12-2010

Toward Automated Detection and Diagnosis of Mammographic Microcalcifications

Imad Mohammad Zyout
Western Michigan University

Follow this and additional works at: https://scholarworks.wmich.edu/dissertations

Part of the Electrical and Computer Engineering Commons

Recommended Citation
Zyout, Imad Mohammad, "Toward Automated Detection and Diagnosis of Mammographic Microcalcifications" (2010). Dissertations. 638.
https://scholarworks.wmich.edu/dissertations/638

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact maira.bundza@wmich.edu.
TOWARD AUTOMATED DETECTION AND DIAGNOSIS OF MAMMOGRAPHIC MICROCALCIFICATIONS

by

Imad Mohammad Zyout

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Philosophy
Department of Electrical and Computer Engineering
Advisor: Ikhlas Abdel-Qader, Ph.D.

Western Michigan University
Kalamazoo, Michigan
December 2010
TOWARD AUTOMATED DETECTION AND DIAGNOSIS OF MAMMOGRAPHIC MICROCALCIFICATIONS

Imad Mohammad Zyout, Ph.D.
Western Michigan University, 2010

Mammographic diagnosis is the most effective technique to detect breast cancer in its infancy when it is most responsive to treatment. An early and a significant indicator of breast cancer is the presence of clustered microcalcifications (MCs). Mammographic MCs greatly vary in their appearance and shape, and become indistinguishable when surrounded by dense breast tissue. This makes radiologist’s interpretation of mammograms a tedious and an error prone task.

Although computer aided diagnosis (CAD) methods are being developed to aid radiologist in detecting and analyzing the malignancy of MCs, existing systems have not achieved a satisfactory performance. The specificity of existing methods is low compared to a radiologist’s interpretation. Therefore, there is a need for exploring new detection methods and developing automated, robust feature extraction and selection techniques that support the diagnosis process.

To address these needs, a detection framework that employs a pattern-synthesizing process along with statistical and spectral characterization of mammograms is proposed. A trained statistical Bayesian classifier using synthetic MCs will be used to classify anonymous input patterns into a background or microcalcification classes. Morphological image processing is also proposed in this dissertation to segment and
characterize the shape and the distribution of MCs. Automated nested subsets feature
selection method and heuristic search method are investigated via a full model selection
using PSO-SVM framework. Furthermore, a new approach to extract texture features of
MCs using a multiscale Hessian image analysis is developed and tested.

The detection and diagnosis schemes developed in this dissertation are tested
using mammograms from the Mammographic Image Analysis Society (MIAS) database
and compared to other existing methods. The results indicate that the performance of the
detection scheme is adequate while the performance of the shape-based diagnosis of MCs
scheme is superior and very promising.
ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my dissertation advisor Dr. Ikhlas Abdel-Qader, for her guidance, and constant encouragement and support. This dissertation and all research publications have not been possible without her instructive comments and valuable suggestions. Her mentorship was principal in providing me a varied experience that helped me to shape my long-term career goals in both research and instruction.

I would also like to extend my thanks to Dr. Christina Jacobs for serving on my dissertation committee and for her medical consultation of the work in this dissertation. Also, it is a pleasure to thank my committee members, Dr. Massood Atashbar, Dr. Dionysios Kountanis, and Dr. Liang Dong.

I am extremely grateful to my parents, sisters, and brothers, abroad, for their unconditional love and support. Finally and most importantly, I will not forget to express my sincerest appreciation to my wife for her support and patience during the last three years.

Imad Mohammad Zyout
TABLE OF CONTENTS

ACKNOWLEDGMENTS......................................................................................... ii
LIST OF TABLES................................................................................................... viii
LIST OF FIGURES................................................................................................... ix

CHAPTER
I. INTRODUCTION ................................................................................................. 1
   1.1 Motivation and objectives ........................................................................... 5
   1.2 Dissertation organization .......................................................................... 9

II. COMPUTER AIDED ANALYSIS: THEORETICAL BACKGROUND ....... 11
   2.1 Introduction .................................................................................................. 11
   2.2 Image segmentation ..................................................................................... 12
      2.2.1 Wavelet transforms .............................................................................. 13
      2.2.2 Mathematical morphology .................................................................. 17
      2.2.3 Hessian based image analysis ............................................................... 20
   2.3 Feature extraction ......................................................................................... 22
      2.3.1 Shape analysis ...................................................................................... 22
         2.3.1.1 Regional descriptors....................................................................... 23
         2.3.1.2 Boundary descriptors.................................................................... 24
      2.3.2 Texture analysis .................................................................................... 26
         2.3.2.1 Spectral measures ........................................................................ 28
      2.3.3 Non-image features .............................................................................. 28
   2.4 Feature selection .......................................................................................... 29
Table of Contents—continued

CHAPTER

2.4.1 A heuristic search using PSO ........................................... 30
2.4.2 Univariate based feature ranking ........................................ 32
2.4.3 Outweighed univariate based nested subsets method ............... 33
2.4.4 Feature selection using binary PSO .................................... 34
2.5 Binary classification using supervised learning ....................... 35
  2.5.1 Statistical Bayesian learning ....................................... 36
  2.5.2 k-nearest neighbor (kNN) ......................................... 37
  2.5.3 Support vector machine (SVM) .................................... 38
2.6 Performance evaluation .................................................... 41
  2.6.1 Results analysis ...................................................... 41
  2.6.2 Validation methods .................................................. 44

III. COMPUTER AIDED DIAGNOSIS IN MAMMOGRAPHY .................. 45

  3.1 Breast cancer ............................................................. 45
  3.2 Screening of breast cancer ............................................... 46
    3.2.1 Screen film mammography ...................................... 47
    3.2.2 Full field digital mammography ................................ 49
    3.2.3 Adjunctive breast imaging techniques ......................... 50
  3.3 Breast abnormalities .................................................... 51
    3.3.1 Calcifications ...................................................... 51
    3.3.2 Masses ............................................................... 52
    3.3.3 Architectural distortion ......................................... 54
# Table of Contents—continued

## CHAPTER

3.4 Radiologist based interpretation of mammograms .................................. 54

3.5 Computer aided analysis (CAD) of mammogram ..................................... 56

3.5.1 Computer aided detection (CADe) of microcalcifications .......................... 59

3.5.2 Computer aided diagnosis (CADx) of microcalcifications ......................... 64

3.5.3 Summary and conclusions .............................................................. 68

3.6 MIAS mammogram dataset ...................................................................... 70

### IV. BAYESIAN CLASSIFIER WITH SIMPLIFIED LEARNING PHASE FOR DETECTING MICROCALCIFICATIONS IN DIGITAL MAMMOGRAMS ........................................ 74

4.1 Introduction ............................................................................................. 74

4.2 Segmentation using simplified learning Bayesian classifier ....................... 76

4.2.1 Construction of the synthetic microcalcifications .................................. 79

4.2.2 Feature extraction and formation of a pattern vector ............................. 80

4.2.2.1 Highpass filtering using 2-D wavelet transforms ............................... 82

4.2.2.2 Feature extraction using point discontinuity ..................................... 82

4.2.3 Learning phase using synthetic microcalcifications ............................... 84

4.2.4 Parameter estimation of Bayesian classifier ......................................... 86

4.2.5 Segmentation via Bayesian classifier .................................................. 87

4.3 Experimental results ................................................................................ 88

4.3.1 Mammogram test data ........................................................................ 88

4.3.2 Simulation methods and parameter settings ......................................... 88
Table of Contents—continued

CHAPTER

4.3.3 Experimental results analysis .................................................. 89
4.3.4 Performance evaluation and comparison ...................................... 92
4.4 Discussion and conclusions .......................................................... 97

V. MORPHOLOGY BASED DIAGNOSIS OF CLUSTERED MICRO-CALCIFICATIONS USING PSO-SVM FULL MODEL SELECTION........................................ 98

5.1 Introduction ..................................................................................... 98
5.2 Morphology based CADx of microcalcifications ................................. 101

5.2.1 Morphological based segmentation .............................................. 102

5.2.1.1 Segmentation of MCs using morphological filtering ................. 103
5.2.1.1 Segmentation of MC cluster .................................................. 104

5.2.2 Shape based feature extraction .................................................... 105

5.2.2.1 Region descriptors ................................................................ 106
5.2.2.2 Boundary descriptors .......................................................... 107

5.2.3 Feature selection methods ............................................................ 108

5.2.3.1 Outweighed univariate based nested subsets method .............. 109
5.2.3.2 Feature selection using binary PSO method ............................. 111

5.2.4 PSO-SVM full model selection .................................................... 112

5.3 Experimental results ....................................................................... 114

5.3.1 Mammographic test data .............................................................. 114
5.3.2 Experimental setup ...................................................................... 114
Table of Contents—continued

CHAPTER

5.3.3  Impact of MC segmentation ........................................ 115
5.3.4  Results on feature selection ........................................ 116
5.3.5  Results on classifier model selection .............................. 120
5.3.6  Impact of a human based interpretation ........................... 123
5.3.7  Comparison with other CADx ......................................... 128
5.4  Discussion and conclusions ............................................. 128

VI.  CHARACTERIZATION OF CLUSTERED MICROCALCIFICATIONS USING MULTISCALE HESSIAN BASED FILTERING .......... 130

6.1  Introduction .............................................................. 130
6.2  Analysis using image Hessian ......................................... 131
6.3  Diagnosis of MCs using Hessian based feature extraction ........ 133
6.4  Experimental results .................................................... 135
6.4.1  Mammogram test data ............................................... 135
6.4.2  Parameters and feature selection ................................... 136
6.4.3  Performance evaluation .............................................. 139
6.5  Discussion and conclusions ............................................. 142

VII.  CONCLUSIONS AND FUTURE WORK .................................. 144

7.1  Summary and conclusions .............................................. 144
7.2  Dissertation contributions ............................................. 147
7.3  Future work ............................................................. 148
# LIST OF TABLES

4.1 Detection results using self-learning BC .................................................. 95
4.2 Comparison with related work ................................................................. 96
5.1 Extracted shape features ............................................................... 108
5.2 Univariate feature ranking using ROC analysis ..................................... 110
5.3 Full model selection using univariate nested subsets and PSO-SVM methods .......................................................... 118
5.4 Full model selection using binary PSO-SVM method ................................... 119
6.1 Analysis of Hessian eigenvalue $\lambda_1$ and $\lambda_2$ ..................................... 131
6.2 Univariate feature ranking using F-score method .................................. 138
6.3 Performance evaluation of Hessian based features ..................................... 140
LIST OF FIGURES

2.1 One-dimensional discrete wavelet transform ........................................... 15
2.2 One level two-dimensional discrete wavelet analysis ................................. 16
2.3 One level two-dimensional discrete wavelet reconstruction ....................... 17
2.4 The confusion matrix ................................................................................. 42
2.5 ROC curve examples .................................................................................. 43
3.1 Illustration of mediolateral oblique (MLO) and craniocaudal (CC) projections a female breast, which are acquired during a screen mammography exam ................................................................. 48
3.2 Craniocaudal (CC) and mediolateral oblique (MLO) view mammograms of a right female breast ...................................................................................... 49
3.3 Examples of most common breast abnormalities ......................................... 53
3.4 Illustration of a typical computer aided diagnosis system, which consists of computer aided detection and computer aided diagnosis subsystems .......... 58
3.5 Malignant microcalcification clusters from the MIAS database .................. 71
3.6 Benign microcalcification clusters from the MIAS database ....................... 72
3.7 Additional eight microcalcification clusters extracted from three MIAS mammograms with calcifications spread over the entire breast region ............. 73
4.1 Segmentation using Bayesian learning ......................................................... 77
4.2 Composition of a pattern vector x using a four-image stack ..................... 78
4.3 Construction of synthetic MCs ..................................................................... 81
4.4 Synthesizing clustered microcalcifications ............................................... 81
4.5 Highpass filtering using wavelet transform ............................................... 83
4.6 Microcalcifications’ point singularity analysis ........................................... 83
4.7 Feature extraction of microcalcifications ..................................................... 84
List of Figures—continued

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.8</td>
<td>Improving the FP rate of the detection results using Otsu’s thresholding</td>
<td>91</td>
</tr>
<tr>
<td>4.9</td>
<td>Detection results of MCs in a dense mammogram using different modeling constant K</td>
<td>92</td>
</tr>
<tr>
<td>4.10</td>
<td>Detection results of MCs in fatty mammogram using different K</td>
<td>93</td>
</tr>
<tr>
<td>4.11</td>
<td>Examples of the detection results using different K values</td>
<td>94</td>
</tr>
<tr>
<td>5.1</td>
<td>Diagnosis of MCs using shape based CADx, which uses SVM classifier generalization error estimated using leave-one-out (LOO) cross-validation method as an objective function for heuristic embedded feature selection using PSO method</td>
<td>102</td>
</tr>
<tr>
<td>5.2</td>
<td>Segmentation of MCs using a dual top-hat filtering scheme</td>
<td>104</td>
</tr>
<tr>
<td>5.3</td>
<td>Segmentation of MC cluster margin</td>
<td>105</td>
</tr>
<tr>
<td>5.4</td>
<td>Impact of the segmentation of the MCs on the classification performance</td>
<td>116</td>
</tr>
<tr>
<td>5.5</td>
<td>Classification performance of feature subsets constructed using conventional ( u=0 ) and outweighed ( u \geq 0 ) nested subsets</td>
<td>117</td>
</tr>
<tr>
<td>5.6</td>
<td>Effect of the selection of the kernel function on the classification performance using SVM</td>
<td>121</td>
</tr>
<tr>
<td>5.7</td>
<td>Impact of the model selection on the classifier generalization performance with feature subsets obtained using modified nested subsets method</td>
<td>124</td>
</tr>
<tr>
<td>5.8</td>
<td>Impact of the model selection on the classifier generalization performance with feature subsets obtained using a heuristic binary PSO method</td>
<td>125</td>
</tr>
<tr>
<td>5.9</td>
<td>Examples of the manual and the automatic delineation of the margin of microcalcification cluster</td>
<td>127</td>
</tr>
<tr>
<td>6.1</td>
<td>Filtering of mammographic microcalcifications using a Hessian based filter</td>
<td>133</td>
</tr>
<tr>
<td>6.2</td>
<td>Classification of MC cluster using Hessian image analysis and a kNN classifier</td>
<td>135</td>
</tr>
<tr>
<td>6.3</td>
<td>The discriminative power of the energy and entropy features</td>
<td>138</td>
</tr>
<tr>
<td>6.4</td>
<td>ROC analysis of the extracted features</td>
<td>141</td>
</tr>
</tbody>
</table>
List of Figures—continued

6.5 Effect of the structureness of the Hessian filter .................................................. 141
CHAPTER I

INTRODUCTION

Breast cancer disease is a tremendous threat and second most deadly cancer for women in the U.S. [1]-[4], which also continues to be a leading cause of death and a significant health problem in Europe [5], Australia [6], and Asia [7]-[8]. In the United States, the American Cancer Society (ACS) estimated that 1 in 8 women is at risk of developing a breast cancer in her life-time, and 192,370 new cases are estimated to be diagnosed with an invasive breast cancer. Also, 40,170 women are expected to die in 2009 [9]. What causes breast cancer remains unexplained and early diagnosis and treatment of the disease is the only feasible medical procedure to minimize breast cancer deaths. Mammography, a low dose x-rays based imaging technology of the breast, is yet the most sensitive, reliable, and low cost imaging technology. Although some limitations of this technology exist when dense breast parenchyma present, screen mammography has been proven to be the most effective tool for early detection of breast cancer, which strongly improves the chances of treatment and the survival rates [1]-[4].

