient vector flow snake.

2.4 Spatial Filtering Techniques

Gulsrud and Husoy [15] used a Fisher criterion method to optimize a filter for texture feature extraction. The extracted features are used to differentiate between texture representing abnormal regions and texture representing normal regions. The feature separation step is achieved through computing the feature mean and variance. To get a signal-enhanced image and a signal-suppressed image, two types of spatial filters are used. By subtracting the suppressed image from the enhanced image, the abnormal tissues are enhanced in the resulting difference image while the normal tissues are reduced.

2.4.1 Directional Filtering

Gabor filters have been used by many researchers in the literature. Most of these methods are associated with segmentation and analysis of texture.

Ferrari et al. [16] proposed a procedure for the analysis of left–right asymmetry in mammographic images. Gabor wavelets are used to detect linear directional components. Two-dimensional Gabor wavelets are applied to the entire image. In order to minimize the redundancy in the wavelet-based representation, the
tuning frequency and the orientation are left variable. A set of Gabor filters is normalized to have zero dc responses and designed to have low redundancy in the representation. The filters are projected to ensure that the half-peak magnitude of the filter responses touch one another. The purpose of the projection step is to ensure that the filters capture most of the information with minimum redundancy. Karhunen–Loeve transform and Otsu’s thresholding method are used to analyze the filter responses for different orientations and scales. To maintain the most significant directional elements, the V principal components of the filter responses are selected. Otsu’s method is then used to threshold the selected principal components. Then, the magnitude and phase of the directional components are estimated by using these selected principal components.

2.4.2 Non-Linear Filters

Lai et al. [24] proposed a technique for detecting circumscribed masses using modified median filtering to enhance the processed image followed by a template matching step to detect abnormal regions. To enhance the images, four selective averaging filters along with a modified linear filter are used. The edge-preserving smoothing method looks for a homogeneous neighborhood in all directions for the current pixel and averages this neighborhood. In contrast, half-neighborhood technique is an un-weighted averaging that operates at pixels within the current region. In case of an edge, the neighborhood is subdivided in different ways, and the half neighborhood with the most different average level from that of the other half is selected for averaging. The proposed k-nearest neighbor method is based on the idea
that pixels within the same region should have similar gray level intensities. Then, a directional smoothing process is used to average pixels in the interior of the processed region using all neighborhood pixels. In case of an edge, the neighbors that lie in the direction along the edge are taken by a directional average. To decrease the computational complexity, a coarse-fine template matching process is implemented. Suspicious areas in template matching step are picked by thresholding the cross-correlated values where a percentile technique is used to determine an appropriate threshold for each image.

Qian et al. [25] developed a new class of nonlinear filters for noise reduction and feature preservation where two filtering blocks are used in the new algorithm. A multistage tree-structured filter based on central weighted median filters is used for enhancing the images. The other used filter is a dispersion edge detector. First, the noise-reduction properties of the multistage tree-structured filter are compared to the median and the central weighted median using square and variable shape adaptive kernels. Second, the performance of the multistage tree-structured filter cascaded with the dispersion edge detector are compared to the performance of the dispersion edge detector, the Sobel edge detector, and the single median filter cascaded with the dispersion edge detector. The results indicate that the proposed filter is better than other filters in terms of feature preservation, noise reduction, and edge detection.

The fundamental property of the weighted median filter is based on noise reduction ability which is decreased by increasing the image feature preservation
ability as the central weight is more emphasized. The weighted median filter with only central weight is called Central Weighted Median Filter (CWMF).

The direction of these kernels is adaptively changed pixel by pixel resulting in an image independent process. For a multistage tree-structured filter with $N$ stages, the first stage of the filter consists of $n$ number of CWMFs which is equal to the number of different shape kernels where the base image is used as an input to each sub filter. The second stage of the filter consists of fewer CWMFs where their inputs are a mixture of an even number of the first stage outputs and the base image. By repeating the previous procedure and by reducing the number of CWMFs every stage, several intermediate stages can be added to the multistage tree-structured filter.

Pettazzoni et al. [67] proposed a modular method which divides the processed data into functionally autonomous modules for automatic microcalcification detection. Spatial non-linear filters are used as modules that perform the selection of ROI. Pixels with specific statistical local features are selected by the first filter. The only pixels that have specific properties on the local standard deviation are preserved. Then, the sharp variations unrelated to small close objects are reduced by using a second filter that checks the local mean values of gradient components. Briefly, the proposed method applies a combination of different non-linear filters in order to simplify system maintenance and consistency and make identification of clustered microcalcifications precise.

Bhangale et al. [68] analyzed image texture using a Gabor filter bank for microcalcification detection. The filtered images are obtained using a set of Gabor
filters bank with different orientations and central frequencies. Then, the binary images are obtained by thresholding the filtered images with a histogram-based threshold method. Then, their features are extracted as feature vectors. The images are segmented using a k-means clustering method with a Euclidean distance.

2.4.3 Iris Filter

Kobatake et al. [27] proposed the iris filter to detect lesions as suspicious regions with a low contrast compared to their background. The proposed filter has the ability of features extraction of malignant tissues. It can enhance rounded convex regions such as suspicious tissues whatever their contrasts. It is sensitive to tumors that have an output greater than or equal to the seventh largest output in each tested image. The results indicate that all tumors are expected and enhanced even if they have a weak contrast compared to their background.

The iris filter uses an orientation map of gradient vectors. It can be explained in a two-dimensional continuous space where it is applied to the image gradient vector field. The gradient vector can be computed using generation of gradients in two orthogonal directions where a 3 x 3 Prewitt edge detector is used for the base images.

2.5 Enhancement Techniques

Most image enhancement techniques are surveyed in this section.
2.5.1 Preprocessing Methods

Image enhancement can be used as a preprocessing step to improve the detection algorithm performance and reduce its computational complexity.

Mudigonda et al. [17] proposed a technique for mass detection. It is based on recursive Gaussian low pass filtering and sub sampling operations as preprocessing steps to reach the required level of image smoothing. By using a separable Gaussian kernel of 15 pixels wide, the image is initially smoothed and reduced to 64 gray levels and threshold at 30 different levels with a step-size decrement of 0.01. The selected Gaussian kernel width refers to the total width of its support. The image is threshold at very small intensity to create a map of iso-intensity contours. The iso-intensity contours map is used to recognize the closed contours that can be used to represent the isolated regions in the image by employing chain code principles. The minimum threshold level is defined as the intensity level at which the masses and other dense tissues appear to merge with the surrounding breast parenchyma. A five texture features based on gray-level co-occurrence matrices are computed, and the features in a logistic regression technique are used to classify the successfully segmented mass regions as malignant or benign.

2.5.2 Conventional Techniques

Most conventional image enhancement techniques are surveyed in this section.
2.5.2.1. Contrast Stretching

There are many methods proposed in the literature to enhance mammographic image contrast. The easiest way to do it is by adjusting its histogram. The resulting image has a greater separation between foreground and background gray-level distributions. A typical contrast stretching transformation is shown below:

\[
y = \begin{cases} 
\alpha x, & 0 \leq x < a \\
\beta(x - a), & a \leq x < b \\
\gamma(x - b), & b \leq x < L 
\end{cases}
\]  

(2)

Where, the slopes \(\alpha\), \(\beta\) and \(\gamma\) are greater than unity in the stretched regions. The parameters \(a\) and \(b\) are estimated by using the image histogram. The parameter \(L\) represents the image maximum gray level.

2.5.2.2 Histogram Equalization

The image is modified to obtain a desired shape for its histogram. The major use of these techniques is to stretch the low contrast images with narrow histogram. The resulting histogram has more uniform distribution of the image gray level intensities. Based on the information theory, the uniform distribution of the image histogram achieves the maximum entropy meaning that the image contains the important information. Therefore, to maximize the tumor’s information, the image gray levels are distributed in a way to have more uniform histogram.
2.5.2.3 Gradient Operators

There are many types of masks (filters) that can be used to enhance the image. One of the most commonly used is the unsharp masking [18]. The others are Sobel gradient operators [98] and Canny edge detector [38]. To enhance the image contrast, the filter function should provide $40^\circ$–$50^\circ$ slopes in the low input range ($0 - 0.1$) to reduce noise. The filtered image will be sharper.

2.5.2.4 Fixed-Neighborhood

The previously stated enhancement techniques are considered as global-based approaches. Since mammograms have inhomogeneous background, local-based enhancement approaches are more preferable. They are based on statistical analysis that estimates the background and then suppresses it to increase the image local contrast.

2.5.2.5 Adaptive Neighborhood

Enhancement techniques can be classified into fixed-neighborhood and global-neighborhood techniques. They may adapt to the local features of the neighborhood, but do not adapt the neighborhood size to the local features. Suspicious regions have features that vary in size and shape. Therefore, their features cannot be enhanced by fixed-neighborhood or global-neighborhood techniques. Adaptive-neighborhood methods enhance the suspicious regions with respect to their local background. These region-based techniques enhance image details and reduce
its artifacts. They can identify microcalcifications more effectively in regions where the contrast between the breast tissues and microcalcifications is very low [20].

2.5.2.6 Background Removal

Background removal method is used to enhance microcalcifications by removing the slowly varying portions of the processed image and increasing gray-level variations. It is done by subtracting the low pass filtered image from the base image. There are many methods used to estimate the image background. The most used ones are Morphological processing and partial wavelet reconstruction.

Mammmographic images are reconstructed from their modified wavelet coefficients at specified levels by using nonlinear operators [21]. Kim et al. [22] proposed an adaptive-enhancement method by using first derivative and local statistics. Li et al. [23] used fractal approach to enhance microcalcifications. The proposed approach is based on the property that states that the breast background tissues have high local self-similarity than microcalcifications.

2.5.3 Adaptive Enhancement

Kim et al. [26] developed an image enhancement technique based on local statistics and image gradient. The proposed technique includes three steps. The first step reduces image artifacts that may be misclassified as microcalcifications. The first derivative is in the second step to estimate the gradient images. The third one adds the adaptively weighted gradient images together to highlight the important features.
Image enhancement can be adaptively done by using local statistics of the image. Therefore, image noise is reduced and important features are enhanced.

2.5.4 Density-Weighted Contrast Enhancement

Petrick et al. [32] proposed a new method to segment and extract suspicious regions by implementing a new adaptive Density-Weighted Contrast Enhancement filter in combination with a Laplacian-Gaussian edge detector. First, images are enhanced by using the new filter. Second, the Laplacian-Gaussian edge detection algorithm is used to identify the boundaries of the suspicious regions. Third, a classification algorithm is used to extract their morphological features. These features are used to classify regions within the image.

2.5.5 Modal-Based Image Enhancement

Highnam et al. [33] proposed a model to estimate the signal's scatter component of the image pixels. Then, this estimation is used to enhance the image. First, breast tissues are classified into suspicious and fat tissues. Second, the theory states that the scattered radiation amount is associated with the energy conveyed to the surrounding neighborhood used to estimate the scatter model. Third, a published empirical data is used to approximate this complex relationship. This approximation, which takes the form of a weighted mask, is convolved with the total scattered image. The scattered function takes the result of the previous step as an input, and approximates it using three reference variables and produces a scatter estimate. By
using this scatter estimation, principal components can be computed and used to form an image recognizable by a radiologist.

2.6 Noise Equalization

Image noise equalization is considered an important preprocessing step for any automatic detection system of microcalcifications.