When radiologists interpret screen mammograms, they usually seek the detection and diagnosis of a group of mammographic abnormalities such as clustered microcalcifications (MCs), circumscribed and speculated masses, architectural distortion, asymmetry between left and right breasts, breast edema, and Lymphadenopathy [7]. Clustered microcalcifications, microscopic deposits of calcium that show as bright spots on a mammogram, are the most important breast abnormality to be interpreted because their
occurrence represents a significant indicator of an early stage of breast cancer [1]. MCs also appear more frequently than other breast abnormalities in that they show on 30% -50% of screened mammograms [1], [3].

Interpretation of mammograms by radiologists tends to be a difficult, tedious, subjective, and error prone task. This causes 10% to 30% of all cancers to be missed by radiologists. Furthermore, only 30% of marked cancers are turn out to be truly malignant lesions after breast biopsies [10]. This low positive prediction value (PPV) is mostly because of the subtle nature of the mammographic abnormalities at early stages of breast cancer that increases the possibility of overlooking and misclassifying even if the diagnosis process is accomplished by an expert radiologist. Other causes of this low PPV are the poor image quality of screen mammography and the low specificity of mammography in distinguishing between malignant and benign abnormalities.

Aiming at addressing some limitations of mammography and its interpretation, computer aided diagnosis (CAD) technology is being developed over the last two decades to radiologists with a second opinion that might help in interpreting a digital mammogram.

Considering the main function of different CAD systems, CAD approaches are divided into two types: computer aided detection (CADe), which is intended to help radiologist in localizing abnormal lesions in mammogram and to reduce the number of missed cancers, and a computer aided diagnosis (CADx) system, which is mainly designed to help radiologists in discriminating benign from malignant abnormalities.

Investigators have developed several CADe systems. Among these systems, CADe systems based on wavelet decomposition, feature extraction using statistical modeling, and supervised learning machines as pattern classifiers are the most effective [4]. Over the last 20 years, CADe technology has achieved a noticeable success and several CADe methods
have received FDA approvals, are commercially implemented, and are currently in clinical use in USA. According to some recent studies, CADe can achieve an average increase of 10% of the number of detected cancers [2],[4]. However, the low specificity (i.e. many false alarms compared to a radiologist) is a major shortcoming of current CAD technology that impairs the confidence in the positive role of CADe [11].

Both healthy breast tissue and abnormal breast lesions have non-stationary natures, which make the efficiency of machine-learning based detection dramatically sensitive to the nature and size of training samples used to estimate the learning model. This fuzzy nature of the breast tissue also represents a great challenge for detecting MCs using a simple template matching. In addition, it requires a huge computational complexity and large volume of templates to detect fuzzy patterns such as MCs. Another limitation of template matching based detection is the poor robustness to the presence of noisy and distorted real MCs [1]. Hence, this dissertation attempts to overcome the limitations experienced by simple template matching and the state of the art detection using a supervised learning by using synthetic patterns of MCs along with real patterns of normal breast tissue to train a statistical Bayesian classifier. This proposed self-learning process adapts the learning model using synthetic training patterns constructed form each input mammogram, which is expected to provide acceptable detection sensitivity of MCs even when surrounding breast tissue is dense. However, the drawback of this proposed approach are: Bayesian learning need to be repeated for each mammogram and sometimes patterns of healthy breast tissue might be corrupted by anonymous real MCs.

Automated diagnosis of mammograms can significantly improve the sensitivity of interpreting mammograms because about 50% of all cancers are missed because of misclassification rather than overlook [10]. Unlike automated detection of mammographic
abnormalities that can be accomplished using different image processing methods including machine based learning methods, automated diagnosis of mammographic lesions usually is modeled as a binary classification problem solved by using a supervised learning scheme. Investigators have developed numerous computerized algorithms for diagnosis of MCs. In such methods, texture and shape based feature extraction methods have been used to characterize mammographic MCs, important features are selected using automated and semi-automated methods, and various supervised learning machines were used to classify lesions into malignant and benign classes.

Although some observer studies have demonstrated the effectiveness of CADx, no single CADx scheme is clinically in use. This situation is mostly because CADx technology, if approved, can play a significant role in the process of the medical diagnosis. Therefore, an adopted CADx scheme must achieve a satisfactory and an optimal performance. Existing CADx methods have not achieved this perfect performance yet. Some studies achieved 100% true-positive rates by sacrificing the false positive rates. Research demands of CADx technology include performing large-scale observer studies to validate the positive role of CADx and finding solutions to several unsolved problems and limitations such as robust feature extraction scheme, automatic feature selection method, and developing and testing new supervised learning machines.

In this dissertation, I decided to address some limitations of existing CADx technology by developing different algorithms for automatic region selection and microcalcification segmentation, shape and texture based feature extraction. I also have developed a heuristic framework using PSO-SVM for accomplishing a full model selection of CADx system to optimize the performance of SVM based classification. Moreover, this
PSO-SVM framework represents a unified scheme for accomplishing feature selection and classifier parameter selection.

1.1 Motivation and objectives

Shape manifestation of mammographic MCs is the key method commonly used by radiologists to detect and non-invasively diagnose of breast cancer. Limitations of mammography: the subtle nature of MCs, radiologist’s experience, and inter-and intra observer variation of human based diagnosis lead to a low positive predictive value (PPV) of mammographic interpretations. To improve the analysis of mammography, CAD technology, including both CADe and CADx systems, is being developed. Although some of these systems have shown to improve the detection and diagnosis of breast cancer, the high number of false breast biopsies is the main drawback.

Microcalcification detection task can be modeled as a two-class pattern recognition problem. Hence, learning machines can play a significant role in the success of CADe technology. Supervised learning with a good generalization performance requires an efficient learning process that involves effective pattern representation and estimation of the learning model. Supervised learning of non-stationary and fuzzy patterns such as mammographic MCs and normal breast tissue requires as many training examples as possible, which are extracted and labeled manually. This situation might be impractical and might lead to a high computational complexity of the process. Over-fitting and poor generalization of most learning machines is possible because the learning model of these methods is commonly estimated using small number of training patterns, which mostly covers a limited spectrum of the fuzzy mammographic patterns. In other words, a mammogram, a 2D image formed as
results of the attenuation of x-rays beams passing a human breast, is significantly affected by the internal structure of the breast that might be affected by other attributes of the body like weight, age, and ethnicity. Moreover, the description of mammographic MCs becomes more difficult when MCs are surrounded by dense breast mammogram tissue.

Considering these challenges, I have raised the following question: Could synthetic and adaptive patterns of MCs lead to an efficient machine learning based approach for segmentation and detection of clustered MCs? This research question has been answered by developing a semi-supervised CADe scheme where the breast region is manually extracted for segmentation and detection of MC clusters. The proposed scheme first attempts to model the real MCs (unknown location) in a given region and creates synthetic MCs patterns. The distributions of both MCs patterns and the background breast tissue can be modeled as two different Gaussian distributions, which might justify the use of optimal Bayesian classifier, for distinguishing between MCs and healthy breast tissue. In this approach, the Bayesian classifier is trained using synthetic patterns of MCs and real patterns of background breast tissue.

Although the problem of developing a computer aided diagnosis (CADx) for mammographic MCs is being investigated for more than two decades, which is even earlier than CADe. CADx area of research is demanding and no single CADx scheme has been approved for clinical use. This situation is mostly because diagnosis of MCs is more challenging and has a more serious impact on the medical decision process than the detection task. Moreover, the number of published studies that addressed the problem of computer-aided diagnosis is relatively less than those devoted to detect MCs.

CADx technology, meanwhile, involves several problems and demands which need further investigation such as the developments of robust feature extraction approach,
automatic feature selection, and pattern recognition using supervised learning machines with good generalization ability. This dissertation aims to address these research problems as follows:

- Morphological and shape analysis of MCs plays a significant role in distinguishing between malignant and benign cases. This shape analysis requires an efficient segmentation method that preserves the shape of MCs while attaining high specificity of the segmentation process. In this dissertation, I employ morphological image processing for accomplishing multiscale image filtering and segmentation and for characterizing the shape of MCs and their cluster.

- The main objective of CADx is to help radiologists in discriminating between malignant and benign MC clusters. Hence, incorporating human based interpretations and utilizing any available clinical data (patient age, family history) is necessary to improve the final diagnosis performance of a CADx scheme. In this dissertation, I utilize the ground truth of the MC cluster, which describes the location and size of each MC cluster that is provided as image annotations, to improve MC segmentation and shape feature extraction.

- Differentiating between malignant and benign MCs is usually modeled as a two-class pattern recognition problem, which is solved via different supervised learning machines. An important aspect of an adopted classification scheme is the generalization performance that ensures a good classification performance on the test dataset that is relatively very large compared to a training set. Support vector machine (SVM) is demonstrated to outperform other popular learning machines such as artificial neural networks (ANN) and a k-nearest neighbor (kNN) in many
applications. However, for the diagnosis of MCs, an SVM approach is still less popular than ANN and kNN classifiers.

- Model selection for the SVM classifier is a necessary step to improve the generalization performance, which involves selecting both hyper-parameters (feature selection, kernel function) and parameters (regularization constant and kernel's parameter). A simple grid search for accomplishing classifier model selection in a real parameter space is not practical because in most cases the region of feasibility is not predetermined. In this dissertation, I adopt a heuristic search using particle swarm optimization (PSO) method that is less complex and more efficient than the grid-search and other heuristic methods such as genetic algorithms (GAs).

- Feature selection is a necessary step when the extracted features are redundant, irrelevant, noisy, and a large number of features is used for pattern recognition. Optimal feature selection using an exhaustive search method is mostly simple but not feasible when the feature space is large, which tends to over-fit the input data. Hence, sup-optimal and more efficient feature selection techniques using heuristic methods are used. This dissertation develops a PSO-SVM framework for accomplishing an embedded feature selection that is a unified scheme for selecting the best feature subset and for optimizing the classification performance of the SVM classifier.

- Shape analysis of MCs is very sensitive to a prior segmentation step. This segmentation becomes difficult when the image quality of a mammogram is poor or when dense breast tissue surrounds MCs. Texture analysis of mammographic regions depicting MCs is a popular alternative, which does not require a prior
segmentation of MCs. This work characterizes MC clusters using spectral features that are extracted from a multiscale image Hessian.

1.2 Dissertation organization

The remaining chapters of this dissertation are described as follows:

In Chapter II, I present some theoretical background on image processing, pattern description and classifications. Section 2.2 discusses different image filtering techniques will used thought this work. Section 2.3 describes shape and texture techniques usually used for describing image pattern. In section 2.4, I present the concept of feature selection while focusing on a single variable feature ranking, nested subsets methods, and heuristic methods. Binary classification using supervised learning machines, the performance evaluation and the result analysis will be presented in Section 2.5.

In Chapter III, I briefly introduce breast cancer, the common techniques for breast cancer screening, breast abnormalities, and the research efforts are being developed to computerize the detection and diagnosis of mammographic microcalcification. Breast abnormalities that are commonly shown on a screen mammogram are described in Section 3.3. In Section 3.4, a brief description of the process of the diagnosis of screen mammography is presented. Section 3.5 provides a survey on CAD algorithms that have been developed for the detection and the diagnosis of mammographic MCs. The status of the computer aided diagnosis in mammography including the limitation and future research demands is briefly addressed in Section 3.5. The mammogram dataset will be examined in this work is described in Section 3.6.
In Chapter IV, I develop a new framework for segmentation and detection of mammographic MCs. The proposed CADe scheme first constructs simulated patterns of MCs. Then, four features (a gray-level or intensity value, a measure of the local statistics, a measure of point’s singularity, and a spectral feature using wavelet based filtering) are used to describe the texture of each pixel. This approach employs synthetic patterns of MCs to simplify the learning phase of the statistical Bayesian classifier. The estimated learning model of Bayesian classifier will be used to classify anonymous pixels into a background or target (MCs) pixels.

In Chapter V, I present a four-stage shape based CADx system in which microcalcifications are segmented using a new multiscale morphological filter followed by extracting 44 shape descriptors. In this work, I also developed a PSO-SVM full model selection to optimize the generalization performance of the SVM classifier by selecting the best feature subset and the classifier learning model. Furthermore, the proposed PSO-SVM framework employs and examines two different methods for accomplishing feature search process including univariate based nested subset methods and a heuristic approach using a binary PSO method.

In Chapter VI, I develop a new method for extracting texture features of MCs. In this method, a multiscale image Hessian is constructed and used to characterize the texture of a given mammographic region by computing a set of spectral measures including normalized energy and entropy. Extracted texture features are first evaluated individually using a Fisher-score criterion. Selected features are used with a k-nearest neighbor classifier to distinguish between benign and malignant MCs.