Veldkamp and Karssemeijer [28] proposed an adaptive method for noise equalization as a preprocessing step for automatic detection of microcalcifications. Other different methods are investigated to estimate the image noise, and used to optimize the proposed method. A model for additive noise is used, and the image grayscales are divided into sample intervals to estimate noise as a function of the grayscales. The results indicate that the proposed method has better detection results than fixed noise equalization methods. By obtaining noise as a function of the gray levels, local contrast features can be normalized.

McLoughlin et al. [29] extended the square root method for noise equalization in mammographic images. Based on the property that quantum noise is prevalent in mammographic images, a noise model is established. Noise estimation as a function of the gray levels is enhanced by using a truncated distribution technique to compute the noise statistics. The results indicate that the extended method improves image noise estimation. Using a theoretical approach for noise variance dependency on gray level intensities enhances noise equalization. The maximum mean squared error is reduced by excluding pixels near the breast edge from the noise estimation stage.
2.7 Fractal Modeling

Fractal Modeling is used to model the mammographic images. Therefore, it provides the ability of using a matched filtering stage to enhance microcalcifications against the background.

Bocchi et al. [30] developed an algorithm for microcalcification detection and classification. By analyzing the spatial arrangement of the detected suspicious regions, the presence of microcalcification clusters can be determined. The existing tumors are detected by using a region-growing method along with a neural network-based classifier. Then, microcalcification clusters are detected and classified by using a second fractal model. The matched filtering stage is implemented by using a Fractional Brownian Motion filter. This filter, along with an artificial neural network-based classifier, shows high accuracy and sensitivity in microcalcification detection, followed by a correct segmentation procedure. By modeling microcalcification clusters as a fractal dust, effective descriptors are extracted from their spatial arrangements.

Li et al. [31] proposed a fractal modeling algorithm for microcalcification enhancement and to model the parenchymal and ductal regions. The fractal image modeling theory states that deterministic fractal objects can be used to model images. Fractal image modeling is based mathematically on the iterated function systems and the collage theorem. Image background regions are modeled and analyzed by using the proposed method. Affine transformation parameters are used to model
parenchymal and ductal regions. By subtracting the modeled image from the base image, microcalcifications are enhanced. After comparing the results with those of morphological operation and partial wavelet reconstruction methods, results indicate that fractal modeling is an effective method for microcalcification enhancement.

2.8 Image Segmentation

Microcalcification segmentation has two advantages. One is to determine the locations of suspicious regions in order to assist radiologists as a computer aided detection system. The other one is used for abnormalities classification into benign or malignant as a computer aided diagnosis system. Segmentation is based on selection of threshold parameters and filter window size. Local thresholding is done by choosing suitable threshold values for sub-images.

Kallergi et al. [34] used region-growing and local thresholding methods to measure the parenchymal densities. First, these methods are evaluated and optimized on images with varying contrast and X-ray exposure conditions. Then, they are compared. The results show that the local thresholding algorithm has the greatest stability. However, it is more dependent on parameter selection.

Many approaches can be used for image segmentation. One approach is based on edge detection methods like Robert operator [35], Sobel operator [36], Prewitt operator [37], Canny operator [38], and Laplacian operator. Another approach is based on mathematical morphological operations like top-hat transformation, erosion, complicated morphological filters, and multi-structure elements [39-41]. And third,
stochastic approaches that deal with geometrically analytic aspects of image analysis problem have been used to segment microcalcifications [42-43]. Bayesian and stochastic methods provide a framework for images modeling and in expressing prior knowledge. Markov Random Field model is used to deal with the spatial relations between the obtained labels from an iterative segmentation process [44]. Region-based methods are other approaches that group pixels into homogeneous regions [45]. The fractal objects in a fractal model approach, which are attractors of sets of 2-D affine transformations, are used to model image context [46]. In multiscale analysis approach, the wavelet-based filters are used to transform the image from spatial domain into frequency domain for further processing [47].

2.8.1 Segmentation and Thresholding

Byrd et al. [49] implemented three different segmentation algorithms on mammographic images. Their performance is compared against manual segmentation results produced by two expert radiologists. They are Region Growing Combined with Maximum Likelihood Modeling, Snake Model, and a Standard Model. To evaluate their performance against expert segmentation results, a comprehensive statistical validation protocol is used.

Melloul and Joskowicz [50] developed a new method for microcalcification segmentation that consists of two steps. First, a multiscale morphological operation is used to remove image background. Second, an entropy based thresholding method that employs a 3-D co-occurrence matrix is applied. The proposed method is fully automatic, independent of local statistics, and parameter-free.
Hatanka et al. [70] proposed a thresholding algorithm to detect masses in mammographic images. This method can detect masses with a partial loss of region especially the ones located close to the edge. During the template matching process, the masses that have partial loss are recognized by their similarity to a sector-form model. Four features are applied to compute this similarity. They are standard deviation, mean, standard correlation coefficient defined by a sector-form model, and concentration feature determined from the density gradient.

2.8.2 Markov Random Field

Li et al. [51] developed a method for tumors detection that consists of a segmentation step followed by a classification step. In the segmentation step, an adaptive thresholding method is used to extract regions of interest from the images. Then, a modified Markov random field procedure is used for additional segmentation. In the classification step, a fuzzy binary decision tree based on a series of radiographic and density-related features is used to classify the segmented regions into suspicious and normal regions.

2.8.3 Deformable Models

Valverde et al. [52] proposed an algorithm to segment vessels in order to reduce number of vascular false positives during microcalcification detection stage. Some of the difficulties for fully automatic vessel detection procedure are the presence of noise, the low and varying contrast of vessels, and the variability of the background. The noise reduction algorithm consists of two steps. First, the optimum
edge detector and threshold value are selected by using parameter analysis from edge detection step. In order to improve the signal to noise ratio, the previous step is done on the whole image. Second, the local approach removes the remaining noise from the previous step and performs vessel segmentation by using a deformable model.

2.8.4 Pectoral Muscle Segmentation

The pectoral muscle represents the largest density region in mammographic images. Its presence may affect the results of intensity-based detection systems. After extracting the segmented pectoral muscle, its local analysis may recognize the presence of abnormal auxiliary lymph nodes which is the only proof of the occult breast carcinoma presence.

Ferrari et al. [54] developed a method for pectoral muscle segmentation based on a multiresolution technique using Gabor wavelets. The developed method gives better accuracy than Hough transform does. First, a set of Gabor filters is convolved with the region of interest for the purpose of enhancing the pectoral muscle edge. Second, the magnitude and phase images are computed using a vector-summation method. Third, the value of each pixel is propagated in the phase direction. Then, the initial pectoral muscle edges are detected. Finally, the true pectoral muscle edge is detected using a post-processing step.

2.9 Wavelet Transform

A powerful framework for multiresolution analysis can be provided through Wavelet transform. The ROIs are mapped into a series of coefficients, which
constituted a multiscale representation of them, using the discrete wavelet transform. A set of features can be extracted from each scale in order to obtain the features reflecting scale-dependent properties. The most commonly used features are energy, entropy, and norm of the coefficients [55- 58].

Bruce and Adhami [53] used the modulus-maxima technique of discrete wavelet transform as a feature extraction technique. These features are fed into a Euclidean distance classifier in order to classify masses as round, nodular, or stellate. Three new features are developed; standard deviation, mean, and the Lipchitz sum. These features are compared with the traditional uniresolution shape features in order to determine their class discriminating abilities. These features provided various scales of evaluating the masses shapes. By utilizing a statistical classification system along with Euclidean distance measures, the performance of multiresolution features improves significantly the classification accuracy.

Chang and Laine [55] developed an artifact-free enhancement method based on multiscale representations of the image. First, a fast wavelet transform method is used to decompose the image. The energy and phase information, at each level of analysis, are calculated via a set of independent steerable filters. Then, the energy measure is weighted with the ratio of projections of local energy within a specified region to measure the coherence within each level. Then, each projection is calculated into the mask central point with respect to the total energy within that mask. Finally, a nonlinear operation is applied to adapt transform coefficients within different levels...
of analysis. An inverse fast wavelet transform is used to reconstruct the modified coefficients resulting in an improved visualization of suspicious regions.

Lemaur et al. [58] used high Sobolev regular wavelets to detect clustered microcalcifications. The new wavelet is compared with classical wavelets assuming the same support width. The Sobolev regularity property refers to the fractional derivatives of the image and to its singular spectrum which represents more sophisticated structure that is contrary to the classical smoothness of wavelet transform.

Yu et al. [69] proposed an automatic detection procedure of clustered microcalcifications. The target for the proposed method is one false positive per image. First, wavelet features and gray level statistical features are used to segment out possible microcalcification clusters and then they are labeled according to their spatial connectivity. Second, the extracted features from labeled microcalcification clusters are used to detect the individual microcalcifications. Then, a feed forward neutral network algorithm is used as a classifier.

2.9.1 Wavelet Expansion

Heine et al. [59] developed a statistical multiresolution technique to separate normal regions from abnormal regions. First, a wavelet expansion is used to decompose base images into a set of independent images of which each one contains different levels of features. If microcalcifications are detected, this means that some of the image components are used in suspicious regions detection stage. A simple probability distribution function is used to model the statistical analysis for each
selected expansion component. This function is used to develop a statistical analysis that allows for the identification of normal regions. Only one parameter is used to represent the distribution function. By using its values, a statistical method can be developed to set detection error rates. After determining the summary statistic, spatial filters are applied separately for each selected expansion image. Then, the correlated regions that have the normal statistical model are rejected and the other regions are flagged. The flagged regions are merged to generate a detected output image that consists only of suspicious regions.

2.9.2 Wavelet Sub-Band Decomposition

Wang et al. [60] proposed a method for microcalcification detection based on wavelet-sub-band decomposition. Since microcalcifications look as small clusters with higher intensities compared to their neighbors, a detection system can preserve their features by employing a suitable image transform that localize the image features in the base and the transform domain. Since microcalcifications appear as high-frequency components in the image spectrum, their detection can be done by decomposing the image into different frequency sub-bands, rejecting the low-frequency sub-bands, and then, the image are reconstructed from its high frequency sub-bands.

Strickland and Hahn [61] used wavelet transform to develop a two-stage algorithm for microcalcification detection and segmentation. First, an undecimated wavelet transform is used since it is simpler than conventional filter bank transformations with no down sampling. This transform keeps the low-low (LL), low-
high (LH), high-low (HL), and high-high (HH) sub-bands at full size. Microcalcification detection occurs in the HH and the combination of LH+HL sub-bands. For finer resolution, four octaves are estimated with 2 inter-octave voices. By suitable selection of the wavelet basis, microcalcification detection can be optimized. The second stage is designed to detect microcalcification boundaries and to reduce the limitations of the Gaussian assumption. Before computing the inverse wavelet transform, detected pixel locations are dilated and then weighted. The resulting image shows that the individual microcalcifications are greatly enhanced so a simple threshold algorithm can be applied to segment them.

2.10 Fuzzy Logic Techniques

Cheng et al. [65] proposed a method for microcalcification clusters detection. First, images are normalized. Second, they are fuzzified by using fuzzy set theory and fuzzy entropy principle, and then enhanced. Then, microcalcification sizes and locations are detected by using scale-space and Laplacian-of-a-Gaussian filter methods. The algorithm performance is evaluated using a free-response operating characteristic curve. Experimental results indicate that the proposed method is able to detect microcalcifications in very dense images.