In Chapter VII, a summary of the proposed methods, contributions, and plans for extending methods of this dissertation are presented.
CHAPTER II

COMPUTER AIDED ANALYSIS: THEORETICAL BACKGROUND

2.1 Introduction

A computer aided analysis scheme in general and a computer aided diagnosis system intended to computerize the analysis of mammograms in particular, is a multidisciplinary system. Mammographic CAD systems integrates outcomes from clinical examinations such as screen mammography and radiologists’ interpretations, image formation mechanisms, digital signal processing, computer vision, pattern representation and analysis, and pattern recognition and artificial intelligence. Hence, one can conclude that CAD system of mammogram is a result of collaborative efforts of the state-of-the art technologies and developments in medicine, engineering, mathematics, statistics, computer science, and data mining.

This chapter presents some concepts and background materials on the image processing and artificial intelligence methods will be used throughout this thesis. Section 2.2 addresses different image segmentation techniques. In section 2.3, methods for pattern analysis and feature extraction techniques are discussed, embedded feature selection using heuristic PSO is presented in Section 2.4. Section 2.5 covers different supervised learning machines for solving two-class pattern recognition problem.
2.2 Image segmentation

Segmentation of a digital image (e.g. digital mammogram) is the process of dividing a given image into distinct objects or regions sharing common patterns such as gray-level intensity, morphology, local statistics, and spectral features [12]. Segmentation process is usually designed to achieve the following: 1) producing high specificity of the detected targets by attaining a high number of the desired (true positive) targets and minimum false positive results 2) shape preserving and accurate delineation of the segmented targets.

An early segmentation approach is based on a region growing method. A major limitation of this technique when used for detection is the need for predetermined seeds points. However, region growing based segmentation is still applicable when the purpose of segmentation is a pattern analysis rather than a detection task. While the region approach naturally produces a binary representation of the grown regions, a thresholding process is usually required to accomplish the segmentation task and to produce a binary representation of an image.

A simple thresholding method is based on using a single and a global gray-level or intensity value to mark all image pixels as a target or background class if a given pixel has a gray-level value more or less than a threshold value. A popular example of using a global thresholding method to segment gray-level images is Otsu’s method [13]. Otsu’s thresholding method implicitly assumes that a given digital image can be divided into two classes (i.e. foreground and background), and the method uses the gray-level histogram of a given image to find a single gray-level that minimizes the within-class variance or maximizes the between-class variance.
Sometimes, using a global threshold for segmenting an image with inhomogeneous texture (background intensity or even that of the desired regions) is usually not sufficient and leads to poor segmentation outcomes. One approach to solve this problem is based on using an adaptive thresholding scheme that divides an input image into sub-images, and then a single threshold is used to segment each sub-image.

Image segmentation via filtering and thresholding is a very popular approach that does not require seed points of the objects to be segmented, which also improves the segmentation using global threshold. Commonly, this segmentation scheme is modeled as a two-step process. In the first step, the desired targets are made more distinguishable from a background region via enhancement, filtering, and background suppression methods. In the second step, an adaptive or a global threshold is used to produce a binary representation of the filtered image. Examples of image filtering are spatial filtering using Laplacian of Gaussian and difference of Gaussian, morphological operators, wavelet-based methods, and image Hessian analysis. When binary representation is not required for image analysis, one can drop a thresholding stage and use a filtered image for further pattern analysis and feature extraction.

In the coming subsections, we will provide some theoretical background of discrete wavelet transform, morphological image processing, and Hessian based image analysis, which will be used throughout this work.

2.2.1 Wavelet transforms

Over the last decade, wavelet theory and its multiresolution analysis (MRA) ability [14] have been recognized as the most powerful tools in signal processing. Unlike Fourier
analysis, multiresolution representation of wavelet transform provides spatial–frequency localization which enables the analysis of both local and global features of the processed signal. Wavelet transforms are a set of basis functions derived by translation and dilation of a single function, the mother wavelet, $\psi$ which has the general form of

$$\psi_{a,b}(x) = \frac{1}{\sqrt{a}} \psi \left( \frac{x - b}{a} \right)$$ \hspace{1cm} (2.1)

Equation (5) shows that the frequency and spatial resolution of the wavelet function $\psi_{a,b}$ are functions of the translation and dilation parameters $b$ and $a$ respectively.

A special case of equation (1) is obtained when translation and dilation parameters are integers with a scaling parameter $a$ as an integer of base 2, resulting in the dyadic wavelet transforms and leading to the construction of orthonormal wavelet basis $\psi_{j,k}$:

$$\psi_{j,k}(x) = 2^{-j/2} \psi(2^{-j} x - k)$$ \hspace{1cm} (2.2)

Moreover, MRA using wavelet transform is based on the existence of two unique functions called wavelet and scaling functions. The scaling function is defined as

$$\varphi_{j,k}(x) = 2^{-j/2} \varphi(2^{-j} x - k)$$ \hspace{1cm} (2.3)

where the wavelet function is defined as given in equation (2.2).

An efficient algorithm for computing discrete wavelet transform of a given discrete signal is introduced in [14] by which, each stage of wavelet decomposition process involves extracting an approximate (low pass) version and a detail (highpass) version of the signal. This can be easily implemented using a set of finite impulse response (FIR) filter banks followed by sub-sampling as shown Figure 2.1.a.
Figure 2.1: One-dimensional discrete wavelet transform, (a) and (b) are decomposition and reconstruction using analysis and synthetic filter banks, respectively.

The wavelet synthesis process as shown in Figure 2.1.b is accomplished by first filtering the up-sampled \( c \) and \( d \) using the synthesis lowpass \( \tilde{h} \) and highpass \( \tilde{g} \) filters, respectively. Then, given that the set of analysis and synthesis filters satisfying perfect reconstruction conditions, an original signal \( \tilde{C}_{j+1} \) is obtained by adding the output of each filter, \( \tilde{h} \) and \( \tilde{g} \) [14].

Wavelet transforms are one dimensional in nature but easily extended to analyze 2-D discrete signals or digital images. Separable two-dimensional wavelet transform of an image is constructed by applying 1-D wavelet transform along the image rows and columns as illustrated by Figure 2.2. The 2-D wavelet and scaling functions derived from 1-D wavelet \( \psi(x) \) and scaling \( \phi(x) \) functions, are expressed as [12]

\[
\begin{align*}
\phi(x, y) &= \phi(x)\phi(y) \\
\psi^H(x, y) &= \phi(x)\psi(y) \\
\psi^V(x, y) &= \psi(x)\phi(y) \\
\psi^D(x, y) &= \psi(x)\psi(y)
\end{align*}
\]  

(2.4)
where $\phi(x,y)$ represents a 2-D separable lowpass filter applied along the horizontal and vertical directions. $\psi^H(x,y)$, $\psi^V(x,y)$, $\psi^D(x,y)$ are 2-D separable highpass filters extracting the signal details along the horizontal, vertical, and diagonal directions, respectively.

Similar to a one-dimensional discrete wavelet transform, a two-dimensional discrete wavelet transforms is implemented using a set of FIR filters followed by down-sampling as shown in Figure 2.2. The result of applying this process to decompose a discrete image $f_{j+1}(x,y)$, as shown in Figure 2.2, is a set of four sub-images: one approximate or coarser version $f_j(x,y)$, and three detail subbands $f^H_j(x,y)$, $f^V_j(x,y)$, and $f^D_j(x,y)$. Again, by extending the reconstruction process of the one-dimensional discrete wavelet transforms, the reconstruction process of decomposed image $f_{j+1}(x,y)$ is accomplished as shown in Figure 2.3.

Figure 2.2: One level two-dimensional discrete wavelet analysis.
Variety of wavelet transforms have been proposed and used in the literature [12], [14]-[15] in many applications. These transforms have different features such as regularity, number of vanishing moments, orthogonality, symmetry, and compact support. However, the selection of a wavelet transform with certain features is an application dependant.

![Diagram](image)

Figure 2.3: One level two-dimensional discrete wavelet reconstruction.

2.2.2 Mathematical morphology

Mathematical morphology, a set-theory technique, is a very powerful tool commonly used for shape description and preservation in digital images [12],[16]. This method is mainly based on using a binary kernel, called a structuring element (SE) of an arbitrary shape and size, which is passed over a digital image in similar manner to a windowing method, to describe geometrical structures may exist.
The basic operations of mathematical morphology are the dilation and erosion operations, which have an effect that is equivalent to a region’s expansion and shrinking, respectively. Before proceeding with the details of morphological operation, some basics concepts from set theory will be introduced.

Let $B$ be a set in $\mathbb{Z}^2$. Then, any element $b \in B$ is expressed as $(b_1, b_2)$.

A translation of a set $B$ by a point $z = (z_1, z_2)$ is given by:

$$(B)_z = \{b + z \mid b \in B\}$$  \hspace{1cm} (2.5)

The reflection of set $B$, denoted as $\hat{B}$, is defined as

$$\hat{B} = \{-b \mid b \in B\}$$  \hspace{1cm} (2.6)

The complement of the set $B$ is expressed as

$$B^c = \{x \in \mathbb{Z}^2 \mid x \notin B\}$$  \hspace{1cm} (2.7)

The fundamental operations of mathematical morphology are the dilation and erosion operations of binary images.

For sets $B$ and $A$ in $\mathbb{Z}^2$, the erosion of $B$ using a structuring element $A$, denoted $B \Theta A$, is the removal of all boundary pixels of $B$ as

$$B_E = \{(x, y) \mid A \subseteq B\}$$  \hspace{1cm} (2.8)

The dilation operation of $B$ using the same structuring element $A$, denoted $B \oplus A$, is the dual operation of the erosion process, which expands the boundary of an object as follows

$$B_D = \{(x, y) \mid A \cap B \neq \phi\}$$  \hspace{1cm} (2.9)

Using these erosion and dilation or shrinking and expansion operations, we can define two morphological operations: opening and closing.
Morphological opening tends to eliminate objects smaller than a structuring element, removes image peaks, and breaks narrow connectivity between objects. Opening of a set or binary image \( B \), denoted \( B \circ SE \), is usually implemented by using same structuring element to perform a successive erosion and dilation operations of \( B \), which can be expressed as
\[
B_{\text{Opening}} = (B \Theta A) \oplus A
\]  
(2.10)

Morphological closing operation can fill existing valleys in the processed image, smooth contours, and close small holes. Which can viewed as a dual operation of an opening process that is a morphological dilation followed by erosion operations using same structuring element \( SE \). Using morphological dilation and erosion operations, the closing of a set \( B \), referred to as \( B \bullet A \), is written as
\[
B_{\text{Closing}} = (B \oplus A) \Theta A
\]  
(2.11)

To this point, morphological processing is applied to binary images or sets. The extension to gray-level images is done by a considering a gray-level image \( I \) as digital function \( I(x, y) \), and a structuring element \( SE \) represented by a \( A(x, y) \) that is also a sub-image function. Morphological dilation and erosion of gray-level image \( I \) using is computed as [12]
\[
I \oplus A(a, b) = \max \{ f(a + (s - x), b + (t - y)) \mid (s, t) \in D_f; (x, y) \in D_a \}
\]  
(2.12)

Similarly, a gray-level erosion of an image \( I \) is defined as
\[
I \ominus A(a, b) = \min \{ f(a - (s - x), b - (t - y)) \mid (s, t) \in D_f; (x, y) \in D_a \}
\]  
(2.13)

Similarly, these gray-level dilation and erosion operation can be used to construct opening and closing operations. A very popular application of gray-level morphology is the peaks detection using a top-hat morphological operator [16]. A top-hat operator is usually implemented by first removing the peaks from a given gray-level image via a gray-level
morphological opening process. Then, the output of the opening step is subtracted from the original image to obtain a new image with suppressed background and enhanced peaks. Since a gray-level morphological image opening process can be implemented by successive gray-level erosion and dilation operations, a top-hat filtering of a gray-level image can be expressed as follows

\[ I_{peaks} = I - (I \ominus SE) \oplus SE \]  

(2.14)

2.2.3 Hessian based image analysis

A common approach to characterize the local behavior of an image point \((x_0, y_0)\) is by using a second order Taylor series expansion of the function \(f(x, y)\) [17] as follows

\[ f(x_0 + \Delta x, y_0 + \Delta y) \approx f(x_0, y_0) + [\Delta x \Delta y]^T \nabla + [\Delta x \Delta y]^T H \Delta x \Delta y \]  

(2.15)

where \(\nabla\) and \(H\) are the gradient vector and Hessian matrix of image pixel \(f(x_0, y_0)\).

Computing an image's derivatives (gradient and Hessian) is a noise sensitive process. It is also necessary to ensure that image structures of various sizes will produce a strong response. These challenges can be addressed by computing a multiscale derivative that is less sensitive to the presence of the noise. Practically, this can be implemented by convoluting an input image with a derivative of a Gaussian kernel of certain scale.

According to the scale space theory [18], one can implement a linear, a scale, and a rotation invariant process using the principle of a normalized derivative as follows

\[ f_x = \sigma^r f(x, y) * G_{x, \sigma} \]

\[ f_y = \sigma^r f(x, y) * G_{y, \sigma} \]  

(2.16)
where $G_{y,\sigma}$ and $G_{x,\sigma}$ are directional derivatives of a standard Gaussian kernel of size $\sigma$, denoted as $G(x, y, \sigma)$, along $x$ and $y$ directions, respectively. $\tau$ is a normalization constant proven to be $\frac{1}{4}$ for a fair combination of derivatives at different scale[19]. In the case of a single scale analysis, $\tau$ can be set to one.