2.10.1 Normalization

Since images have different contrast and brightness, these images are first normalized, as a preprocessing step, in order to achieve computational consistency and decrease their variations. First, images are mapped into a fixed intensity range
Some degree of fuzziness is associated with mammographic images such as ill-defined shapes, indistinct borders, and different densities. In order to reduce number of FN and FP rates, the tested mammograms are enhanced.

2.10.2 Fuzzy – Genetic

Pena-Reyes and Sipper [71] applied a combined fuzzy-genetic approach with new methods to the Wisconsin Breast Cancer Database as a computer aided diagnosis system. This combined approach achieves the highest classification performance and a confidence measure to the output image. Also, this approach employs a few simple rules. Therefore, it can be regarded as human interpretable.

2.11 Statistical Analysis

Kim and Park [66] developed a surrounding region-dependence method as a new texture-analysis method and compared it with other conventional texture-analysis methods such as the gray-level run-length method, the spatial gray level dependence method, and the gray-level difference method. The extracted features by the previously mentioned methods are exploited in order to classify suspicious regions as either negative regions containing normal tissues or positive regions containing abnormal tissues. The suspicious regions are classified by using a three-layer back propagation neural network classifier. A receiver operating-characteristics curve is used to evaluate the neural network classifier results. Based on classification accuracy and computational complexity, the developed method is superior to the conventional texture-analysis methods.
2.12 Classifiers

Most classification techniques are surveyed in this section.

2.12.1 Artificial Neural Networks

Zheng and Chan [72] combined artificial intelligent methods with the discrete wavelet transform (DWT) to build an algorithm for mass detection. The combined artificial intelligent methods include: multiresolution Markov random field, fractal dimension analysis, and dogs-and-rabbits algorithm. The fractal dimension analysis method is used as a preprocessing step to approximate the locations of suspicious regions. The segmentation stage at the Low-Low sub-band of a three-level DWT decomposition is done by using the dogs-and-rabbits clustering algorithm. Each region is classified into suspicious or normal ones by using a tree-type classification strategy.

Setiono [73] used a simple pre-processing technique to improve the accuracy of the trained neural networks and their classification rules. In the data pre-processing step, the image attributes are selected and then the regions with missing attribute values are removed. The extracted classification rules are more accurate than those generated by other rule generating methods.

Bocchi et al. [74] developed a method for microcalcification detection and classification. The analysis of spatial distribution of the detected lesions is used to determine the type of microcalcification clusters. Each image is described by using a fractal model. Then, a matched filtering stage is used to enhance microcalcification clusters. The existing lesions are detected by using a region growing algorithm along
with a neural classifier. Then, their spatial distribution is analyzed by using a second fractal model. Therefore, the existence of microcalcification clusters can be classified and detected. The results indicate that the proposed method achieves high correct classification rates.

Verma and Zakos [76] used fuzzy-neural and feature extraction techniques to develop an algorithm for detecting and classifying of microcalcification clusters. A combination of three feature extraction methods is used to differentiate between benign and malignant microcalcification clusters. They are entropy, number of pixels, and standard deviation. The microcalcification clusters are detected by using a fuzzy technique in combination with the three mentioned features. Then, a neural network is used to classify the pattern into benign or malignant. These intelligent systems give the user the ability to identify, detect, enlarge, zoom, and measure distances of areas in mammographic images.

Zadeh et al. [77] developed two methods of classifying microcalcifications as benign or malignant, and compared them with each other. The first method extracts 17 shape features from each image. Then, the extracted features are related to the shapes of individual or clustered microcalcifications. The second method extracts 44 texture features from each image using co-occurrence matrix of Haralick. Then, a genetic algorithm is used to select suitable features from each set. A k-nearest neighbor classifier and a malignancy criterion are used to create ROC curve that will be used to compare performance of the two methods. The results indicate that the performance of the feature shape based method is better than that of Haralick method.
2.12.2 Pattern Recognition Algorithms

Difference in image analysis encourages utilizing computerized pattern recognition techniques to assist in the assessment of radiographic features.

Fogel et al. [78] used artificial neural networks (ANNs) to analyze radiographic features from mammographic images. For each image, twelve radiographic features are quantified by a human expert depending on predefined rules. Evolutionary programming is used to train ANNs with suspicious regions of various complexities. A statistical cross validation procedure is used to indicate the presence of a malignancy region given a vector of scored input features. The results indicate that ANNs with two hidden nodes performed better than ANNs with only one hidden node. Therefore, small ANNs can be used successfully in breast cancer diagnosis.

2.12.3 Hybrid Neural Networks

Papadopoulos [80] proposed a hybrid intelligent method for the detection of microcalcification clusters consisting of three main steps. Preprocessing and segmentation, detecting regions of interest, and feature extraction and classification. An intelligent system is used to reduce number of false positive cases. This system includes a rule-based and a neural network algorithm. The classification stage computes 22 features that refer to microcalcification clusters type. Then, principal component algorithm is used for additional reduction of features.
2.12.4 Modular Neural Networks

Li et al. [81] proposed a method that enhances suspicious microcalcification clusters and then detect them. The proposed method can be decomposed into three machine learning tasks: building of the featured knowledge database, classified and/or unclassified data points mapping, and intelligent user interface development. Then, a decision support algorithm is constructed in order to assist the radiologist in clusters detection. A mathematical feature extraction module is used to build the knowledge database from all the enhanced microcalcification clusters locations. Then, the normal clusters mixtures and decision boundaries are learned in order to obtain the best mapping of the data points. As a decision support, a visual clarification of the decision making is further invented based on the knowledge base projections using PCA.

2.12.5 Neuro-Fuzzy

Grohman and Dhawan [82] proposed a new convex-set-based neuro-fuzzy procedure for classification of suspicious regions. Based on its estimation of the feature space, the proposed procedure offers better performance than the backpropagation (BP) algorithm. In order to evaluate its performance; the classification performance, computational complexity, and structural efficiency are analyzed and compared with those of the BP algorithm. The results show that this procedure ability to infer knowledge is better than that of backpropagation algorithm in all the tested images.
2.12.6 Fuzzy K-Nearest Classifier

Sekar et al. [83] used the fuzzy k-nearest neighbor classifier as a fuzzy logic technique for predictive decision and evaluation of the tumor signs. Then, it is compared to a statistical method, namely logistic regression, and an artificial neural-network tool, namely multilayer feedforward backpropagation neural networks. The results indicate that their proposed method produce the highest predictive accuracy among the three tested methods.

2.12.7 Convolution Neural Networks

Gurcan et al. [84] evaluated the effect of selecting optimal neural network algorithm on the performance of microcalcification detection system. A computer program is developed for an automatic detection of microcalcification clusters. In a previous work, the authors found that the proper selection and training of convolution neural network (CNN) parameters reduce false-positive rates and enhance the diagnosis system accuracy. The effectiveness of using an automated optimization method for the CNN is evaluated in improving the accuracy of the detection system. Then, it is compared with a manually selected CNN. The results indicate that an optimized CNN can efficiently decrease the number of FP rates and improve the accuracy of the detection system.

Christoyiani et al. [85] used a radial basis function neural network for fast detection of circumscribed masses. The proposed method can differentiate between suspicious and normal tissues, classify the images as suspicious or not, and perform sub-image windowing analysis in order to detect the mass locations. Then, by
applying a set of criterion, square regions that include suspicious masses are marked
as regions of suspicion. In order to reduce the overall algorithm processing time, fast
feature extraction module and a neural classifier are used.

2.12.8 Self-Organizing Map

Markeya et al. [86] used a self-organizing map algorithm to identify
microcalcification clusters in a huge heterogeneous images database. Then, a
constraint satisfaction neural network is used to describe the resulting clusters. These
clusters could be architectural distortions, masses, or microcalcifications. The masses
and microcalcifications groups are divided into seven clusters of masses and three
clusters of microcalcifications. To identify benign clusters that will be candidate for
further analysis by a radiologist rather than biopsy, a feed-forward backpropagation
artificial neural network is used.

2.12.9 Enhanced Rough Set Approach

Hassanien and Ali [87] proposed an enhanced rough set technique for feature
reduction and classification rules extracted from a database. In the first step, the
contrast and edges of the images are enhanced as a pre-processing step. Then, the
regions of interest are extracted by using a segmentation algorithm. In the second step,
the co-occurrence matrix is used to extract five features and store them in a feature
vector. Then, the reducts with minimal number of features are extracted. In the third
step, the generated reduct sets are used to extract decision rules. Finally, a classifier
model and a quadratic distances similarly function are used during the matching
process. The significance of the rules is evaluated using a proposed statistical test. The experimental results indicate that the total number of decision rules is less than that of a well-known decision trees and neural network classifier models.

2.12.10 Fuzzy Classifier with ICA

Lin and Cheng [100] developed a method based on fuzzy neural networks that uses small number of fuzzy rules for handling the embedded information of a system using the trained data. The developed method applies Takagi-Sugeno-Kang model to initialize fuzzy rules and a neural network algorithm for the learning stage. ICA model is used to select the proper number of clusters for the fuzzy neural network. Experimental results indicate that the developed method improves the convergence speed, the accuracy, and simplified the neural network structure.

In general, fuzzy neural network is used for the classifying stage after the feature selection stage by ICA in the literature. In this thesis, a novel algorithm is developed to utilize a simple fuzzy classifier, but proved to be very powerful with the adaptation of ICA model for the detection of suspicious regions in mammography.

2.13 Linear Transformation Methods

2.13.1 Principal Component Analysis-PCA

Abdel-Qader et al. [98] proposed an algorithm using principal component analysis to identify suspicious regions in mammographic images. Linear structure versus curve modeling (Hough transform) is employed as a pre-processing step for
PCA. Euclidean versus Chebyshev distance measures are used as a classifier. The proposed algorithm performance is based on percentage of correct classification, false positive and false negative rates. Experimental results indicate that using Hough transform prior to PCA is better than linear modeling. Also, Euclidean distance measure improves the algorithm accuracy better than Chebyshev distance measure.

2.13.2 Independent Component Analysis

Bell and Sejnowski [94] applied the infomax learning algorithm to a set of natural scenes to estimate their localized and oriented visual filters (features). The applied method proved that their independent components are edge filters. Compared to results from decorrelation-based methods such as PCA, the ICA filters are sparsely distributed outputs and similar to Gabor filters.

Koutras et al. [89] proposed a new feature extraction method for mammographic images. The proposed method uses ICA to extract region features. The suspicious tested regions are described by using the linear transformation coefficients that result from the original trained regions. First, PCA is used as a preprocessing step for dimensionality reduction of features vector. Then, a Radial Basis Function neural algorithm is used as a classifier. The proposed method estimates the unsupervised learning algorithm parameters by choosing the marginal p.d.f. of the source regions that follow the hyperbolic cosine distribution.

Üzümçü et al. [90] used ICA as an alternative shape decomposition algorithm for PCA. The results indicate that ICA algorithm has two advantages over PCA. First, ICA does not require a Gaussian distribution for the data. Second, it has the ability to
describe localized shape variations. However, a method for ordering the resulting independent components is required. The results indicate that the ICA accuracy in border localization is better than that of PCA.

2.13.3 Feature Extraction Methods

Mudigonda et al. [63] developed an algorithm for mass detection. Gaussian smoothing and sub-sampling operations are used as preprocessing steps. The masses are segmented by creating intensity links from the mass central portions into their neighbors. New methods for analyzing oriented textural information are proposed. Extracted features based on pixels orientation across the mass margins are used to classify the detected regions. The successfully segmented mass regions are classified as either benign or malignant by using gray-level co-occurrence matrices. These matrices calculate five texture features and use the features of a logistic regression technique.