To illustrate image filtering using Hessian, we can assume that a 2-D typical nodular object $f(x, y)$ of size $\sigma_0$, denoted as $f(x, y, \sigma_0)$, is modeled as

$$f(x, y) = F_0 e^{-\frac{x^2 + y^2}{2\sigma_0^2}}$$  \hfill (2.17)

where $F_0$ is the brightness at the origin (structure midpoint), and $\sigma_0$ is the size of the structure in pixels. Convoluting a model of nodular object described in equation (2.17), with a second derivative of Gaussian at scale $\sigma$ is equivalent to computing a second derivative of another nodular object $f(x, y, \sigma_0')$ of size $\sigma_0' = \sqrt{\sigma^2 + \sigma_0^2}$. Then, second derivatives $f_{xx}$, $f_{yy}$, and $f_{xy}$ of the model $f(x, y, \sigma_0')$ are obtained as

$$f_{xx} = \frac{x^2 - \sigma_0'^2}{\sigma_0'^4} f(x, y, \sigma_0')$$

$$f_{yy} = \frac{y^2 - \sigma_0'^2}{\sigma_0'^4} f(x, y, \sigma_0')$$

$$f_{xy} = \frac{xy}{\sigma_0'^4} f(x, y, \sigma_0')$$  \hfill (2.18)

In the case of a 2D bright nodular structure, clearly, $f_{xy}$ and $f_{xx}$ are negative for $|y|<\sigma_0'$, and $|x|<\sigma_0'$, which are significantly larger than values of $f_{xy}$ and $f_{yy}$. Using second derivatives computed in equation (4.4), a $2 \times 2$ point-wise Hessian matrix is formed as
\[ H(x, y) = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix} \]  \hspace{1cm} (2.19)

Then, one can solve \( H \) for two eigenvalue \( \lambda_1 \) and \( \lambda_2 \) at each pixel \((x, y)\).

### 2.3 Feature extraction

An important step for solving pattern recognition and classification problem is representation and description of each pattern using a set of discriminative attributes. The most common approaches for characterization image patterns are shape or morphology based analysis, and texture analysis of an image region depicting each pattern. For analysis of mammogram, some studies used none-image features such as the age of the patient and the family history are also used to diagnose different mammographic breast lesions [10].

#### 2.3.1 Shape analysis

This method differentiates between segmented objects, represented by binary regions, by measuring three main groups of shape descriptors. The first group included regional descriptors such as perimeter, area, compactness or circularity, orientation, extent, convex area, and eccentricity. The second group includes boundary descriptors that analyze the regularity of the object's boundary such as fractal dimension, normalized boundary moments, and Fourier descriptor. While the first and groups of shape features analyze objects individually, the third groups attempts to characterize the all objects by measuring their distribution spread from the centroid, area sum, and number of object in the entire region.
2.3.1.1 Regional descriptors

What follows describes a group of regional descriptors that commonly used for shape analysis.

*Area* $(A)$ of the region that represents a simple and straightforward shape measure, which is computed as number of pixel contained within an object margin including the boundary pixels.

The *perimeter* $(P)$ of a region, a measure of the region circumference, is estimated using a chain code method. Using a chain code method [12], segments connected the boundary pixels are labeled from 0 to 7 or 0-3, if 4- or 8-connectivity criterion is used to produce chain code of an object. Then, $P$ of an object is estimated from chain code as

$$P = n_e + n_o \sqrt{2}$$ \hspace{1cm} (2.20)

where $n_e$ and $n_o$ are the number of even and odd labels of the chain code.

Using the area $A$ and perimeter $P$ of a region, the *Compactness* of the region, which is a dimensionless, rotation, and scale invariant measure, is estimated as follows

$$Compactness = \frac{P^2}{4\pi A}$$ \hspace{1cm} (2.21)

Theoretically, a circular object is the most compact one, which produces the minimum value of *Compactness* that is equal to one. Hence, the *Compactness* of the region increases as the shape approaches a line-like structure.

Eccentricity $(\varepsilon)$ of a region, is a dimensionless descriptor that is typically between 0 and 1, is a measure of the degree of which the mass of the region is concentrated along a specific axis. A region’s eccentricity $\varepsilon$ is computed as
\[ \varepsilon = \frac{(m_{0,2} - m_{2,0})^2 + 4m_{1,1}}{(m_{0,2} + m_{2,0})^2} \]  

(2.22)

where \( m_{p,q} = \sum_x \sum_y x^p y^q f(x, y) \) represents a 2-D moment of order \( (p + q) \) of an image function \( f(x, y) \) with \( p, q = 0, 1, 2, \ldots \), [12].

### 2.3.1.2 Boundary descriptors

Several approaches have been introduced to describe the boundary of a region [12]. Among these methods, moments of the shape boundary and Fourier descriptors are very common methods, which will be used to characterize the malignancy of the shape of microcalcifications in Chapter V of this dissertation.

#### 2.3.1.2.1 Moments of the shape boundary

Moments of the region boundary are defined as moments of the Euclidian distances between ordered pixel sequences located on boundary and the centroid of the region. Let the Euclidian distances of region contour pixels \( x(i), y(i) \) and its centroid as \( z(i), \ i = 1, 2, \ldots, N \).

Then, a \( pth \) moment can be expressed as [20]

\[ m_p = \frac{1}{N} \sum_{i=1}^{N} [z(i)]^p \]  

(2.23)

Moreover, one can define a set of the \( pth \) central, and transition invariant moments as

\[ \mu_p = \frac{1}{N} \sum_{i=1}^{N} [z(i) - m_1]^p \]  

(2.24)
A set of low order moments ($F_1', F_2'$, and $F_3'$) [20] have been demonstrated to be more robust to the presence of the noise than high order moments, which can be appropriate for classifying different object using their shape. These rotation, scale, and translation invariant low order shape moments are estimated as follows [20]

$$F_1' = \left[ \frac{1}{N} \sum_{i=1}^{N} [z(i) - m_i] \right]^{1/2}$$

$$F_2' = \left[ \frac{1}{N} \sum_{i=1}^{N} [z(i) - m_i] \right]^{1/3}$$

$$F_3' = \left[ \frac{1}{N} \sum_{i=1}^{N} [z(i) - m_i]^{4} \right]^{1/4}$$

(2.25)

The difference moment $F_3' - F_1'$, denoted $F_4'$, is another shape boundary moment that is shown very efficient for describing the shape irregularity [20]-[21].

2.3.1.2.2 Fourier descriptors

An important technique to describe the shape of an object is using Fourier descriptors [20]. Using normalized Fourier descriptors (NFD) of the region boundary, a shape descriptor FF is defined as follows [20]

$$FF = \left[ \frac{N/2 \sum_{\mu=-N/2+1}^{N/2} \|NFD(\mu)\|}{\sum_{\mu=-N/2+1}^{N/2} \|NFD(\mu)\|} \right]$$

(2.26)
A Fourier descriptor $FD$ of a given boundary pixel $z_i$ at coordinate $x_i, y_i$ is computed by defining a complex number $z_i$

$$z_i = x_i + jy_i, \quad i = 0, 1, ..., N - 1$$  (2.27)

Then, one can compute a discrete Fourier transform of the contour coordinates, usually implemented using FFT algorithm, as

$$FD(i) = \frac{1}{N} \sum_{i=0}^{N-1} z_i \exp[-j2\pi i / N]; \quad i = 1, 2, ..., N - 1$$  (2.28)

FDs can be modified to be invariant to a change in scale, rotation, and translation (position of the object). These normalized FDs are modified using three steps

- $FD(0)$ is set to 0 to obtain descriptors that are invariant to the position of the object.
- $FDs$ are normalized by $FD(1)$ to ensure scale invariant property.
- The magnitude of $FD$ is used to obtain a rotation invariant descriptor.

$$NFD(u) = \begin{cases} 
0 & u = 0 \\
FD(u) / FD(1); & u = 1, 2, ..., N / 2 \\
FD(u + N) / FD(1); & u = -1, -2, ..., -N / 2 + 1
\end{cases}$$  (2.29)

2.3.2 Texture analysis

The texture analysis of a given image can be accomplished using pixel- and region-based description and feature extraction. A straightforward example of pixel based texture analysis is based on using a gray-level intensity of each pixel to characterize a given region.

Region based texture analysis is based on describing each region using different attributes (gray-level average, entropy) or by using measuring one attribute from multiscale or sub-band representation of each image.
The most popular examples of region based texture analysis are features representing the first order statistics of the gray-level histogram, statistical features from analyzing the local statistics such as measuring high order statistics (skewness, kurtosis) and image modeling using Markov random filed. More popular texture analysis techniques are based on the analysis of the second order histogram using gray-levels co-occurrence matrices (GLCM) or Haralick's measures [22] that is also known as spatial gray-level dependence matrices (SGLD) [23], and texture features from spectral analysis using wavelet transforms and discrete cosine transforms. Other approach to derive texture features are surrounding region dependence method [1], Laws measures of texture [24], and gray-level run length method [25].

Region based characterization is computationally more attractive and more feasible when a large image is being investigated or if a multiscale image representation (e.g. dyadic wavelet transform) leads to several scales of different sizes. However, the selection of a pixel or a region based texture analysis depends on the pattern recognition problem itself and that the region based method is more suitable when the goal is to draw a single decision on the whole region (e.g. a given region is malignant or benign).

In this work, identifying image pixels of the desired patterns is accomplished by using multispectral image representation [12] that consists of pixel’s intensity or gray-level, local statistics estimated using overlapping 9\times9 window centered at each pixel, and responses from a spatial domain and a wavelet based filtering scheme. For characterizing a class label of a given image region (e.g. the malignancy of a given image microcalcification cluster), this thesis employs a multiscale texture analysis and feature extraction, presented in Chapter six, which describe each scale of an a given image region using different spectral
quantities such as energy and entropy [23].

2.3.2.1 Spectral measures

A transform domain approach to analysis the texture of an input image is usually accomplished by construction a multiscale representation or sub-band decomposition of an image and computing normalized energy and entropy or each scale (of frequency sub-band). The normalized energy [23] of each scale of size $M \times N$ is computed as follows

$$E_{\text{norm}} = \frac{E}{M \times N}$$

(2.30)

where $E = \sum_{i} \sum_{j} [x_{ij}]^2$ is the scale’s energy, $x_{ij}$ is the pixel value at location $(i, j)$ in the transformed image.

Moreover, the normalized entropy [23] is defined as

$$Entropy = -\frac{\sum_{i} \sum_{j} \left[ \frac{x_{ij}^2}{E} \right] \log_2 \left[ \frac{x_{ij}^2}{E} \right]}{\log_2 [M \times N]}$$

(2.31)

2.3.3 Non-image features

For applications like computerized interpretation of medical images (e.g. mammograms in this work) none-image features such as human-based interpretation by an expert, patient’s age, family history, and some clinical attributes and fact on the pattern being examined can be integrated with other features to improve the performance of CAD [10].
2.4 Feature selection

Feature selection commonly used to search for an optimal feature subset that improves the classification performance and generalization ability of the classifier, to reduce the dimensionality of the feature space by discarding some inadequate features. Feature selection can also provide some knowledge and better understanding of the significant, irrelevant, and redundant features. This knowledge might be useful for realization of the system. Several feature search techniques have been used in literature. Most popular examples are sequential forward search (SFS) [26], linear discriminate analysis [27] heuristic search using Genetic algorithms (GAs) [23], [27]. Although these techniques do not guarantee a global solution obtained using a computationally expansive exhaustive search method, they are still more effective and produce near optimal solution in most cases.

During feature search process, subsets of features commonly evaluated using feature filters, wrappers [10], [28], and embedded methods [29]. Filters approaches perform feature ranking based on the distribution of the input data independently of the classification process. Examples of feature filters are single variable or univariate techniques such as receiver operation characteristic (ROC) analysis, statistical t-test, and Fisher-score method [28], and multivariate feature filter using stepwise linear discriminate analysis [27]. Wrappers methods accomplish the feature selection task by searching for an optimal feature subset based on the performance of a learning machine with fixed learning model [29]. Embedded methods integrate feature selection process with classifier’s learning and model selection stages [29].

An embedded feature selection strategy combines a feature selection task with the optimization of the classifier’s performance. Hence, in this dissertation a heuristic search
using particle swarm optimization (PSO) [30], presented in Chapter Five of this dissertation, is not only intended to find the best features but also to accomplish parameter selection of a successive classification stage.

In this section, we present different approaches for accomplishing a feature selection task. The first approach is based on using a single variable feature ranking for constructing a feature search space. Then, can use different search methods to select the best feature subset. In the second approach, heuristic search using PSO method is used for accomplishing the entire feature selection process that includes both the generation of the candidate feature subsets and search for best subset or at least for selecting the best feature subset from a given search space.

Since the details of a heuristic search using PSO method is not significantly different from one application to another, in the next subsection, we will present some theoretical background of a PSO heuristic search, which will be used throughout this work for accomplishing heuristic parameter selection.

2.4.1 A heuristic search using PSO

Several studies have selected a heuristic feature search based on particle swarm optimization (PSO) algorithm instead of using a genetic based algorithm since the former is proven to be a more computationally efficient and a very competitive alternative of GAs based methods [30]-[32].

Particle swarm optimization (PSO) [33], introduced by Eberhart and Kennedy in 1995, is a population based heuristic search approach inspired by the social behavior of the flocks of birds and the schools of fish, where a group of individuals (particles) located in the
parameter space of an objective function search for the optimal solution. During the search process, the location of the personal best fitness achieved by each individual as well as the global best fitness achieved by the whole swarm (all particles) are memorized, which will be used to determine the movement (search velocity and direction) of each particle, in the parameter space.

Mathematically, the \( kth \) particle of the swarm that is a candidate solution to a given objective function or a bird of the flock search for food is modeled as a \( d \)-dimensional vector in the search space expressed as \( x_k = [x_{k1}, x_{k2}, \ldots, x_{kd}] \). The location of the personal best fitness (previous best experience) of the \( kth \) particle can be defined as \( x_k^{pBest} = [px_{k1}, \ldots, px_{kd}] \). In addition, the location of the global best fitness that can be achieved by the whole swarm is defined as \( x^{gBest} = [gx_1, \ldots, gx_d] \).

PSO search strategy uses the location of both best personal and global fitness to compute \( ith \) dimensional velocity and the new position of \( kth \) particle as follows

\[
v_i(t+1) = w v_i(t) + c_1 r_1 (x_p(t) - x_i^{pBest}) + c_2 r_2 (x_g(t) - x_i^{gBest}), \quad i = 1,2,..d \tag{2.32}\]

where \( w \) is constant, typically in interval \([0,1]\), represents the inertia of the movement, \( r_1 \) and \( r_2 \) are random numbers between \([0,1]\), and \( c_1 \) and \( c_2 \) are non-negative constants represent learning rates. To control the search speed, the \( ith \) velocity \( v_i(t) \) is constrained by the user to be in the range \([v_{min}, v_{max}]\).

During the search process, the location of each particle is updated using velocity computed in (32) as

\[
x_i(t+1) = x_i(t) + v_i(t+1), \quad i = 1,2,..d \tag{2.33}\]
2.4.2 Univariate based feature ranking

The main objective of evaluating features using a single variable feature filter is to rank a group of features based on their individual discriminative ability. In other words, each feature is considered as simple predictor or classifier. This single variable evaluation helps identifying the most important and irrelevant features that may lead to a better understanding the problem and the significant of each measurements, which also can be useful for successive stages of the feature selection process.

Univariate feature ranking can be accomplished using several methods [28], common approaches are ranking using ROC analysis, and Fisher-score method. According to ROC analysis technique, a feature with a stronger discrimination will produce ROC larger area under curve (AUC) or index $Az$.

Fisher criterion or F-score [34] method evaluates the predictive power ($Z_i$) by measuring the correlation of a feature represented by a real variable $x_i$ and the class label (positive or negative), which computed as follows

$$Z_i = \frac{(\bar{x}_i^+ - \bar{x}_i^-)^2 + (\bar{x}_i^+ - \bar{x}_i^-)^2}{\frac{1}{n^+_M - 1} \sum_{k=1}^{n^+_M} (x_{ik}^+ - \bar{x}_i^+)^2 + \frac{1}{n^-_B - 1} \sum_{k=1}^{n^-_B} (x_{ik}^- - \bar{x}_i^-)^2}$$  \hspace{1cm} (2.34)

where $\bar{x}_i^+$, $\bar{x}_i^-$, and $\bar{x}_i$ are the average of the feature $x_i$ of the positive, negative, and whole samples, respectively. Also, $n^+_M$ and $n^-_B$ are representing the number of samples from malignant and benign classes respectively.

One approach to perform feature selection utilizing the above feature ranking process is simply by forming $N$ feature subsets according to a nested subsets method. This
nested subsets method uses a threshold value to include and exclude features with ranking score larger or smaller than this threshold level. By varying the value of the threshold, one can construct \(N\) feature subsets of size 1 to \(N\) features.

2.4.3 Outweighed univariate based nested subsets method

This method does not rely entirely on PSO method to create candidate feature subsets but it adopts the nested subsets method to generated \(N\) candidate feature subsets from the \(N\) features ranked individually. We follow this by an embedded feature selection procedure using PSO-SVM algorithm. Our nested subset approach can be briefly described as follows:

- Univariate feature ranking criterion such as ROC analysis, which uses each feature as a simple classifier, is employed to evaluate features individually.

- A feature score that is an area under ROC curve, which is typically between 0.5 and 1, is used to form \(N\) nested subsets. By first sorting \(N\) features, in descending order, using results from ROC analysis method. Then, the first subset will include only one feature that achieved the highest score while the second subset will include features ranked at first and second position. This process is repeated until \(N\)th subset is formed.

An advantage of this simple method over exhaustive SFS and GA methods is the size of the search space, which consists of \(N\) subsets in case of univariate based nested subset compared to \((2^N - 1)\) subsets from exhaustive search methods. It is also more computationally attractive than stochastic methods such as real and binary GAs feature search [23].
The main shortcoming of forming different feature subsets using only a single variable evolution is the fact that truly redundant or highly correlated features may exist within subsets. Therefore, this study uses an average cross-correlation between a candidate feature and features already included as an additional criterion to control the redundancy level among selected features. Such process penalizes the ranking score of a potential feature if this feature shows high correlation with others already the subset.

Let $Z^0_n$ be the original ranking score of the $n$th candidate feature to be add to a given subset. Then, the average cross-correlation value between $n$th feature and features already included is used to compute new ranking score $Z_n$ as follows [35]

$$Z_n = Z^0_n (1 - u \frac{1}{n-1} \sum_{i=1}^{n-1} \rho_{ni}) \quad n > 1$$

(2.35)

where $u$ is a weight constant that can be set between 0 (discard the redundancy) and 1 (highest penalty). $\rho_{ni}$ is the cross correlation between $nth$ candidate feature and $ith$ feature previously added.

Using this method, a nonzero value of $u$ might alter the ranking score of the remaining features leading to a new structure of the nested subsets.

2.4.4 Feature selection using binary PSO

PSO based feature selection method [30], [36] is similar in principle to GA based method proposed by Seidlecki and Skalanski [37]. Using binary PSO based feature search method, each particle in the swarm represents a candidate feature subset that coded as N-dimensional binary string with each component randomly assigned a value 0 or 1 [38]. Coordinates of each particle are assumed real valued random variables uniformly distributed
between 0 and 1. Hence, this study converted real representation of each particle into a binary string by assigning a binary 1 to all components larger than a statistical mean of all coordinates that is also a real number between 0 and 1. A binary 0 is also assigned to all coordinates less than this statistical mean. This binary conversion of the coordinates is different from original binary PSO [36], [38], which compared a logistic transformation of the new coordinate velocity $v_\mu(t+1)$ with a random number between 0 and 1 to determine new location $x_\mu(t+1)$ of the corresponding coordinate.

The dimensionality of the feature space of each particle is determined by counting the number of binary 1’s it contains. Each particle is evaluated by first constructing a corresponding feature vector or subset by including a feature if a binary 1 presents and removing a feature if binary 0. Then, the fitness value defined as generalization error (or classification accuracy) of the classifier is estimated for each feature subset.

2.5 Binary classification using supervised learning

Following the human methodology of learning, machine learning or artificial intelligence is the process of developing ability of predicting type or class of a new unseen data using a knowledge gained form memorizing previously seen examples. A supervised learning machine is an algorithm that uses set of observations or measurements chosen and labeled manually to construct a learning model or mapping function. Such learning model can be employed, in a testing phase, to predict (or classify) the label of an unseen pattern.

Differentiating between abnormal and normal breast tissue, malignant and benign abnormality is usually modeled as a two-class pattern recognition or binary classification
problem. An abnormal or malignant target is commonly labeled as a positive class while healthy or benign target is labeled negative.

Supervised learning machines are usually grouped into parametric or non-parametric schemes. A parametric or generative learning machine assumes a prior knowledge on the probability distribution of different patterns, which requires some parameter to be estimated during the learning process. The most popular example of this generative learning approach is a statistical Bayesian classifier. On the other hand, none-parametric methods do not make any assumptions but use different regression and optimization techniques to find an appropriate linear or none-linear decision boundary. The simplest example of non-parametric learning machines is the k-nearest neighbor classifier [39] while the neural network and support vector machine based methods are more efficient examples that are commonly used.

2.5.1 Statistical Bayesian learning

Bayesian classifier (BC) is a statistical method used for classification by maximizing a class a posteriori probability. The application of Bayesian classifiers for pattern recognition assumes a prior knowledge of an analytical expression of the probability density functions of various classes. Using a sufficient statistic sample patterns of each class, one can properly estimate the necessary parameters of its density function. Mathematically, a Bayesian classifier [12] has a decision function of the form

\[ d_j(x) = p(x | \omega_j) P(\omega_j) \quad j = 1, 2, \ldots, M \]  

(2.36)

Where \( p(x | \omega_j) \) is the conditional probability density function of n-dimensional feature vector \( x \) belonging to a class \( \omega_j \), \( P(\omega_j) \) is the priori probability of class \( \omega_j \), and \( M \) is the
number of classes. Assuming the probability density functions (PDF) of the measured features are Gaussian [40], [41] then, the n-dimensional Gaussian density function can be expressed as

\[ p(x \mid \omega_j) = \frac{1}{(2\pi)^{n/2} |C_j|^{1/2}} e^{-\frac{1}{2}(x-m_j)^T C_j^{-1} (x-m_j)} \]  

(2.37)

where \( C_j \) and \( m_j \) are the covariance matrix and the mean vector of class \( \omega_j \), respectively. Also \( |C_j| \) is the determinant of the matrix \( C_j \).

Since the decision function given in equation (2.37) is monotonically increasing and because of the exponential nature of the Gaussian density function of the conditional probability \( p(x \mid \omega_j) \) [12], the decision function can be rewritten as

\[ d_j(x) = \ln p(\omega_j) - \frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln |C_j| - \frac{1}{2} [(x-m_j)^T C_j^{-1} (x-m_j)] \]  

(2.38)

Since the term \( \frac{n}{2} \ln(2\pi) \) is common for all classes and assuming that all classes are equally likely, the decision function \( d_j(x) \) reduces to:

\[ d_j(x) = -\frac{1}{2} \ln |C_j| - \frac{1}{2} [(x-m_j)^T C_j^{-1} (x-m_j)] \]  

(2.39)

A feature vector \( x \) is assigned to a class \( \omega_j \) with a minimum distance \( d_j(x) \), [12].

2.5.2 k-nearest neighbor (kNN)

kNN classifier [39] is a classical and simple approach to accomplish learning task, which uses an experience (a similarity) from previous training patterns to classify a newly test data into a certain class label (a malignant or benign in this study). Examples of the metrics commonly used to measure the similarity between input pattern and previously
received data are: correlation, Euclidian distance. Using Euclidian distance as similarity metric, kNN method classify an input pattern into a certain class that has a majority vote among k neighbors. These neighbors are the set of training patterns with smallest Euclidian distance among all training samples. Another similarity metric also used inverse distance weighted voting to assign higher vote to the label of closest neighbors [42].

2.5.3 Support vector machine (SVM)

Support vector machine (SVM) is a supervised learning machine that utilizes a structural risk minimization principle to reduce the generalization error of learning machines [43]. SVM classifier has been recognized as one of the most powerful supervised learning machines [29], [44] as it has been demonstrated to be more efficient for MCs detection and classification than other popular learning machines [26], [45]-[46]. The basic principle of pattern recognition using SVM is based on finding an optimal hyper-plane in the input feature space that maximizes the separation (geometric margin) among the patterns from different classes.

For a binary or a two-class linear SVM classification, input data \( x \in \mathbb{R}^n \) are two classes with a class label \( y \in \{-1, 1\} \), the learning problem is formulated as convex optimization problem subjective to a set of inequality and linear constraints. Solving linear SVM optimization leads to a decision function or separating hyper-plane in the input feature space that can be expressed as

\[
d(w, x, b) = w^T x + b
\]  

(2.40)

where \( w = [w_1 \; w_2 \; \ldots \; w_n]^T \), \( w \in \mathbb{R}^n \) is the weight vector and \( b \) is a bias constant.
Assuming that all data are linearly separable is not the general case. In most cases, data is either nonlinearly separable (i.e. requires a nonlinear decision boundary) or data is not linearly separable in the original feature spaces but can be linearly separated in a higher dimensional feature space. This higher dimensional feature space is usually obtained using a nonlinear mapping called kernel function \( \Phi(x) \) [47], known as "Kernel trick", which maps the original feature space into higher (even infinite) dimensional feature space. A nonlinear or kernel mapping of the original feature space leads to SVM decision function expressed as follows

\[
d_{\text{Nonlinear}}(w, x, b) = w^T \Phi(x) + b
\]  

(2.41)

Sometimes mapped data remain nonlinearly separable and cannot be classified using hard-margin classifier due to some training errors (i.e. data overlapping). Such situation is resolved using a soft margin classifier and slack variable concept introduced by Cortes and Vapnik [48]. A standard formulation of a soft margin SVM learning problem is a convex optimization problem subject to a set of inequality constraints written as follows

\[
\min_{w, \xi} J(w, \xi) = \frac{1}{2} w^T w + C \sum_{i=1}^{L} \xi_i
\]  

(2.42)

Subject to:

\[
w^T \Phi(x_i) + b \geq 1 - \xi_i, \quad i = 1, 2, ..., L
\]

where \( \xi_i \) is a positive slack variable, \( C \) is a positive regularization or penalization parameter, which corresponds to a training error that must be adjusted during a model selection process, and \( L \) is the number of training examples.

SVM learning problem can be solved using dual or primal Lagrangian formulations of equation (2.42). Although both approaches can lead to the same global solution, the dual formulation is commonly applied because the solution of the dual depends on the
Lagrangian multipliers (finite number that is less or equal to the number of training examples) rather than of the weight coefficients that might be infinite due the high dimensional nonlinear mapping of the feature space.

The dual formulation of SVM optimization problem is expressed as

$$L_D(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$

(2.43)

Subject to the following constraints

$$\sum_{i=1}^{n} \alpha_i y_i = 0 \text{ and } 0 \leq \alpha_i \leq C$$

where $\alpha_i$ is a dual Lagrange multiplier, $K(x_i, x_j) = \Phi(x_i)^T \Phi(x_j)$ is a kernel or nonlinear mapping function, which is simply the dot product $x_i^T x_j$ for the case of a linear SVM.