Gulsrud and Husoy [64] extracted features from images and used them to differentiate between texture representing suspicious regions and texture representing normal regions. An algorithm that extracts texture features and employs a single filter based on Fisher criterion is proposed. Fisher criterion has excellent feature extraction because it uses both the mean and the variance of the image features. The effects of compressing the data on the performance of the proposed method are also explored. Joint Photographic Experts Group compression method is used to compress the tested images at different ratios. The results indicate that the proposed method had a similar
detection performance in both uncompressed and the compressed images by a factor of four.

2.14 Discussion of Literature

Results analysis is based on ROC and FROC curves analysis.

2.14.1 ROC and FROC Curve Analysis

The Receiver Operating Characteristic (ROC) curve is used to measure the performance of any proposed algorithm by plotting its sensitivity (true-positive rate (TP)) versus its specificity (false-positive rate (FP)). The area under this curve is an accepted method of comparing any classifier performance [88]. The FP and TP rates are both specified in the interval [0.0, 1.0]. A perfect detection algorithm has an area under the curve (AUC) of 1.0, as a TP rate of 1.0 and a FP rate of 0.0. The trapezoid rule is used to estimate the AUCs for discrete operating points.

In order to evaluate true-positive detection, the tumor must be localized. Therefore, FROC is used for this case as a plot of the TP rates versus the average number of false positives per image. However, both ROC and FROC do not address the density of images and are not easy to transform the subjective measurements (radiologist observations) to an objective curve.

2.14.2 Sensitivity versus Specificity

The use of overall system accuracy as an assessment approach assumes equal error costs, i.e. that a false negative rates are equally important as a false positive rates. Therefore, the performance of such systems is best evaluated in terms of their
specificity and sensitivity. These assessments are based on the assumption that a
tested sample always results in one of the following four classes:

i. False Positive (FN): the system classifies a negative one as positive;
ii. False Negative (FP): the system classifies a positive one as negative;
iii. True Positive (TP): the system classifies a positive when it is truly positive;
iv. True Negative (TN): the system classifies a negative when it is truly negative.

The sensitivity of a detection system is defined as the ratio between the
number of true positive classifications $T_P$ and the number of positive tissues in the
tested samples where $F_N$ is the number of false negative classifications.

$$\text{Sensitivity} = \frac{T_P}{T_P + F_N}$$  \hspace{1cm} (16)

The specificity of a detection system is defined as the ratio between the number
of true negative classifications $T_N$ and the number of negative tissues in the tested
samples.

$$\text{Specificity} = \frac{T_N}{T_N + F_P}$$  \hspace{1cm} (17)

The overall accuracy of a detection system is defined as the ratio between the
total number of correctly classified tissues and the complete number of tested
samples.

$$\text{Overall Accuracy} = \frac{N_r}{N} \times 100\%$$  \hspace{1cm} (18)
3. LINEAR TRANSFORMATION METHODS

3.1 Principal Component Analysis

Principal Component Analysis (PCA) is sometimes called Karhunen-Loeve transformation. PCA is a decorrelation based technique that finds the basis vectors for a subspace in order to:

i. Maximizing the projected data retained variance.
ii. Results in uncorrelated projected distributions.
iii. Minimize the least square reconstruction errors
iv. Minimize the entropy.
v. Describes the information in its best form.
vi. Choose the most important information.

The PCA algorithm consists of two phases. The first phase finds \( v \) uncorrelated and orthogonal vectors. The second phase projects the given data set into a subspace spanned by these \( v \) vectors [98].

The first phase consists of the following steps.

i. Form \( A_{\text{train}} \) matrix with dimension \( N \times M \) where every row represents a sub-image.

ii. Find its transpose then normalize every row by its mean. The normalized matrix is \( P_{M \times N} \)

iii. Find covariance matrix using the following equation.
\[ C_{NNN} = P^T_{NM}.P_{MN} \]  

iv. Let \( \lambda_i \) and \( E_i, i = 1, 2 \ldots N \) be its Eigen values and normalized Eigenvectors that satisfies the equation \( C.E_i = \lambda_i.E_i \) where \( \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0 \). By discarding Eigenvalues less than a threshold \( T \) and retaining the rest, the selected Eigenvectors are called principal components.

The second phase projects the given testing data \( A_{test} \) into a space spanned by the reduced training matrix \( A_{rN}^R \) using the following equation:

\[ W_{vN} = A_{rM}^R.A_{test}_{MN} \]

Images have oriented lines, edges, and other structures that have dependencies of higher-order statistics. These localized structures can be described mainly by their phase spectrum. For example, an edge occurs locally and has its phase spectrum aligned across different spatial frequencies. Correlation-based approaches like PCA are phase-blind, i.e. they can capture only the power spectrum. Therefore, higher-order methods are needed to account for localized oriented structures.

Also, PCA failed to describe a random variable when its behavior is non-Gaussian. Therefore, higher order statistics, such as ICA techniques, are used to compensate for PCA short comings. Higher-order statistics are based on the use of moments and cumulants up to fourth-order to describe any distribution of a random variable. For a random variable \( x \), its probability density functions (p.d.f.) are
described in terms of a set of discrete parameters called moments. The first order moment is the mean value of the variable. The n-th order moment is defined as:

$$\mu_x(n) = E\{x^n\} \tag{5}$$

Central moments provide a set of parameters that describe the manner in which the distribution is about the variable mean value:

$$m_x(n) = E\{(x - m_x)^n\} \tag{6}$$

Therefore, central moments and moments are identical when the mean is zero. In this work, third and fourth order moments are used. Third order central moment measures the skewness of the p.d.f. about its mean value. The skewness is a measure of the asymmetry degree of a distribution. For any distribution, if its left tail is more pronounced than its right tail, the function has negative skewness. Otherwise, it has positive/zero skewness. While, fourth order central moments measure the excess of flatness (kurtosis) of p.d.f. For a random variable, the kurtosis is a measure of the peakedness of its probability distribution. If the random variable has higher kurtosis, most of its variance is due to infrequent extreme deviations.

3.2 Independent Component Analysis

The literature is rich in linear transformation methods. The most used methods are Principal Component Analysis, Projection Pursuit, Factor Analysis, and Independent Component Analysis (ICA). PCA is used as a statistical shape analysis algorithm to describe the main directions of shape variations in the training data. Unfortunately, PCA assumes that the data must have a Gaussian distribution. Also,
PCA is only able to describe the image global shape variations. For these reasons, ICA has been introduced as an alternative shape decomposition algorithm for PCA for two reasons. Its ability to describe localized shape variations and it does not require a Gaussian distribution of the data. But, the resulting vectors are not ordered as they are in PCA. Therefore, it requires using a method for ordering the resulting vectors.

In digital signal processing field, it is difficult to find an appropriate representation for audio, image or other data for tasks such denoising and compression. Most data representations are often based on discrete linear transformations. The most widely used standard linear transformations are the Cosine, Fourier, and Haar transforms. To ideally adapt the linear transformation to the kind of processed data, it is more suitable to estimate the linear transformation parameters from the data itself.

ICA is a new technique developed recently to find a linear representation of nongaussian data so that the data components are statistically as independent as possible. This representation captures the important features of the data in many applications, including feature extraction and signal separation. In contrast to correlation based transformations such as PCA, ICA not only decorrelates the signals by using second order statistics but also reduces higher-order statistical dependencies among them.

The statistical latent variables model is used to define ICA. Assuming that we have \( n \) linear mixtures \( x_1, \ldots, x_n \) of \( n \) independent components \( s_1, s_2, \ldots, s_n \) according to the following equation.
The mammographic image (x) is regarded as a mixture of linear combination of statistically independent source regions (s) where A is the mixing matrix. Its coefficients describe the mixed source regions in a distinctive way and can be used as extracted features.

ICA method is a generative method because it explains how the observed data are generated by a process of mixing the independent components. The independent components are latent variables, because they cannot be directly observed. Also, the mixing matrix is assumed to be unknown. The only observed data is the random vector x, and by using it, both A and S can be estimated under the following assumptions.

i. The independent components are statistically independent.

ii. The independent components must have non-Gaussian distributions.

Then, after estimating the matrix A, its inverse W is computed. Then, the independent components are estimated by using the following equation.

\[ S = W \cdot X \]  \hspace{1cm} (8)

In order to utilize non-Gaussianity in ICA model, a quantitative measure of non-Gaussianity of the mixture variables and their independent components is established. Therefore, two assumptions are made as follow.

i. Centering the data. If this is not the case, then the observed data is centered by subtracting its mean, which makes the model zero-mean. Therefore, its central moments are equal to its moments.
Data have a unit variance. Since both independent components and the mixing matrix are unknown, any scalar multiplier in one of the independent components is cancelled by dividing the corresponding column of A by the same scalar. The solution is by fixing the magnitudes of the independent components. One method assumes that each independent component has unit variance. Then the mixing matrix is adapted to handle this restriction. But, this restriction does not solve the ambiguity of the sign: any independent component could be multiplied by \(-1\) without affecting the ICA model. This ambiguity is, fortunately, not important in most applications.

Actually, one of the preprocessing steps for ICA algorithms is to make this simplification possible. ICA techniques are used in biomedical data analysis, computational neuroscience, bioinformatics, sensor array processing, and telecommunications. ICA initially is used for blind source separation. The term blind indicates that both, the source signals and the way the signals are mixed, are unknown.
4. BACKGROUND ON FUZZY CLASSIFIERS

Zadeh [99] invented fuzzy logic as a new method of handling uncertainty and imprecise information in data. Fuzzy logic is a powerful method for decision making systems and systems with noisy or imprecise data. Table 1 shows the differences between fuzzy sets and classical (crisp) sets.

<table>
<thead>
<tr>
<th>Fuzzy Set</th>
<th>Crisp Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>The object partially belongs to the given set.</td>
<td>The object either belongs to the given set or does not.</td>
</tr>
<tr>
<td>The object membership function degree has a value in the range [0,1]. Where 1 indicates that the object entirely belongs to the given set, 0 indicates that the object does not belong to the given set, other values indicate that the object partially belong to the given set.</td>
<td>The object membership function degree has a value 0 or 1. Where 1 indicates that the object is entirely in the set, 0 indicates that the object does not belong to the set.</td>
</tr>
</tbody>
</table>

Table 1: A Comparison between Fuzzy Set and Crisp Set Theory

Fuzzy rules are more comprehensible than crisp rules because they can be expressed in terms of linguistic concepts. Fuzzy classifier is ideally suited to the labeled observed data to provide interpretable solutions. It handles imprecise data and the resulting fuzzy rules are interpretable, i.e. fuzzy classifier structure can be analyzed through its semantic structure.

Many researchers have developed data-driven fuzzy algorithms for prediction, control, and pattern recognition. Unfortunately, most of these algorithms are designed
for accuracy and result in complex non-interpretable rule bases. Usually, fuzzy classifier consists of the following four stages:

i. Statistical analysis algorithms are used to initialize fuzzy model structure.

ii. Most important features are selected.

iii. The inconsistency analysis of the selected features is transformed into fuzzy sets.

iv. The parameters of the resulting fuzzy model are adapted to optimize the classifier accuracy.

Fuzzy inference engine is built by using rules based on linguistic statements which allow general understanding of the current problem. It takes a series of inference steps to produce the decision. Fuzzy inference engine consists of four components.

i. A Fuzzifier to map crisp input values into fuzzy values.

ii. A fuzzy inference engine that uses fuzzy reasoning to compute output.

iii. A Defuzzifier to map the output of the previous step into a crisp value.

iv. A knowledge base that contains a set of fuzzy rules or membership functions.

There are two different methods for developing of fuzzy classifiers.

4.1 Approximate Fuzzy Rule Base

Approximate Fuzzy Rule Base defines each fuzzy rule as fuzzy membership functions. This work is based on this method. The object membership degree to a fuzzy set defines a membership function. Its domain is the universe of discourse (all
values the object can take). Its range is the interval [0, 1]. One of the simple and most used membership functions is the triangular. As displayed in figure 7.

The object \( x \) has a degree of membership of 0.7 to the fuzzy set “Small”. This means \( x \) has a probability of 70% belongs to this fuzzy set and 30% that does not belong to it. Fuzzy space is a set of fuzzy sets that defines the fuzzy classes for a particular object as shown in figure 8.

Fuzzy space allows the object to belong partially to different classes at the same time. This idea is very useful in cases where the difference between classes is not well defined. For example in mammographic images, the difference between the
normal and abnormal sub-images is not well defined. Fuzzy membership functions are easy to implement and their fuzzy inference engines are very quick.

4.2 Descriptive Fuzzy Rule Base

It uses fuzzy if-then rules to define the linguistic variables and uses labels $A_{ij}$ to represent a discrete set of linguistic fuzzy sets. Fuzzy rule systems have many advantages. First, they do not require large memory storage. Second, their fuzzy inference engines are very quick. Last, each fuzzy rule is easily inspected by the programmer.

4.3 Fuzzy Classifier Structure

Fuzzy classifier is generated by using several steps as presented below,

1. Fuzzy classification rules that describe each class of sub-images are developed. One rule is developed to represent each class. The degree of fulfillment of each rule relates to the sample membership degree. Fuzzy rules have the form:

$$\text{IF condition THEN consequent [weight]} \quad (9)$$

The two ways to express fuzzy rules are:

$$R_i: \text{If } x_1 \text{ is } A_{i1} \text{ and } ... \text{ x_n is } A_{in} \text{ then } Y = \text{Class}_i \text{ [weight]} \quad (10)$$

Where $A_{ij}$ denotes a fuzzy set defined for the j: 1,..., n-th selected feature in the i: 1,..., R-th rule.

Similarly,

$$R_i: \text{If } x_1 \text{ is } \{a_{i1}, b_{i1}, c_{i1}, d_{i1}\} \text{ and } ... \text{ x_n is } \{a_{in}, b_{in}, c_{in}, d_{in}\} \text{ then } \text{Class}_i \text{ [weight]} \quad (11)$$
Where \{a_{ij}, b_{ij}, c_{ij}, d_{ij}\} represent trapezoidal fuzzy set parameters as figure 9 shows.

![Trapezoidal Fuzzy Set Example](image)

**Figure 9: Trapezoidal Fuzzy Set Example.**

ii. The degree of activation of the developed rules is used to compute fuzzy classifier output.

\[
\mu_i(x) = \sum_{j=1}^{n} A_{ij}(x_j)
\]  

(12)

iii. By assigning the class to the fuzzy-rule with the maximum degree of activation, a crisp decision is made, i.e. normal/abnormal or benign/malignant. There are many methods used to determine to which class an object belongs. A simple one is the maximum algorithm. It classifies the object as the class that has the maximum degree of activation.

\[
C_i = \max(\mu_1(x), ..., \mu_R(x))
\]

(13)

### 4.4 Fuzzy Model Initialization

It is used to adapt the membership functions to the inconsistency of the training data. It consists of the following steps.
i. A feature ranking algorithm is used to select the relevant features. In this work, ICA algorithm is used for this purpose.

ii. A clustering algorithm is used to obtain fuzzy sets by defining membership functions in the product space of the selected features.

\[ A_y(x_j) = \frac{s_i(x_j)}{s(x_j)} \]  \hspace{1cm} (14)

Where \( s_i(x_j) \) represents number of samples of the current feature that is in the class \( j' \), while \( s(x_j) \) represents the total number of all samples of the current feature \( x_j \).

iii. The membership functions are normalized in order to have normal fuzzy sets.

\[ A_y(x_j) = \frac{A_y(x_j)}{\max_y (A_j(x_j))} \]  \hspace{1cm} (15)
5. FRAMEWORK OF PROPOSED CADD ALGORITHM

In this section, a computer aided detection and diagnosis algorithm of suspicious regions on mammographic images is explored. First, PCA algorithm is used as a dimensionality reduction module. Then, ICA algorithm is used as a feature selection module. Finally, a fuzzy classifier is used to classify tested sub-images into normal or abnormal as a detection system and later to classify the abnormal sub-images into malignant or benign as a diagnosis system.

The extracted sub-images are divided into two groups. The main features of the first group, which is used for the training procedure, are extracted. The other one is used for the testing procedure and its main features are extracted. The following figure shows the main steps of the proposed CADD algorithm.

Figure 10: (a) General Block Diagram of the Proposed Algorithm and (b) Detailed Block Diagram of Proposed Algorithm.
The proposed CADD algorithm framework can be presented as follows.

5.1 Sub-Images Generation

1. The MIAS database has a total of 119 ROS divided as 51 malignant and 68 benign. Two different sets, each one consists of 119 ROS, of abnormal sub-images are cropped and scaled to 35x35 and 45x45 pixels based on the center of each abnormality.

2. Five different sets of normal sub-images are cropped and scaled randomly to 35x35 and 45x45 pixels from the normal MIAS mammograms.

3. Each set of abnormal sub-images is mixed with one set of normal sub-images every time and then divided into two groups; one for training phase and the other one for testing phase as shown in table 2.

<table>
<thead>
<tr>
<th>#</th>
<th>Training set</th>
<th></th>
<th>Testing set</th>
<th></th>
<th>Size-pixels</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ROS</td>
<td>Normal</td>
<td>Total</td>
<td>ROS</td>
<td>Normal</td>
</tr>
<tr>
<td>1</td>
<td>60</td>
<td>59</td>
<td>119</td>
<td>59</td>
<td>60</td>
</tr>
<tr>
<td>2</td>
<td>60</td>
<td>59</td>
<td>119</td>
<td>59</td>
<td>60</td>
</tr>
<tr>
<td>3</td>
<td>60</td>
<td>59</td>
<td>119</td>
<td>59</td>
<td>60</td>
</tr>
<tr>
<td>4</td>
<td>60</td>
<td>59</td>
<td>119</td>
<td>59</td>
<td>60</td>
</tr>
<tr>
<td>5</td>
<td>60</td>
<td>59</td>
<td>119</td>
<td>59</td>
<td>60</td>
</tr>
</tbody>
</table>

Table 2: Different Sets Used To Evaluate the Detection Algorithm Performance

5.2 Training Procedure

4. A training matrix $A_{\text{train}}$ is defined, where each row contains a sub-image, with dimension NxM. Where N is the total number of trained sub-images and M is the dimension of each sub-image (either 35x35 or 45x45).
5. $A^T$ is computed as the transpose of the training matrix.

6. All sub-images in the transpose matrix are normalized by their mean.

7. PCA algorithm is used to obtain $V$ principle components from it. The reduced matrix is $A_{MxV}^R$.

8. The covariance matrix is estimated using the following equation.

$$C_{NXV} = A_{train}^R A_{MxV}^R$$

9. The transpose of the reduced matrix $A_{MxV}^R$ is computed $A_{K}^T$. And then, it is used as an input for the following proposed unsupervised learning algorithm which is used to estimate the separating matrix $W$ and the independent source regions matrix $S$ in an unsupervised mode.

5.3 Unsupervised Learning Algorithm

The following learning algorithm is proposed in order to estimate the separating matrix $W$ and the independent source regions $S$.

9.1 The separating matrix $W$ is initialized to the identity matrix.

9.2 The independent source regions are estimated by using the following equation.

$$S = W \cdot A_{R}^T.$$  \hspace{1cm} (20)

9.3 The minimum mutual information algorithm [101] is used to estimate the non-linear function $\Phi(S)$ which is used to approximate the marginal probability density function of the output regions $S$ in order to achieve a maximal statistical independence of the source regions using $3^{rd}$ and $4^{th}$ order moments of the independent source regions $S_i$. The independent source regions are latent
variables, meaning that they cannot be directly observed. Also the mixing matrix is assumed to be unknown. All we observe is the random vector $A_{\text{train}}$, and we must estimate both $A$ and $S$ using it as follows.

i. Third and fourth order central moments of $S_i$ are estimated as the expected value of normalized observed data $x$.

$$m_i = E(x - \mu)^k$$  \hspace{1cm} (21)

ii. Third and fourth order cumulants are estimated using the following equations.

$$k_{3(i)} = m_{3(i)} \quad \text{and} \quad k_{4(i)} = m_{4(i)} - 3$$  \hspace{1cm} (22)

iii. The non-linear function $\Phi(S)$ is estimated using the following equations.

$$\Phi = f_1(k_3, k_4) \circ S^2 + f_2(k_3, k_4) \circ S^3$$  \hspace{1cm} (23)

$$f_1(k_3, k_4) = -\frac{1}{2} k_3 + \frac{9}{4} k_3 k_4$$  \hspace{1cm} (24)

$$f_2(k_3, k_4) = -\frac{1}{6} k_4 + \frac{3}{2} k_3^2 + \frac{3}{4} k_4^2$$  \hspace{1cm} (25)

Where $\circ$ denotes the Hadamard product of two matrices.

$$a \circ b = a_{ij} b_{ij}$$  \hspace{1cm} (26)

9.4 The change in the separating matrix $\Delta W$ is estimated by using the following equation where $\eta(t)$ represents the learning rate.

$$\Delta W = \eta(t) \left[ I - \Phi(S)S^T \right] W$$  \hspace{1cm} (27)

9.5 If $\Delta W \approx 0$ stop the procedure; this means that the target response (I) is almost equal to the actual response ($\Phi(S).S^T$), otherwise go to next step.

9.6 The weights are updated using the following equation.
\[ W_i(t+1) = W_i(t) + \Delta W \]  

(28)

9.7 The above procedure is repeated recursively.

10. The previous steps will estimate \( S_{vxM} \) and \( W_{vxv} \).

11. The reduced dimensionality extracted features which can be used for the training procedure can be estimated as follows.

i. From Eq. (19), the training matrix can be reconstructed as follows.

\[ A_{\text{rec}} = C_{NxV} \cdot A^T_R \approx A_{\text{train}} \]  

(29)

ii. From Eq. (20),

\[ A^T_R = W^{-1} \cdot S \]  

(30)

iii. Substituting Eq. (20) into Eq. (19).

\[ A_{\text{rec}} = C_{NxV} \cdot A^T_R = C_{NxV} \cdot W^{-1} \cdot S \]  

(31)

iv. There, the reduced dimensionality extracted features can be estimated using the following equation.

\[ R_{\text{train}_{vxy}} = C \cdot W^{-1} \]  

(32)

5.4 Testing Procedure

12. A testing matrix \( A_{\text{test}} \) is defined, where each row contains a sub-image, with dimension \( NxM \).

13. The rows of \( A_{\text{test}} \) are normalized by their mean.

14. The regions in \( A_{\text{test}} \) are projected on the reduced data from the training procedure using the following equation.
15. The reduced dimensionality extracted features which can be used for the training procedure can be estimated using the following equation.

\[ B_{\text{test}, V} = A_{\text{test}, V} A_{\text{MxV}}^R \]  

(33)

\[ R_{\text{test}, V} = B_{\text{test}, W}^{-1} \]  

(34)

5.5 Fuzzy Classifier Modeling

The matrices \( R_{\text{train}} \) and \( R_{\text{test}} \) contain \( N \) rows and each row contains \( V \) selected features of the corresponding sub-image. The proposed algorithm consists of the following steps.