The results of solving SVM dual optimization is the decision function described by a set of Lagrange multipliers $\alpha_i$ and a bias constant $b$, which can be used to compute the class label $\tilde{y}_p$ of an input test pattern $x_p$ as follows

$$\tilde{y}_p = sign(\sum_{i}^{N_{sv}} \alpha_i K(x_p, x_i)y_i + b)$$

(2.44)

In equation (2.44) $N_{sv}$ is the number of support vectors, which represents the complexity of the decision function that consists of all training patterns located close to the boundary with nonzero $\alpha_i$ and achieve an optimality condition $\alpha_i (w^T \Phi(x_i) + b - 1) = 0$. It is worth noting that classifying a given test pattern using nonlinear decision boundary given in equation (2.44) does not require explicit computation of the nonlinear mapping $\Phi(x_i)$ but the outcome of a kernel function $K(x_p, x_i)$. Any nonlinear mapping function can be used as SVM kernel provided that such function is positive definite and satisfies Mercer's condition [47]- [48]. Several kernel functions commonly reported in literature such as the Gaussian,
also known as the radial basis functions (RBF), and the polynomial kernel [47]. A Gaussian or RBF kernel with real and positive parameter $\sigma$ is expressed as

$$K_{RBF}(x, y) = \exp\left(-\frac{(x - y)^2}{2\sigma^2}\right)$$ (2.45)

While a polynomial kernel with an integer parameter, $P$, can be written as

$$K_{poly}(x, y) = (1 + x^T y)^p$$ (2.46)

2.6 Performance evaluation

2.6.1 Results analysis

Binary classification problem associated with a detection or diagnosis scheme is usually evaluated by examining the accuracy of the decision making process. A simple and mostly an early evaluation measure of the classifier performance is using an overall classification Accuracy or equivalently the generalization error ($1.0 - Accuracy$) of the classification process [49]-[50], is usually defined as follows

$$Accuracy = \frac{N_r}{N}$$ (2.47)

where $N_r$ is the total number of correctly classified patterns or examples, and $N$ is the total number of tested patterns.

Measures commonly used to evaluate the classification results are the rate of the correct classifications of positive (malignant or abnormal) and negative (benign or healthy) patterns, which are known as Specificity and Sensitivity, respectively. Computing the Sensitivity and Specificity measures requires us to define some necessary parameters by
interpreting the outcome of the classification process. These parameters include true positive (TP), false negative (FN), false positive (FP), and true negative (TN) rates which are defined as follows [50]

- A true positive rate $TP$ represents the probability of the correct prediction of an abnormal or positive target, which can be used also to compute the $FN$ rate as $1 - TP$.
- A false positive $FP$ rate is the probability of incorrect prediction of the negative (i.e. Benign or background) target. Similarly, we can compute a true negative $TN$ rate is computed from a $FP$ rate as $1 - FP$.

Using $TP, FN, TN,$ and $FP$ rates of a given predictor or classifier, a $2 \times 2$ confusion matrix can be constructed in which the sum of each column or row being equal to 1, as shown in Figure 2.4.

<table>
<thead>
<tr>
<th>Actual value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted value</td>
</tr>
<tr>
<td>True Positive</td>
</tr>
<tr>
<td>False Negative</td>
</tr>
</tbody>
</table>

Figure 2.4: The confusion matrix.

Using the confusion matrix, $Sensitivity$ and $Specificity$ metrics are computed as follows

\[
Specificity = \frac{TN}{TN + FP} \tag{2.48}
\]

\[
Sensitivity = \frac{TP}{FN + TP} \tag{2.49}
\]
Although the use of true positive fraction (TPF) and false positive fraction (FPF) pairs, computed as $1.0 - \text{Specificity}$, or Sensitivity and Specificity pairs to measure the performance of a supervised learning machine is more efficient than the classification accuracy. The former metrics do not solely describe the performance of the classifier because their values depend on the selection of the threshold or the bias of the decision function. A statistically more convenient and widely used technique is the receiver operating characteristic (ROC) curve [49]. An ROC curve method, which was originally proposed by signal processing committee for analyzing radar based detection systems [51], represents a 2-D plot of the Sensitivity or true positive fraction (TPF) vs. false positive fraction (FPR) computed as $1.0 - \text{Specificity}$, obtained by varying the value of a classification threshold. Examples of ROC curve of different classifiers are shown in Figure 2.5.

![ROC curve example](image)

Figure 2.5: ROC curve examples. Curve A corresponds to a random guessing classifier with area under ROC curve (AUC) of 0.5, curves B and C represent two classifiers with AUCs larger than 0.5.

Sometimes, it is not possible to accurately estimating the false positive rate instead the average number of false positive target per image to construct an alternative curve
commonly called free response receiver operating characteristic (FROC). For evaluation and comparison purposes area under ROC curve (AUC) or an index Az, computed using trapezoidal method is commonly used. An ideal ROC curve of an optimal classification performance leads to $Az = 1.0$. Since a random guessing produces ROC curve represented by a diagonal line that produces $Az$ of 0.5, any realistic classifier must have $Az$ value between 0.5 and 1 [51].

2.6.2 Validation methods

Cross validation (CV) is a very common approach for estimating the generalization ability (i.e. classifier’s performance on the test dataset) and for selecting appropriate learning model of a supervised learning machine [52]. A cross validation technique, basically, divides the data population into $k$ folds or bins, each time $k-1$ folds are used for training and 1-fold hold out for testing. This process is repeated $k$ times, and the generalization performance computed as average of $k$ testing experiments. The selection of number of folds $k$ mainly depends on the number of data available for testing and training, which usually set to 5 or 10. An extreme case of the $k$-fold cross-validation scheme occurs when $k$ is equal to the number of examples, which is known as the leave-one-out (LOO) cross-validation case. One advantage of the LOO training and testing scheme is that such a method represents an unbiased estimator of the generalization performance of a given classifier. Also, the LOO method is very attractive for small scale studies like the dataset considered in this work. For large-scale problems, one can partition the dataset into training, validation, and testing subsets [26].
CHAPTER III

COMPUTER AIDED DIAGNOSIS IN MAMMOGRAPHY

3.1 Breast cancer

Cancerous cells, which might exist in a human, distinguish themselves from normal cells by growing without any biological control, by developing and by clustering to form tumors that have different nature from healthy cells. Whereas the growth rate of the normal cells gradually decreases as a human ages, abnormal cells continue to divide indefinitely. Based on the risk these abnormal cells may cause, cancerous cells are divided into malignant and benign cells. Malignant cells usually tend to invade and destroy surrounding tissue. Sometimes, malignant cells spread from the organ they originated from to reach other body organs through the blood or lymph system and cause what is known as metastasis [53], which might extend to the entire body if cancer reaches an incurable and deadly stage. Benign tumors, on the other hand, are recognized by their limited spread and mostly remain localized where they initiated. Benign tumors are in most cases less danger, curable, and mostly irreproducible after appropriate removal and treatment.

Cancer of all types continues to be a serious threat of a human life regardless of the age with higher cancer’s risk as a human gets old. According to statistics from World Health Organization (WHO) [54], cancer is responsible for 13 % of all deaths in the world in 2004. Among all cancers, breast cancer is a major cause of death among women in the US and other parts of the world [4]. Statistical studies indicated that the spread of breast cancer as
well as other types of cancer differ from one geographical place to another. For example, the incident rate of breast cancer is much higher in the developed countries than in other parts of the world such as Far East and Africa [9].

Attempts by researchers to link the genetics, diet, and environmental factors with the developing of a breast cancer indicated that the age and family history are factors of highest risk of developing breast cancer. Unlike the lung cancer that has been strongly linked to tobacco smoking as a significant cause, researchers could not find a similar cause of breast cancer. However, studies reported that obesity of the body, and some diet elements such as drinking alcohol slightly increase the risk of breast cancer. In particular, women who daily consume 2-5 alcoholic drinks are at one and half time risk of women drink no alcohol [55].

Since the development of breast cancer remains unexplained and there is no medical procedure to prevent its occurrence, mammography based screen programs have been established in many countries to fight against breast cancer. These screening programs mainly aim at detecting breast cancer in its infancy and curable stage by encouraging women at age 40 and older with no previous history of developing breast cancer to undergo screening mammography. Mammogram screening efficacy relies on two facts. The first fact accounts for the high risk of breast cancer as woman gets old. The second fact is based on the proven effectiveness of mammography in detecting early stage breast cancer.

3.2 Screening of breast cancer

Radiographic manifesting of the breast cancer, namely, mammography was the earliest non-invasive medical imaging system employed to screen the breast cancer. Meanwhile, several none-ionized medical imaging systems, such as magnetic resonance
imaging (MRI) and ultrasound (US), are used for imaging various parts of a human including the breast. However, these systems are currently used as adjunctive tools and mammography is the only imaging system approved by the US Food and Drug Administration (FDA) as a screening device. Screen mammography has been shown to detect breast cancer two years earlier before it becomes palpable [56]. Approximately, mammography can detect up to 85% of breast cancers [57] and can lead to a 30% reduction of the deaths from breast cancer [2]. Moreover, clinical studies claimed that performing screening mammography on all women could reduce deaths from breast cancer by 36% to 44% [57].

3.2.1 Screen film mammography

Mammography is x-rays based diagnostic imaging of the breast, which projects the structure of a 3-D object (i.e. female breast) onto 2-D receptor (i.e. screen film). Briefly, a low energy x-ray beam, generated from specially designed x-ray tube, is used to image a compressed breast at different points and field of views. Unlike a general radiography system, mammography uses a special x-ray tube designed to generate low kilovolts (kV) and a low energy x-ray beam with a small focal spot to ensure the best image quality of the breast at low radiation dose. Several factors are usually considered to reduce imaging artifacts and amount of x-ray exposure such as breast compression, geometrical setting such as appropriate source to film distance, source to image distance, and focal spot size.
Figure 3.1: Illustration of mediolateral oblique (MLO) and craniocaudal (CC) projections of a female breast, which are acquired during a screen mammography exam. Figure was obtained from [57].

During the screen mammography of the breast, two projections (point of view of a breast), as illustrated in Figure 3.1, are commonly acquired. The first projection is a craniocaudal (CC) projection in which an imaged breast is located vertically underneath an x-ray tube. The second projection is mediolateral oblique (MLO) view where the mid-axis of an x-ray tube is angled between 30°-70° from the longitudinal axis of the imaging system [5]. An Example of MLO and CC screen mammograms of a right female breast is shown in Figure 3.2.

In screen film mammography, a produced x-ray film is developed like a regular photograph and viewed on a light-box to be interpreted by radiologist. The current technology of computer vision and digital image handling (view, process, and archive) require screen films to be digitized, which is accomplished using a high resolution film digitizer to produce a high resolution digitized mammogram with minimal information loss. The availability of digital mammograms enables a smooth migration to filmless radiology department, digital image archive and transmission, and application of up-to-date digital
image processing and computer vision algorithms.

Figure 3.2: Craniocaudal (CC) and mediolateral oblique (MLO) view mammograms of a right female breast. Image courtesy of Dr. Desai, Radiopaedia.org [58].

3.2.2 Full field digital mammography

Similar to the conventional screen film mammography, full field digital mammography (FFDM) uses x-rays to generate a high quality image of the breast. Digital mammography is mainly different from a screen film mammography in that it uses a digital receptor or sensor and computer based image formation techniques rather than film cassette and chemical process to generate images [59]. Obviously, the benefits of the digital mammography are mostly faster image acquisition and construction, short scan time that might minimize scan repetition and minimize patient’s exposure to x-rays. Additionally,
high-resolution digital images produced by digital mammography enable radiologists to magnify, orient, and adjust image’s contrast and brightness even after an exam is completed [60]. Digital mammography also provides more convenient image review, archive, transmission, and integration with recent developments such as Picture Archiving, and Communication Systems (PACS), remote diagnosis, hospital information system (HIS), and radiology information system (RIS).

After receiving FDA approval, FFDM has become a standard tool for breast cancer screening. There are many advantages of digital mammography over a screen film method [61]. However, several studies have applied both methods and demonstrated that digital mammography can provide a comparable performance but not a significant improvement of the diagnosis accuracy [4].

3.2.3 Adjunctive breast imaging techniques

Mammography is exclusively proven effective for breast cancer screening, the specificity and sensitivity of mammography degrades significantly as the breast tissue gets denser in young women. These limitations in addition to the ionizing radiation used by mammography have motivated the application of other existing technologies such as MRI, US, positron emission tomography (PET) imaging, and the exploration into newly developed tools such as microwave imaging. Among these adjunctive breast imaging methods, MRI is very promising, which can effectively diagnose some cases of breast cancer. However, MRI breast imaging is not yet proven to be a substitute of mammography due to several limitations of breast MR imaging [62]. Compared to mammography, current breast MRI is more expansive, imaging time is relatively longer, and a contrast agent needs to be taken
before image acquisition. In addition, breast MRI does not provide sufficient specificity to
distinguish between a benign and malignant tumor [57]. Another limitation of breast MRI is
the deficiency for imaging the breast calcifications, which are strongly correlated with an
early-stage breast cancer. However, American Cancer Society (ACS) has recently supported
the use of breast MR for an annual screening of young women with high risk of breast
cancer.

3.3 Breast abnormalities

Breast abnormalities that commonly appear on a mammogram are masses, clusters
of calcifications, distortions in the breast architecture, and breasts’ asymmetry. Among these
abnormalities, calcifications and masses are the most important ones because they are signs
of early stage cancer and can be shown on a mammogram before physician can feel them.
Examples of different breast abnormalities are presented in Figure 3.3, which presents
benign microcalcification clusters, circumscribed and speculated masses, and architecture
distortion.

3.3.1 Calcifications

Breast calcifications, including macrocalcifications and microcalcifications (MCs),
are tiny deposits of calcium that appear as bright spots on a mammogram, which form in
breast as a woman gets older. Breast calcifications vary in shape, size, and density.
However, not all breast calcifications or deposits of calcium are cancerous because some
calcifications might be results of breast inflammation and trauma. In addition, there is no
clinical evidence of any relation between women’s diet and the presence of breast
calcifications. Compared to malignant calcifications, macrocalcifications mostly turned out to be benign and relatively they have large size, regular and almost round shape. Macrocalcifications also have uniform density and sharp outline, which might appear isolated with a scattered distribution. Microcalcifications, on the other hand, have higher degree of suspiciousness and malignancy, which a usually appear on mammogram in form of clusters, which is defined as mammographic region of an area of 1 cm² with 3 or more calcifications. Malignant MCs vary greatly in size, morphology and shape irregularity, distribution, and orientation. Mostly, they have a small size that is between 0.1-1 mm with an average size of 0.3 mm [1]. In young women where breast tissue is denser, microcalcification can be of low contrast and indistinguishable from their surrounding that might impair segmentation and interpretation using various techniques.