1. Four activation functions \( \mu_{\text{at}}, \mu_{\text{nt}}, \mu_{\text{as}}, \mu_{\text{ns}} \), each one of size \( N \times 1 \), are initialized to 1. Each one represents the degree of activation of the selected features. \( \mu_{\text{at}} \) represents the degree of activation for the abnormal trained sub-images, \( \mu_{\text{nt}} \) represents the degree of activation for the normal trained sub-images, \( \mu_{\text{as}} \) represents the degree of activation for the abnormal tested sub-images, and \( \mu_{\text{ns}} \) represents the degree of activation for the normal tested sub-images.

2. Since the sub-images have different intensities and the goal is to reduce the variation and the computational complexity, the selected features of \( R_{\text{train}} \) and \( R_{\text{test}} \) are mapped to a limited range \([r_1, r_2]\).

3. A clustering algorithm is used to obtain fuzzy sets by defining membership functions in the product space of the selected features using equation 14.
\[ A_y(x_j) = \frac{s_i(x_j)}{s(x_j)} \]

Where \( S_i(x_j) \) represents number of samples of the current feature \( x_j \). \( S(x_j) \) represents total number of all samples in the current feature \( x_j \) or the product space of the current feature.

4. The membership functions are normalized by using equation 15.

\[ A_y(x_j) = \frac{A_y(x_j)}{\max_y \{ A_j(x_j) \}} \]

5. The degree of activation of the developed membership functions is computed for each one of the four proposed activation functions using equation 12.

\[ \mu_i(x) = \sum_{j=1}^{n} A_y(x_j) \]

6. By assigning the class to the fuzzy-rule with the maximum degree of activation, a crisp decision is made, i.e. normal or abnormal and benign or malignant. There are many methods used to determine to which class an object belongs. A simple one is the maximum algorithm. It classifies the object as the class that has the maximum degree of activation. From equation 13, we have

\[ C_i = \max(\mu_i(x),...,\mu_R(x)) \]
6. EXPERIMENTAL RESULTS

6.1 Parameters Selection

Several parameters that impact the proposed CADD algorithm accuracy are investigated in the following sections. Parameters such as number of selected principal components, the learning rate, and mapping range.

6.1.1 Number of Selected Principal Components

Using PCA algorithm to reduce data dimensionality as a preprocessing step for ICA algorithm will affect the total algorithm accuracy. When large number of principal components is selected, the extracted features will have large dimensionality and therefore, will increase the classifier task complexity. However, if a small number is selected, the independent source regions cannot be estimated precisely and fuzzy classifier performance will be degraded. The following table shows simulation results for five tested sets where PC indicates number of selected principal components. For first set, the best learning rate is \( \eta = 0.001 \) and best mapping range of the sub-images is \([0, 16]\). For second set, the best learning rate is \( \eta = 0.0045 \) and best mapping range of the sub-images is \([0, 11]\). For third set, the best learning rate is \( \eta = 0.0011 \) and best mapping range of the sub-images is \([0, 11]\). For fourth set, the best learning rate is \( \eta = 0.0047 \) and best mapping range of the sub-images is \([0, 7]\). For fifth set, the best learning rate is \( \eta = 0.0014 \) and best mapping range of the sub-images is \([0, 9]\). The results from this table indicate that selecting less than 11 principal components achieves acceptable results in all cases which mean less than 0.81% of principal
components are selected for sub-images of size 35x35 pixels and less than 0.5% of principal components are selected for sub-images of size 45x45 pixels.

6.1.2 Learning Rate

The learning rate for computing the change in W for ICA algorithm as stated by equation 27 will determine the speed of convergence for \( \Delta W \). Also, it will impact the total algorithm accuracy. From the results in the previous section, 6 principal components along with a mapping range [0, 11] produce acceptable results for all sets. These parameters are kept fixed for all sets.

<table>
<thead>
<tr>
<th>PC</th>
<th>Set #1</th>
<th>Set #2</th>
<th>Set #3</th>
<th>Set #4</th>
<th>Set #5</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>61.35%</td>
<td>65.55%</td>
<td>84%</td>
<td>69.75%</td>
<td>81.52%</td>
</tr>
<tr>
<td>6</td>
<td>55.46%</td>
<td>64.71%</td>
<td>79.83%</td>
<td>68.07%</td>
<td>84.03%</td>
</tr>
<tr>
<td>7</td>
<td>62.19%</td>
<td>66.39%</td>
<td>78.99%</td>
<td>68.91%</td>
<td>82.35%</td>
</tr>
<tr>
<td>8</td>
<td><strong>66.39%</strong></td>
<td>66.39%</td>
<td>78.99%</td>
<td>65.55%</td>
<td>78.15%</td>
</tr>
<tr>
<td>9</td>
<td>59.66%</td>
<td>66.39%</td>
<td>70.59%</td>
<td>67.23%</td>
<td>76.47%</td>
</tr>
<tr>
<td>10</td>
<td>58.82%</td>
<td><strong>68.91%</strong></td>
<td>80.67%</td>
<td>66.39%</td>
<td>74.79%</td>
</tr>
<tr>
<td>11</td>
<td>63.03%</td>
<td>65.55%</td>
<td>69.75%</td>
<td>63.87%</td>
<td>75.63%</td>
</tr>
<tr>
<td>12</td>
<td>58.82%</td>
<td>60.5%</td>
<td>72.27%</td>
<td>63.03%</td>
<td>78.15%</td>
</tr>
<tr>
<td>13</td>
<td>62.19%</td>
<td>63.87%</td>
<td>70.59%</td>
<td>63.87%</td>
<td>74.79%</td>
</tr>
<tr>
<td>14</td>
<td>63.87%</td>
<td>62.19%</td>
<td>73.95%</td>
<td>62.19%</td>
<td>77.31%</td>
</tr>
<tr>
<td>15</td>
<td>57.98%</td>
<td>63.03%</td>
<td>69.75%</td>
<td>63.03%</td>
<td>73.95%</td>
</tr>
<tr>
<td>16</td>
<td>62.19%</td>
<td>59.66%</td>
<td>68.07%</td>
<td>63.03%</td>
<td>76.47%</td>
</tr>
<tr>
<td>17</td>
<td>62.19%</td>
<td>67.23%</td>
<td>72.29%</td>
<td>63.87%</td>
<td>77.31%</td>
</tr>
<tr>
<td>18</td>
<td>63.03%</td>
<td>60.5%</td>
<td>71.43%</td>
<td>62.19%</td>
<td>76.47%</td>
</tr>
<tr>
<td>19</td>
<td>64.71%</td>
<td>67.23%</td>
<td>72.29%</td>
<td>64.71%</td>
<td>79.83%</td>
</tr>
<tr>
<td>20</td>
<td>62.19%</td>
<td>64.71%</td>
<td>79.83%</td>
<td>62.19%</td>
<td>73.95%</td>
</tr>
<tr>
<td>21</td>
<td>60.5%</td>
<td>66.39%</td>
<td>74.79%</td>
<td>61.35%</td>
<td>80.67%</td>
</tr>
<tr>
<td>22</td>
<td>63.03%</td>
<td>66.39%</td>
<td>78.15%</td>
<td>62.19%</td>
<td>71.43%</td>
</tr>
<tr>
<td>23</td>
<td>63.03%</td>
<td>63.87%</td>
<td>80.67%</td>
<td>63.87%</td>
<td>74.79%</td>
</tr>
<tr>
<td>24</td>
<td>58.82%</td>
<td>60.5%</td>
<td>73.95%</td>
<td>62.19%</td>
<td>80.67%</td>
</tr>
<tr>
<td>25</td>
<td>59.66%</td>
<td>60.5%</td>
<td>68.91%</td>
<td>63.03%</td>
<td>79.83%</td>
</tr>
</tbody>
</table>

Table 3: Number of Selected Principal Components Impact on Algorithm Accuracy Where Learning Rate and Mapping Range of Each Set are Kept Fixed
The figures 11-15 show learning rate impact on five tested sets. It can be concluded that choosing a learning rate close to 0.0045 is acceptable for all tested sets.

Figure 11: Learning Rate Impact on Algorithm Accuracy for Test Set #1 Where Other Parameters Kept Constant.

Figure 12: Learning Rate Impact on Algorithm Accuracy for Test Set #2 Where Other Parameters Kept Constant.
Figure 13: Learning Rate Impact on Algorithm Accuracy for Test Set #3 Where Other Parameters Kept Constant.

Figure 14: Learning Rate Impact on Algorithm Accuracy for Test Set #4 Where Other Parameters Kept Constant.
6.1.3 Moment Algorithm

There are many methods that can be used to compute central moments. Each method affects the algorithm accuracy.

6.1.4 Mapping Range

Selecting a mapping range for sub-images during classification stage will affect the algorithm accuracy. The figures 16-20 show accuracy versus mapping range for five tested sets with 6 principal components along with \( \eta = .0045 \) kept fixed for all sets. It can be concluded from these figures that choosing a mapping range equal to 15 is acceptable for all tested sets.
Figure 16: Mapping Range Impact on Algorithm Accuracy for Test Set #1.

Figure 17: Mapping Range Impact on Algorithm Accuracy for Test Set #2.
Figure 18: Mapping Range Impact on Algorithm Accuracy for Test Set #3.

Figure 19: Mapping Range Impact on Algorithm Accuracy for Test Set #4.
6.1.5 Cropping and Scaling Criterion

The way the normal and abnormal sub-images are cropped and scaled affects the algorithm accuracy. Table 4 shows that each test set has different algorithm accuracy.

6.2 Discussion of Experimental Results

Table 4 shows the proposed CADD algorithm accuracy versus PCA accuracy for the same tested data using a fuzzy classifier. Table 5 shows the ICA algorithm accuracy versus the proposed CADD algorithm accuracy. The algorithm accuracy is defined as the ratio between the total number of correctly classified sub-images and the total number of tested sub-images.
Table 4: FP and FN and Total PCA, CADD Algorithms Accuracy

<table>
<thead>
<tr>
<th>Set</th>
<th>PC</th>
<th>FP</th>
<th>FN</th>
<th>Accuracy</th>
<th>PC</th>
<th>FP</th>
<th>FN</th>
<th>Accuracy</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>19</td>
<td>17.65%</td>
<td>25.21%</td>
<td>57.14%</td>
<td>25</td>
<td>15.13%</td>
<td>18.48%</td>
<td>66.39%</td>
<td>16.19%</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>26.05%</td>
<td>10.08%</td>
<td>63.87%</td>
<td>10</td>
<td>12.61%</td>
<td>18.48%</td>
<td>68.91%</td>
<td>7.89%</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>10.08%</td>
<td>14.29%</td>
<td>75.63%</td>
<td>5</td>
<td>9.24%</td>
<td>6.73%</td>
<td>84.03%</td>
<td>11.11%</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>12.61%</td>
<td>21%</td>
<td>66.39%</td>
<td>5</td>
<td>20.17%</td>
<td>10.08%</td>
<td>69.75%</td>
<td>5.06%</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>11.75%</td>
<td>7.58%</td>
<td>80.67%</td>
<td>6</td>
<td>7.56%</td>
<td>8.41%</td>
<td>84.03%</td>
<td>4.17%</td>
</tr>
</tbody>
</table>

Table 5: FP and FN and Total ICA, CADD Algorithms Accuracy

<table>
<thead>
<tr>
<th>Set</th>
<th>FP</th>
<th>FN</th>
<th>Accuracy</th>
<th>PC</th>
<th>FP</th>
<th>FN</th>
<th>Accuracy</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10.08%</td>
<td>40.34%</td>
<td>49.58%</td>
<td>25</td>
<td>15.13%</td>
<td>18.48%</td>
<td>66.39%</td>
<td>33.91%</td>
</tr>
<tr>
<td>2</td>
<td>10.08%</td>
<td>40.34%</td>
<td>49.58%</td>
<td>10</td>
<td>12.61%</td>
<td>18.48%</td>
<td>68.91%</td>
<td>38.99%</td>
</tr>
<tr>
<td>3</td>
<td>10.08%</td>
<td>40.34%</td>
<td>49.58%</td>
<td>5</td>
<td>9.24%</td>
<td>6.73%</td>
<td>84.03%</td>
<td>69.48%</td>
</tr>
<tr>
<td>4</td>
<td>10.08%</td>
<td>40.34%</td>
<td>49.58%</td>
<td>5</td>
<td>20.17%</td>
<td>10.08%</td>
<td>69.75%</td>
<td>40.68%</td>
</tr>
<tr>
<td>5</td>
<td>10.08%</td>
<td>40.34%</td>
<td>49.58%</td>
<td>6</td>
<td>7.56%</td>
<td>8.41%</td>
<td>84.03%</td>
<td>69.48%</td>
</tr>
</tbody>
</table>

Table 4 demonstrates that using ICA for feature extraction after the dimensionality reduction step using PCA improves the total algorithm performance in all test sets rather than using PCA algorithm only. The best result for PCA algorithm for all test sets is 80.67% while 84.03% for the proposed CADD algorithm. The
proposed algorithm improved PCA algorithm accuracy with average 8.88% for all test sets.

Table 5 shows the experimental results of ICA algorithm versus the proposed CADD algorithm. The best result of applying ICA algorithm is 49.58%. In contrast, the best result of applying the proposed CADD algorithm is 84.03%. These results indicate that using PCA algorithm for dimensionality reduction before ICA algorithm improves the ICA algorithm accuracy with average 50.51%. The results for ICA algorithm show that fuzzy classifier performance degraded when size of sub-images increase. A fuzzy classifier requires features reduction method in order to minimize total number of membership functions and improves its accuracy. For ICA algorithm, each sub-image has large number of selected features that did not reduced using a dimensionality reduction module like PCA algorithm. Therefore, fuzzy classifier performance degraded in all tested sub-images for ICA algorithm.

Table 6 shows experimental results using the proposed CADD algorithm as a computer aided diagnosis system. The best result is 78% where 15 malignant sub-images out of 25 are correctly classified and 31 benign sub-images out of 34 are correctly classified.
<table>
<thead>
<tr>
<th>Set</th>
<th>Training set</th>
<th>Testing set</th>
<th>Size-pixels</th>
<th>P</th>
<th>C</th>
<th>CADD Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Benign</td>
<td>Malignant</td>
<td>Total</td>
<td>Benign</td>
<td>Malignant</td>
<td>Total</td>
</tr>
<tr>
<td>1</td>
<td>34</td>
<td>26</td>
<td>60</td>
<td>34</td>
<td>25</td>
<td>59</td>
</tr>
<tr>
<td>2</td>
<td>34</td>
<td>26</td>
<td>60</td>
<td>34</td>
<td>25</td>
<td>59</td>
</tr>
</tbody>
</table>

Table 6: Computer Aided-Diagnosis Using CADD Algorithm
7. CONCLUSIONS AND FUTURE DIRECTIONS

A computer aided detection and diagnosis systems have been developed and implemented. The performance of the proposed CADD algorithm is compared against the PCA and ICA algorithms' performance. Extensive simulations using 1400 sub-images are produced. These results indicate that using ICA algorithm for feature selection after the dimensionality reduction step using PCA algorithm improves the total PCA algorithm accuracy about 8.88% for all test sets and the total ICA algorithm accuracy about 50.51%. The best results are obtained with cropped images of size 45x45 pixels. Using ICA algorithm for features selection without using a preprocessing module to reduce sub-images size degrades fuzzy classifier performance. Using a fuzzy classifier following the ICA modeling results are excellent and inline of other reported systems' accuracy. Issues with parameter selection and the robustness of the proposed system are to be an interesting investigation of our future directions. Also, Hough transform will be tested as preprocessing model of the sub-images since it is regarded as a powerful tool for detecting features with various orientations and sizes. Also, a noise reduction module should be investigated to decrease the noise impact on the results. Moreover, the membership functions can be modeled using mean and standard deviation of features values.
APPENDIX

Simulation Results for the Proposed CADD Algorithm Set #5

The following matrices will be used.

V: # of selected principal components for all sub-images.

$\mu_{nt}$: Activation function of normal trained sub-images.

$\mu_{at}$: Activation function of abnormal trained sub-images.

$\mu_{nt}$: Activation function of normal tested sub-images.

$\mu_{as}$: Activation function of abnormal tested sub-images.
Activation Functions

The yellow rectangles represent a false positive rate which indicates a normal sub-image is classified into abnormal one. The red rectangles represent a false negative which means an abnormal sub-image is classified into normal one.