3.3.2 Masses

According to BI-RAD™ [63] a mammographic mass is defined as a space occupying lesion seen in two different projections. If a potential mass is seen only in a single projection, it should be called a "Density" until its three-dimensionality is confirmed. The degree of the speciousness of masses is usually characterized by analyzing their shapes and margin properties. Two types of breast masses are commonly seen on mammograms: circumscribed and spiculated masses, which are shown in Figure 3.3 (c) and (d). A Circumscribed mass is a breast lesion that is high likely benign when appears round, oval, and lobulated, with a well defined margin that is strongly distinguished from surrounding breast structure. However, even a round and well defined lesion might be diagnosed as cancer, which makes a circumscribed lesion one of the most controversial mammographic findings [64].
Figure 3.3: Examples of the most common breast abnormalities. (a) A benign microcalcification cluster, (b) a benign mass (circumscribed lesion), (c) a malignant mass (spiculated lesion), and d) an architectural distortion (AD). Images are from the MIAS database.

A *speculated* mass, also called a satellite lesion because it looks like a star object, is recognized as a mass of irregular margin or boundary with radiating patterns from its periphery. Compared to circumscribed masses, speculated masses are more likely malignant [65]. Speculated masses also might have different sizes, which range from a few millimeters to several centimeters. Another breast abnormality that looks like a mass on a mammogram is a breast *cyst* that is a collection of fluid. Fortunately, breast cysts are commonly turned out
to be benign, which can be further diagnosed using ultrasound and fine needle aspiration exams [57].

3.3.3 Architectural distortion

Architectural distortion is recognized by the presence of many speculations radiating along duct lines from a focal point rather than a central visible mass [63], [66]-[67], if the flow of the normal structure is directed toward the nipple, suspiciousness of malignancy is high and biopsy is recommended. However, benign distortion might occur as a result of superimposition of normal breast structures.

3.4 Radiologist based interpretation of mammograms

If a screening mammography indicates any suspected breast abnormality, a woman is directed to undergo a diagnosed mammography that includes additional mammographic views, magnified and/or spot mammography, and sometime additional examinations using an adjunctive system such as ultrasound or magnetic resonance imaging is called. Based on the radiologist interpretation of these exams (mammogram and other supplemental medical tests), a breast biopsy might be requested if a given breast abnormality shows a high level of the suspected malignancy. Invasive breast biopsy involves obtaining anatomical specimens of the suspected breast lesion via minor surgical operation and a histological testing to prove if the breast lesion is a benign or malignant one. It is worth noting that invasive breast biopsy is the only examination can provide an evidence of the malignancy of a breast lesion.

Although mammography is recognized as the most sensitive tool for early detection of breast cancer[4],[10]-[11], mammography and its interpretation by radiologists suffer
from several limitations. One of the main shortcomings of mammography is its poor discrimination between malignant and benign lesions because a mammogram is a 2-D projection of 3-D breast structure, which produces some false alarms as results of the superimposition of normal breast tissue. Mammography is also challenged by the structure of a female-breast and by the subtle nature of various breast abnormalities such as calcifications and malignant masses. Diagnosis and analysis of mammograms by radiologist is a difficult and a time-consuming process. Even when the interpretation of mammograms is done by expert radiologists, inter-variability (same radiologist interprets a mammogram several times) and intra-variability (same mammogram is interpreted by different radiologists) is relatively high [2].

These limitations lead to a low positive predictive value (PPV) of mammography which is defined as a percentage of breast biopsies turned out to be positive. In United States, mammography PPV has not exceeded 50% [2], which means a high number of women undergoes unnecessary invasive breast biopsies as results of miss-interpreting mammograms. Furthermore, studies reported that about 50% of all cancers are missed not because of visibility problem but because radiologist fail to classify them correctly [10].

One approach to improve the interpretation of mammogram is via a double reading stage. Double reading of mammogram is usually accomplished by an expert radiologist, which proven to improve the sensitivity of breast cancer detection from 5-15% [11]. However, establishing a double reading in many parts of the world might be difficult because of the availability of specialized radiologists [11].

Early detection and diagnosis of breast cancer is crucial step toward better treatment and reduction of women death due to breast cancer. A key sign of an early stage of breast cancer is the presence of granular clustered microcalcifications [1]. Compare to other
mammographic abnormalities, MCs appear more frequently on mammograms, which represent an early sign to 30-50% of breast cancers detected using mammography. Histological examinations of the breast also indicate that 60-80% of breast cancer include MCs. This strong correlation between the presence of mammographic MCs and the development and breast cancers, along with the difficulties and challenges of interpreting mammogram by radiologists demonstrates the importance of mammographic computer aided diagnosis systems, in general, and more specifically automated detection and diagnosis algorithms for mammographic MCs [1].

3.5 Computer aided analysis (CAD) of mammogram

The main objective of computer aided analysis technology of digital mammogram, including detection and diagnosis, is not to replace radiologists but mainly to provide them with a second opinion that might increase the number of detected cancers and to improve the radiologist's ability in discriminating between benign and malignant breast lesions shown on mammograms. Hence, the final diagnosis decision on a given mammographic region is a radiologist's responsibility and any diagnosis result from CAD system is only a supplementary one.

A typical computer aided analysis of digital mammograms, as illustrated in Figure 3.4, is briefly described as follows: a digital mammogram, obtained from a digitized screen film or directly from a FFDM system, first fed to a preprocessing stage of a computer aided detection subsystem (noise removal and/or enhancement using background suppression). Detection of mammographic MCs is commonly accomplished using two stages. The first stage applies various image processing and segmentation techniques to filter mammographic
regions and detect suspicious regions of MCs. The second stage, as shown by Figure 3.4, is a machine learning based detection scheme, which consists of feature extraction, feature selection, and a binary or two-class classification using a supervised learning machine. The main purpose of this stage is to reduce the FP rate (i.e. improve detection’s specificity) without increasing the FN rate or detection’s sensitivity, which commonly uses a supervised learning machine to identify a true microcalcifications by classifying a suspected breast tissue into normal (i.e. healthy) or abnormal (MCs ) class.

ROIs contain detected microcalcification clusters, either obtained from a previous CADe system or manually selected by a radiologist, which are analyzed and classified into malignant and benign classes by CADx subsystem that is also a supervised learning machine.

As shown by Figure 3.4, a CADx subsystem is typically modeled and solved using four steps: segmentations of MCs, characterization or feature extraction, feature selection and dimensionality reduction, and classification. Clearly, the structure of CADx is similar to a machine learning based detection stage of a CADe system. The main difference between CADe and CADx is that a CADx scheme is completely modeled and solved as a binary class classification problem while many CADe systems use a supervised learning machine as a secondary stage as to improve the FP rate. Additionally, when a shape based of CADx is implemented, a CADe subsystem is needed to segment MCs. However, this stage is usually dropped when a texture based scheme is used to diagnose MCs.
Figure 3.4: Illustration of a typical computer aided diagnosis system, which consists of a computer aided detection and a computer aided diagnosis subsystems.

The Earliest attempts to automate the diagnosis of breast cancer was introduced by Winsberg and others [68]. Following Winsberg’s work several unsuccessful attempts to automate the detection and diagnosis of breast cancer can be found in literature. The failure of these studies was mostly because they planned to replace radiologist in analyzing mammograms. Realizing that automated diagnosis of mammogram is mainly to help radiologists in interpreting mammography rather than replacing them, in the mid-eighties, research groups have accomplished two studies, which have been credited for introducing a new vision and raising confidence in computer aided diagnosis of breast cancer [2]. The first study by Getty et al. [69], demonstrated that diagnosis of breast lesions can be improved
through a computer based processing of the radiologists’ interpretation, which are manually input to a computer system as a checklist. The second study by Chan et al. demonstrated that automated detection (CADe) can improve the sensitivity of mammography by detecting cancers overlooked by a group of fifteen radiologists [70]. Following these two groundbreaking studies, many researchers have devoted their efforts to develop algorithms for computer aided detection and diagnosis. Detailed review studies of CAD systems being developed can be found in [1], [4], [10], [71].

3.5.1 Computer aided detection (CADe) of microcalcifications

During the last decades, investigators have developed different algorithms to assist radiologists in the detection of microcalcification clusters on mammogram either by making such subtle breast abnormalities more visible or by providing radiologist with an automated scheme to localize MCs on a given mammogram. These methods mainly vary with respect to the techniques used in the implementation, the performance they achieved, and mammogram datasets used to evaluate each approach. However, tackling the problem of computer aided detection of MCs and other breast abnormalities continues to be a research demand [4].

Early approaches to design a computer aided detection of MCs were enhancement-based methods, which aimed at making MCs or regions enclosing MCs more distinguishable from surrounding tissue. Such an enhancement process might improve the interpretation by radiologists as well as segmentation of MCs using a threshold. The most straightforward and earliest enhancement methods are histogram equalization, contrast stretching, unsharp masking, and spatial filtering. The great variety of the appearance of MCs and the contrast between MC cluster and surrounding breast tissue limit the effectiveness of enhancement
using a set of global features and a fixed size neighborhood, which can cause undesired image artifacts (e.g. over-enhancement) and sometimes lead to missing some important regions. This fuzzy appearance of MCs, which is characterized by irregular shapes and sizes has guided researchers to use adaptive [72] and region based enhancement methods [73]-[74].

Instead of making MCs more visible by manipulating their appearance in the image domain, a feature based approach attempts to enhance the contrast of MCs by measuring their statistical features (kurtosis, skewness, and local maxima) in spatial and spectral representations of mammograms [50],[75].

Alternatively, background suppression to enhance MC has been employed in many studies [26], [76]-[79] to replace both conventional and direct enhancement methods. In this approach, clustered MCs, bright and tiny deposits of calcium, are modeled as highpass anomalies laying on a slowly varying background. Therefore, filtering using a highpass filter or subtracting a lowpass version model of a given image from an original one can lead to a substantial enhancement of MCs.

Detection and enhancement of MCs using multiscale and subband image decomposition using wavelet transforms is a very popular and effective example of this approach, which has been been employed by many investigator [41],[50], [75], [77]-[82]. Lain et al. [80] accomplished contrast enhancement of MCs by applying wavelet reconstruction after modifying the wavelet coefficient. Strickland et al. [77] concluded that by using an appropriate wavelet filter, one could easily detect and segment MCs within the wavelet domain by thresholding the wavelet coefficients before the reconstruction process. Wang and Karayiannis [78] applied wavelet reconstruction approach to detect MCs by applying a threshold to a filtered mammogram. This wavelet based reconstruction method
decomposes a digital mammogram into approximate and details subbands and uses only the
detail subbands to obtain a highpass filtered version of the input image. Following [78],
several studies used this wavelet filtering method for detecting suspected MCs [82] and to
reduce false positive results [50]. Some studies demonstrated that least asymmetric
Daubechies wavelet transforms are more suitable for enhancement of the mammogram
images such as in microcalcification detection [15] while other works demonstrated that the
design of a spatial wavelet filter with high regularity is more successful in detecting
microcalcifications than conventional wavelet filters such as the orthogonal Daubechies db4
[8]. Moreover, the non-stationary nature of mammogram image texture motivated many
researchers to design wavelet transforms using adaptive filters, which has been reported to be
more efficient than fixed FIR filters in the detection of low contrast MCs that may be present
in the denser breast tissue [82].

Moreover, few studies accomplished the enhancement and the detection of MCs
using multiscale image Hessian [83], [84]. Li et al. [83] presented a preliminary work for
enhancement of MCs using Hessian based filtering. Nakayama et al. [84] computed
elements of the Hessian matrix using perfect reconstruction filter banks and demonstrated
that the proposed scheme preserved the shape of MCs and might achieve better detection of
MCs. Another approach that is proven efficient for segmenting and preserving the shape of
MCs is based on a graylevel morphological image processing such as image filtering using
top-hat and watershed operators [16]. Fu et al. [26] detected suspected MCs by first using a
top-hat graylevel morphological transform followed by an edge detection. The outputs from
Canny and Sobel edge detectors, applied to a filtered region, were combined to produce
suspicious MC regions.
Since the texture of the breast's background tissue is more self-similar than that of MCs, fractal modeling was employed in [85] to model the background breast tissue. Then, enhanced MCs were obtained by subtracting a modeled background from the original image. Moreover, fuzzy logic and meta-heuristic techniques were used for the detection of MCs in [86]-[87].

Feature extraction and classification via supervised learning machines have become essential components of recent CADe schemes, which have been employed to accomplish MC detection with reduced false positive results. This approach has been employed in several CADe systems [40], [41], [88]-[101]. In these CADe systems, shape and texture features are used to characterize image pixels or regions. These texture features are usually obtained from the first and the second order statistics of the graylevel histogram of an image, local statistics, and the spectral representation using wavelet and discrete cosine transforms. Then, extracted features or a subset of features are used as inputs of a supervised learning machine to classify input patterns into MC or background regions. Whether a given CADe system is a two-stage or a three-stage system, feature extraction and classification stages have been either employed to reduce the false positive results and to accomplish the final MC detection [26], [41],[93],[95] or to accomplish the entire MC detection task [88], [91], [96]-[97], [101].

Using wavelet coefficients and two statistical descriptors of the gray-level histogram, Yu and Guan [92] detected clustered MCs in 40 mammograms at 1 FP/image and 93% TP rates. In a successive study, Yu and Guan [93] improved the specificity of detection by extracting an additional 31 features from analyzing the shape of MCs and the texture of their region, which are used as a feature vector of a general regressive neural network. These
additional features led to a 0.5 FP/image and 90% detection's sensitivity compared to 1 FP/image and 93% levels from [92].

Veldkamp and Karssemeijer [94] detected MCs by using shape features and a k-nearest neighbor classifier. Zhang et al. [96] proposed a two-stage MC detection scheme in which they used texture features obtained from the gray-level histogram of MC and background regions as inputs to a neural network classifier to detect suspected MCs. In the second stage, the authors employed two shape features describing the entire MC cluster and a neural network classifier to improve the specificity of the detection.

Early in this decade, El-Naga et al. [97] proposed a successive enhancement learning method to achieve better learning of SVM classifier and obtained detection results with a better false positive rate. Fu et al. [26] also studied the impact of selection of the supervised learning machine on the performance of CADe and reported the superiority of the SVM over a neural network method.