<table>
<thead>
<tr>
<th>Image#</th>
<th>µnt</th>
<th>µat</th>
<th>µns</th>
<th>µas</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>N</td>
<td>5.917</td>
<td>0.231</td>
<td>5.135</td>
<td>1.115</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>1.737</td>
<td>4.494</td>
<td>3.026</td>
<td>3.122</td>
</tr>
<tr>
<td>3</td>
<td>N</td>
<td>5.000</td>
<td>1.224</td>
<td>3.038</td>
<td>3.135</td>
</tr>
<tr>
<td>4</td>
<td>A</td>
<td>0.891</td>
<td>5.378</td>
<td>2.071</td>
<td>4.077</td>
</tr>
<tr>
<td>5</td>
<td>N</td>
<td>5.538</td>
<td>0.641</td>
<td>3.891</td>
<td>2.429</td>
</tr>
<tr>
<td>6</td>
<td>A</td>
<td>1.372</td>
<td>4.917</td>
<td>2.718</td>
<td>3.442</td>
</tr>
<tr>
<td>7</td>
<td>N</td>
<td>3.615</td>
<td>2.558</td>
<td>4.410</td>
<td>1.827</td>
</tr>
<tr>
<td>8</td>
<td>A</td>
<td>1.333</td>
<td>4.872</td>
<td>2.917</td>
<td>3.385</td>
</tr>
<tr>
<td>9</td>
<td>N</td>
<td>4.692</td>
<td>1.410</td>
<td>5.224</td>
<td>1.045</td>
</tr>
<tr>
<td>10</td>
<td>A</td>
<td>2.750</td>
<td>3.423</td>
<td>2.808</td>
<td>3.372</td>
</tr>
<tr>
<td>11</td>
<td>N</td>
<td>5.103</td>
<td>1.019</td>
<td>3.385</td>
<td>2.872</td>
</tr>
<tr>
<td>12</td>
<td>A</td>
<td>1.045</td>
<td>5.224</td>
<td>1.026</td>
<td>5.122</td>
</tr>
<tr>
<td>13</td>
<td>N</td>
<td>5.199</td>
<td>0.955</td>
<td>3.526</td>
<td>2.705</td>
</tr>
<tr>
<td>14</td>
<td>A</td>
<td>1.288</td>
<td>4.987</td>
<td>0.788</td>
<td>5.365</td>
</tr>
<tr>
<td>15</td>
<td>N</td>
<td>5.064</td>
<td>1.141</td>
<td>3.218</td>
<td>3.045</td>
</tr>
<tr>
<td>16</td>
<td>A</td>
<td>1.705</td>
<td>4.590</td>
<td>0.788</td>
<td>5.365</td>
</tr>
<tr>
<td>17</td>
<td>N</td>
<td>3.904</td>
<td>2.218</td>
<td>4.372</td>
<td>1.968</td>
</tr>
<tr>
<td>18</td>
<td>A</td>
<td>1.365</td>
<td>4.910</td>
<td>0.558</td>
<td>5.615</td>
</tr>
<tr>
<td>19</td>
<td>N</td>
<td>5.064</td>
<td>1.135</td>
<td>1.923</td>
<td>4.308</td>
</tr>
<tr>
<td>20</td>
<td>A</td>
<td>0.949</td>
<td>5.288</td>
<td>2.308</td>
<td>3.846</td>
</tr>
<tr>
<td>21</td>
<td>N</td>
<td>5.596</td>
<td>0.551</td>
<td>5.141</td>
<td>1.135</td>
</tr>
<tr>
<td>22</td>
<td>A</td>
<td>1.205</td>
<td>5.064</td>
<td>2.545</td>
<td>3.583</td>
</tr>
<tr>
<td>23</td>
<td>N</td>
<td>5.032</td>
<td>1.109</td>
<td>4.814</td>
<td>1.436</td>
</tr>
<tr>
<td>24</td>
<td>A</td>
<td>1.526</td>
<td>4.744</td>
<td>1.109</td>
<td>5.032</td>
</tr>
<tr>
<td>25</td>
<td>N</td>
<td>5.365</td>
<td>0.788</td>
<td>3.968</td>
<td>2.353</td>
</tr>
<tr>
<td>26</td>
<td>A</td>
<td>1.904</td>
<td>4.340</td>
<td>1.436</td>
<td>4.724</td>
</tr>
<tr>
<td>27</td>
<td>N</td>
<td>5.365</td>
<td>0.788</td>
<td>4.910</td>
<td>1.365</td>
</tr>
<tr>
<td>28</td>
<td>A</td>
<td>3.705</td>
<td>2.333</td>
<td>3.910</td>
<td>2.308</td>
</tr>
<tr>
<td>29</td>
<td>N</td>
<td>2.923</td>
<td>3.276</td>
<td>3.917</td>
<td>2.282</td>
</tr>
<tr>
<td>30</td>
<td>A</td>
<td>0.782</td>
<td>5.442</td>
<td>1.122</td>
<td>5.045</td>
</tr>
<tr>
<td>31</td>
<td>N</td>
<td>4.519</td>
<td>1.558</td>
<td>4.647</td>
<td>1.583</td>
</tr>
<tr>
<td>32</td>
<td>A</td>
<td>1.135</td>
<td>5.141</td>
<td>2.929</td>
<td>3.179</td>
</tr>
<tr>
<td>33</td>
<td>N</td>
<td>2.981</td>
<td>3.173</td>
<td>5.224</td>
<td>1.045</td>
</tr>
<tr>
<td>34</td>
<td>A</td>
<td>0.872</td>
<td>5.372</td>
<td>0.641</td>
<td>5.538</td>
</tr>
<tr>
<td>35</td>
<td>N</td>
<td>3.821</td>
<td>2.295</td>
<td>4.647</td>
<td>1.583</td>
</tr>
<tr>
<td>36</td>
<td>A</td>
<td>1.981</td>
<td>4.256</td>
<td>0.865</td>
<td>5.282</td>
</tr>
<tr>
<td>37</td>
<td>N</td>
<td>3.615</td>
<td>2.558</td>
<td>1.987</td>
<td>4.154</td>
</tr>
<tr>
<td>38</td>
<td>A</td>
<td>1.538</td>
<td>4.873</td>
<td>1.494</td>
<td>4.622</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>39</td>
<td>N</td>
<td>2.833</td>
<td>3.253</td>
<td>3.276</td>
<td>2.891</td>
</tr>
<tr>
<td>40</td>
<td>A</td>
<td>1.436</td>
<td>4.814</td>
<td>0.968</td>
<td>5.212</td>
</tr>
<tr>
<td>41</td>
<td>N</td>
<td>4.821</td>
<td>1.372</td>
<td>3.455</td>
<td>2.782</td>
</tr>
<tr>
<td>42</td>
<td>A</td>
<td>1.103</td>
<td>5.109</td>
<td>2.628</td>
<td>3.526</td>
</tr>
<tr>
<td>43</td>
<td>N</td>
<td>5.615</td>
<td>0.558</td>
<td>5.224</td>
<td>1.058</td>
</tr>
<tr>
<td>44</td>
<td>A</td>
<td>1.942</td>
<td>4.359</td>
<td>1.115</td>
<td>5.051</td>
</tr>
<tr>
<td>45</td>
<td>N</td>
<td>5.615</td>
<td>0.558</td>
<td>3.974</td>
<td>2.179</td>
</tr>
<tr>
<td>46</td>
<td>A</td>
<td>1.199</td>
<td>5.064</td>
<td>5.455</td>
<td>0.808</td>
</tr>
<tr>
<td>47</td>
<td>N</td>
<td>5.615</td>
<td>0.558</td>
<td>3.571</td>
<td>2.756</td>
</tr>
<tr>
<td>48</td>
<td>A</td>
<td>3.558</td>
<td>2.506</td>
<td>0.936</td>
<td>5.179</td>
</tr>
<tr>
<td>49</td>
<td>N</td>
<td>5.615</td>
<td>0.558</td>
<td>5.224</td>
<td>1.058</td>
</tr>
<tr>
<td>50</td>
<td>A</td>
<td>3.423</td>
<td>2.846</td>
<td>2.974</td>
<td>3.218</td>
</tr>
<tr>
<td>51</td>
<td>N</td>
<td>3.910</td>
<td>2.340</td>
<td>3.167</td>
<td>3.135</td>
</tr>
<tr>
<td>52</td>
<td>A</td>
<td>4.397</td>
<td>1.737</td>
<td>0.712</td>
<td>5.449</td>
</tr>
<tr>
<td>53</td>
<td>N</td>
<td>4.167</td>
<td>1.987</td>
<td>4.891</td>
<td>1.353</td>
</tr>
<tr>
<td>54</td>
<td>A</td>
<td>1.282</td>
<td>4.987</td>
<td>3.840</td>
<td>2.365</td>
</tr>
<tr>
<td>55</td>
<td>N</td>
<td>5.032</td>
<td>1.109</td>
<td>5.058</td>
<td>1.199</td>
</tr>
<tr>
<td>56</td>
<td>A</td>
<td>2.429</td>
<td>3.885</td>
<td>2.224</td>
<td>3.923</td>
</tr>
<tr>
<td>57</td>
<td>N</td>
<td>5.295</td>
<td>0.885</td>
<td>2.327</td>
<td>3.897</td>
</tr>
<tr>
<td>58</td>
<td>A</td>
<td>3.782</td>
<td>2.404</td>
<td>2.237</td>
<td>3.936</td>
</tr>
<tr>
<td>59</td>
<td>N</td>
<td>4.897</td>
<td>1.276</td>
<td>4.391</td>
<td>1.795</td>
</tr>
<tr>
<td>60</td>
<td>A</td>
<td>1.929</td>
<td>4.346</td>
<td>3.154</td>
<td>3.141</td>
</tr>
<tr>
<td>61</td>
<td>N</td>
<td>4.872</td>
<td>1.263</td>
<td>3.359</td>
<td>2.846</td>
</tr>
<tr>
<td>62</td>
<td>A</td>
<td>1.385</td>
<td>4.910</td>
<td>2.878</td>
<td>3.295</td>
</tr>
<tr>
<td>63</td>
<td>N</td>
<td>4.660</td>
<td>1.519</td>
<td>2.186</td>
<td>4.045</td>
</tr>
<tr>
<td>64</td>
<td>A</td>
<td>2.167</td>
<td>4.109</td>
<td>3.026</td>
<td>3.154</td>
</tr>
<tr>
<td>65</td>
<td>N</td>
<td>4.167</td>
<td>1.987</td>
<td>5.378</td>
<td>0.891</td>
</tr>
<tr>
<td>66</td>
<td>A</td>
<td>1.135</td>
<td>5.141</td>
<td>2.590</td>
<td>3.679</td>
</tr>
<tr>
<td>67</td>
<td>N</td>
<td>5.045</td>
<td>1.122</td>
<td>4.801</td>
<td>1.423</td>
</tr>
<tr>
<td>68</td>
<td>A</td>
<td>2.436</td>
<td>3.891</td>
<td>4.160</td>
<td>2.058</td>
</tr>
<tr>
<td>69</td>
<td>N</td>
<td>3.385</td>
<td>2.917</td>
<td>4.724</td>
<td>1.506</td>
</tr>
<tr>
<td>70</td>
<td>A</td>
<td>0.981</td>
<td>5.308</td>
<td>1.968</td>
<td>4.135</td>
</tr>
<tr>
<td>71</td>
<td>N</td>
<td>5.615</td>
<td>0.558</td>
<td>4.997</td>
<td>1.218</td>
</tr>
<tr>
<td>72</td>
<td>A</td>
<td>3.308</td>
<td>2.987</td>
<td>3.699</td>
<td>2.564</td>
</tr>
<tr>
<td>73</td>
<td>N</td>
<td>3.891</td>
<td>2.308</td>
<td>1.301</td>
<td>4.923</td>
</tr>
<tr>
<td>74</td>
<td>A</td>
<td>1.256</td>
<td>4.955</td>
<td>3.026</td>
<td>3.122</td>
</tr>
<tr>
<td>75</td>
<td>N</td>
<td>5.212</td>
<td>0.968</td>
<td>4.410</td>
<td>1.827</td>
</tr>
<tr>
<td>76</td>
<td>A</td>
<td>0.795</td>
<td>5.442</td>
<td>2.468</td>
<td>3.673</td>
</tr>
<tr>
<td>77</td>
<td>N</td>
<td>4.846</td>
<td>1.391</td>
<td>4.051</td>
<td>2.276</td>
</tr>
<tr>
<td>78</td>
<td>A</td>
<td>2.000</td>
<td>4.263</td>
<td>5.147</td>
<td>1.141</td>
</tr>
<tr>
<td>79</td>
<td>N</td>
<td>4.308</td>
<td>1.936</td>
<td>2.141</td>
<td>3.981</td>
</tr>
<tr>
<td>80</td>
<td>A</td>
<td>1.212</td>
<td>5.071</td>
<td>3.064</td>
<td>3.212</td>
</tr>
<tr>
<td>81</td>
<td>N</td>
<td>3.282</td>
<td>2.865</td>
<td>3.353</td>
<td>2.795</td>
</tr>
<tr>
<td>82</td>
<td>A</td>
<td>1.122</td>
<td>5.141</td>
<td>2.853</td>
<td>3.269</td>
</tr>
<tr>
<td>83</td>
<td>N</td>
<td>1.981</td>
<td>4.244</td>
<td>0.987</td>
<td>5.231</td>
</tr>
<tr>
<td>84</td>
<td>A</td>
<td>0.872</td>
<td>5.372</td>
<td>4.051</td>
<td>2.276</td>
</tr>
<tr>
<td>85</td>
<td>N</td>
<td>2.103</td>
<td>4.199</td>
<td>4.567</td>
<td>1.590</td>
</tr>
<tr>
<td>86</td>
<td>A</td>
<td>1.122</td>
<td>5.128</td>
<td>2.942</td>
<td>3.179</td>
</tr>
<tr>
<td>87</td>
<td>N</td>
<td>1.596</td>
<td>4.654</td>
<td>5.218</td>
<td>1.051</td>
</tr>
<tr>
<td>88</td>
<td>A</td>
<td>1.519</td>
<td>4.737</td>
<td>0.801</td>
<td>5.365</td>
</tr>
<tr>
<td>89</td>
<td>N</td>
<td>3.199</td>
<td>2.942</td>
<td>3.583</td>
<td>2.769</td>
</tr>
<tr>
<td>90</td>
<td>A</td>
<td>1.821</td>
<td>4.317</td>
<td>2.891</td>
<td>3.359</td>
</tr>
</tbody>
</table>
Activation Functions of Set # 5

The false positive, false negative and algorithm accuracy for set #5 can be computed using the following equations.

False positive rate is computed using the following equation.

\[ F_p = \frac{B}{N} \times 100 \]

\[ = \frac{9}{119} \times 100 = 7.56\% \]

Where B represents number of false positives and N represents total number of tested sub-images.
False negative rate is computed using the following equation.

\[ F_N = \frac{C}{N} \times 100 = \frac{10}{119} \times 100 = 8.4\% \]

Where \( C \) represents number of false negatives and \( N \) represents total number of tested sub-images.

Algorithm accuracy is computed using the following equation.

\[ Acc = \frac{N_r}{N} \times 100 = \frac{100}{119} \times 100 = 84.03\% \]

Where \( N_r \) represents number of correctly classified sub-images.
BIBLIOGRAPHY

Detection and Diagnosis in Mammography,”
eee.sharif.edu/~miap/Files/MammographyReview1.pdf
3. C.H. Lee,” Screening mammography: proven benefit, continued controversy,”
4. C. Smart, R. Hendrick, J. Rutledge, R. Smith,” Benefit of mammography
screening in women ages 40 to 49 years: Current evidence from randomized
medical imaging,” IEEE Transactions on Medical Imaging, vol. 20, pp. 1205-
field for tumor detection in digital mammography,” Medical Imaging, IEEE
Diagnosis in Mammography", Handbook of Image and Video Processing , 2nd
Detection of Microcalcification in Mammograms– A Review," ICGST, GVIP.
11. AJ. Mendez, Pablo G. Tahocesb, Maria J. Lado, Miguel Souto, Jose L. Correa,
Juan J. Vidal, “Automatic Detection of Breast Border and Nipple in Digital
Mammograms”, Computer Methods and Programs in Biomedicine, pp: 253–
262, v.49, 1996.
12. Michael A. Wirth, Alexei Stapinski, “Segmentation of the Breast Region in
Mammograms Using Active Contours”, Dept. of Computing and Information
Science, University of Guelph.
14. Ruey-Feng Chang, Member, IEEE, Wen-Jie Wu, Chih-Chi Tseng, Dar-Ren
Chen, Woo Kyung Moon, “3-D Snake for US in Margin Evaluation for
Malignant Breast Tumor Excision Using Mammothome”, IEEE Transactions
15. Thor Ole Gulsrud and John Hakon Husoy, “Optimal Filter-Based Detection of
Microcalcifications”, IEEE Transactions on Biomedical Engineering, pp:
Asymmetry in Mammograms via Directional Filtering With Gabor Wavelets”,


56. F. Rafiee-Rad, H. Soltanina-Zadeh, M. Rahmati, S. Pour-Abdollah, "Microcalcification classification in mammograms using multiwavelet