Statistical modeling as a feature extractor has been used in [40]-[41], [88] to improve the false positive rate of the proposed MC detection. Caputo et al. [40] detected MCs by characterizing the histogram of mammographic regions using a statistical spin-glass Markov random field (SG-MRF) and optimal Bayesian classifier. Yu et al. [41] proposed a two-stage scheme for detecting MCs by which they used wavelet based filtering and global thresholding to identify suspicious MC regions. Then, they identified regions of true MCs by using texture features extracted using a statistical Markov random field (MRF) modeling and other image processing techniques, which have been used as inputs of Bayesian and back propagation neural network (BPNN) classifiers.
3.5.2 Computer aided diagnosis (CADx) of microcalcifications

Aiming at increasing the positive predictive value (PPV) of mammography, reducing the high volume of unnecessary and false invasive breast biopsies, and aiding radiologists in interpreting mammograms, researchers continue to develop computerized algorithms for characterizing and discriminating mammographic MCs into benign and malignant cases [10]. Following the radiologists’ methodology, CADx algorithms commonly model the diagnosis of MCs as a two-class or a binary pattern recognition problem accomplished using feature extraction and supervised learning machines.

Analyzing the shape of mammographic abnormalities in general and MC clusters in particular, which mimics the radiologist method in differentiating between malignant and benign MCs, is a significant approach in the diagnosis of mammographic lesions (mass and calcifications) [10],[67]. This approach has been applied by the earliest CADx systems[102]-[104] and continues to be a standard and important characterization technique in numerous studies [20]-[21],[27],[45]-[46],[105]-[110].

Wee et al. [102] characterized MCs using the standard deviation of the distances between boundary points and the cluster centroid. Sickles [103] demonstrated the role of the shape in differentiating between malignant and benign MCs. Magnin and others [104] analyzed the shape of MCs by measuring several regional descriptors such as area, eccentricity, compactness, and perimeter. In a seminal paper by Shen and his colleagues [20], three shape descriptors (compactness, normalized moments, and a Fourier descriptor) were developed and used to characterize the irregularity of the contour of mammographic microcalcifications. These proposed shape factors were evaluated using both simulated and 143 real microcalcifications. Classifying the test dataset using the three shape features and a
k-nearest neighbor (kNN) classifier achieved 100 % accuracy. Jiang et al. [105] used eight shape features and a neural network classifier to analyze the malignancy of 100 mammograms, which produced Az of 0.92. Using shape descriptors extracted and evaluated in [105], Wei et al [45] investigated the performance of several state-of-the-art supervised learning machines for classifying MCs as malignant and benign cases. Examining various classifier using 697 mammograms indicated that the set of kernel based classifiers (SVM, KFD, and RVM) outperformed a neural network approach, and the best classification result of Az of 0.85 was obtained using the SVM method.

Papdopoulos et al. [46] used 54 shape features to characterize MC clusters from MIAS and Nijmegen datasets. Using univariate feature ranking and a rule based expert system, a subset of 37 features was used as inputs to SVM and neural network classifiers. Results indicated the superiority of the SVM classifier that achieved Az of 0.79 and 0.81 using mammograms from MIAS and Nijmegen datasets, respectively. Wang et al. [110] presented a fully automated detection and diagnosis scheme that first employed CADe system from [98] to detect MCs. Then, each detected MC cluster is described using 34 shape descriptors. Rather than performing an explicit feature selection and a dimensionality reduction of the feature space, the authors transformed the original feature space using a principle component analysis (PCA) method. Then, GA was not only employed to select the best principle components and to reduce the dimensionality of the feature space but also to improve the generalization of the non-linear SVM classifier. Evaluating the proposed CADx using mammograms from the MIAS database produced Az of 0.86. Kallergi et al. [21] classified 100 MC clusters by extracting 14 shape descriptors used as inputs to a neural network. Using leave-one-out cross-validation, the proposed scheme achieved a classification sensitivity of 100 % and specificity of 85% that corresponded to ROC curve with Az of 0.98.
Moreover, in [111], the authors developed and applied a new band-pass filter, called “donut filter”, for detection and segmentation of MCs in digital mammograms.

Indeed, shape based diagnosis has proven to be very effective for discriminating a malignant from benign MC cluster and remains the most popular and the closest to the radiologist’s approach, which [20],[112]. A key process for analyzing clustered microcalcifications using their shape is the segmentation stage, which remains a challenging and an unsolved problem [10]. An alternative and promising method for characterizing MCs is by analyzing the texture of mammographic regions using different textural and statistical techniques [23], [27], which override the need for a prior segmentation of MCs. Another advantage of texture based diagnosis is its ability to characterize texture dependency and spectral properties, which are invisible to human eyes or cannot be described using shape measures. Commonly used texture features are Haralick measures of texture. These features analyze the second order statistics of the gray-level histogram of the selected region and are usually derived using gray-level co-occurrence (GLCM) or spatial graylevel dependence matrices.

Dhawan et al. [113] used textural based analysis of mammographic regions to characterize hard-to-diagnose clustered MCs. Extracted texture features included first order statistics of the gray-level histogram of MCs, number of MCs and their distribution in a cluster, second order statistics using the GLCM method, and spectral features from wavelet decomposition. Various feature subsets selected using a global heuristic GA search and multivariate clustering analysis methods were used for classification using a back-propagation neural network (BPNN) classifier, a k-nearest neighbor classifier, and a parametric statistical Bayesian classifier. The authors tested their methods using 191 MC clusters, which indicated the superior performance of BPNN.
Chan et al. [27] classified MCs using texture features derived using GLCM matrices and shape features. Comparing the performance of feature selection using GAs and linear discriminate analysis (LDA) methods indicated the effectiveness of GAs. Classifying 145 MC clusters using combined shape and texture features achieved $A_z$ of 0.89 that was better than 0.84 and 0.79 obtained from texture and shape features, respectively.

Zadeh et al. [107] demonstrated the superior classification performance of the shape features over texture features computed using the gray level co-occurrence matrices (GLCM) technique. Features selected through binary and real Genetic algorithms were employed to classify MC clusters using the k-nearest neighbor classifier, which produced an area under ROC curve ($A_z$) of 0.82 from shape features compared to 0.72 from texture features. In a more detailed study [23], Zadeh et al. evaluated the classification performance of spectral features derived from a multiscale analysis of mammographic regions using wavelet packets and multi-wavelet transforms, contrast and shape features, and statistical features derived using graylevel co-occurrence matrices. Results of this study indicated that texture features derived using multi-wavelet transforms produced the best area under the ROC curve of 0.89, which not only outperformed features from wavelet packets and the GLCM method but also they produced calcification’s performance that was even better than shape features.

Zadeh et al. [23] also indicated that including the background texture when extracting GLCM features of MCs produced better results than only characterizing the texture of the regions representing individual MCs. However, other studies [112], [114] demonstrated that it is not the texture of MCs objects but it is the texture of breast tissue surrounding MCs that can be useful for cancer diagnosis. Thiele et al.[114] classified 54
MC clusters by extracting texture and fractal features of the region surrounding each cluster and reported a classification sensitivity of 89% and specificity of 83%.

Karahaliou et al. [112] diagnosed clustered MCs by analyzing the surrounding texture by excluding image locations that correspond to individual MCs. The residual texture was analyzed using four techniques: first order statistics of the gray-level histogram, Laws’ measures of texture, second order statistics of the gray-level histogram using GLCM method, and the run length statistics of the gray-levels. Different sets of texture features were used to classify mammographic regions into a region of malignant MCs and benign one using the kNN method. Comparing the performance of various feature extraction methods indicated the superiority of texture features extracted using Laws’ texture energy measure techniques that produced classification of 89% accuracy compared to 82% accuracy from the GLCM method, which was also better than other techniques. The promising results obtained in this study suggested that analyzing the texture of tissue surrounding MCs can be very useful for computer aided diagnosis of breast cancer and might provide a diagnosis method that can avoid segmentation of MCs.

3.5.3 Summary and conclusions

Computer aided detection has been thoroughly studied in the past twenty years, and several CADe systems have received FDA approval, and they are clinically in use. Examples of these systems are: Image Checker (R2 technology, USA), Second Look (iCAD, Canada), and Kodak’s System(Kodak, USA). However, the role of CADe remains questionable. Many observer studies have examined the efficacy of CADe technology and its role in screening the breast cancer, some studies demonstrated that CADe technology can improve the
sensitivity of the detection of breast cancer by 10% [2] without a significant increase of the recall rate [3]-[4]. In addition, they reported that the improvement of the performance using CADe is close to the detection gain from a double reading stage [2]. On the other hand, other studies have reported no improvement is achieved by using CADe. Additionally, a recent study by Boyer et al. reported that current CADe technology cannot replace a second reader but it can be beneficial if employed as a supplement stage [11]. Moreover, some CADe schemes have achieved a high detection sensitivity at expense of low specificity that may increase the number of false biopsies. Therefore, future research work needs to continue investigating CADe to improve the performance of existing technology, and to explore new detection methods.

Although the efforts of developing computer aided diagnosis (CADx) schemes have been started earlier than CADe technology and the results of observer studies have shown relatively higher confidence in the positive impact of using CADx on radiologists’ performance [10], [106], CADx technology has not been commercially realized. In addition, the overall performance of existing CADx methods including the robustness of the feature extraction and selection methods, and the specificity level of the classification stage has not met the radiologists’ expectation. Hence, future CADx research needs to answer several research demands including the development of efficient and robust feature extraction techniques, automatic feature selection methods, and classification schemes with good generalization’s ability and high specificity and sensitivity levels.

Moreover, a major shortcoming of the CAD community is the lack of availability of common and large online digital mammography datasets for evaluating newly developed approaches and for comparing them with existing CAD algorithms. Meanwhile, some screening mammography datasets, in digital format, are available online at no cost [64].
This dissertation selects the dataset provided by Mammographic Image Analysis Society (MIAS) [115], which has been used by many researchers.

3.6 MIAS mammogram dataset

CAD algorithms, presented in this dissertation, will be tested using a dataset of 23 screen film mammograms from a mini MIAS database [115]. This mini database composed of 322 MLO screen film mammograms represent left and right breasts of 161 women. These mammograms are of size 1024×1024 pixels with a spatial resolution of 200µm/pixel, which are sub-sampled of higher resolution mammograms that have been digitized at 50µm/pixel.

MIAS dataset contains 20 mammograms with 25 limited spread MC clusters of which 13 are benign cases and 12 are malignant ones, and three mammograms with islands of malignant calcifications spread over the entire breast gland. Additional 8 malignant clusters, confirmed by an expert radiologist, were extracted from three mammograms with islands of MCs, which leads to a total of 33 MC clusters that are shown in Figures 3.5, 3.6, and 3.7.

Moreover, each mammogram from the MIAS database is annotated with radiologists' interpretations that include a description of breast tissue (fatty, fatty-glandular, and dense-glandular), type of the pathology (malignant or benign) of breast lesion. Additionally, a ground truth file that accompanies each image provides information about the location of each abnormality in the image, the centroid and the size of the region the best encloses each lesion. This ground truth file is utilized in this work, as radiologist's input, to
extract a region of $128 \times 128$ pixels centered at each cluster's centroid and to design an efficient segmentation and feature extraction scheme.

![Images of malignant microcalcification clusters](image)

**Figure 3.5:** Malignant microcalcification clusters from the MIAS database.
Figure 3.6: Benign microcalcification clusters from the MIAS database.
Figure 3.7: Additional eight microcalcification clusters extracted from three MIAS mammograms with calcifications spread over the entire breast region.
CHAPTER IV

BAYESIAN CLASSIFIER WITH SIMPLIFIED LEARNING PHASE FOR DETECTING MICROCALCIFICATIONS IN DIGITAL MAMMOGRAMS

4.1 Introduction

Individual calcifications, including microcalcifications and macrocalcifications, are tiny calcium deposits. Mammographically, calcifications appear as small bright spots, which greatly vary in sizes and shapes. When calcifications are surrounded by dense breast tissue, the detection and segmentation of calcifications become very difficult. Compared to other types of breast structures, calcifications are commonly modeled as impulse-like structures, which can be detected by searching for highpass anomalies in the digital image. Since calcifications are generally shown on a mammogram as bright spots, it is more convenient to apply a set of image processing techniques that is proven efficient for handling a point singularity in the digital image. Examples of calcification detection and segmentation tools are spatial filtering using Laplacian of Gaussian and difference of Gaussian kernels, morphological operators, and wavelet transforms.

Among these approaches, multiscale image analysis using wavelet theory is the most effective approach, which been used for enhancement [80], segmentation and detection [41], [50], [77], [78], [101], and for characterizing the malignancy of mammographic microcalcifications [23], [27]. Furthermore, the gray-levels of the healthy breast tissue and
microcalcifications as well as their wavelet representations have been demonstrated to have Gaussian-like distributions [77], [82]. These statistical traits of mammographic MCs motivated several researchers to employ local statistics using measures like skewness and kurtosis measures [50],[75], probabilistic modeling using Markov random field [41], and a parametric classifier such as Bayesian learning to distinguish microcalcifications in digital mammogram.

The detection and segmentation of calcifications using a filtering stage only is very sensitive to the selection of the threshold, which is usually used to generate a binary representation of the segmented image. To overcome this problem, several studies have used a second stage that is a supervised learning machine (i.e. feature extraction and binary classification stages) to reduce the false positive signals, which might result from using low threshold levels mostly intended to produce a high detection sensitivity at price of a low specificity.

A supervised learning machine in general and a Bayesian classifier in particular, require pre-labeled training examples to be extracted and used for estimating a learning model. Opposed to this approach, we present a method that estimates the parameters of the learning model (i.e. the mean vectors and covariance matrices for Bayesian learning) by constructing synthetic training patterns of the class representing calcifications and by using these synthetic patterns for accomplishing the Bayesian learning.

This new framework is a single stage detection scheme consisting of two phases, namely, feature extraction and feature classification. In the feature extraction stage, each image pixel is represented using four features: a gray-level value or intensity, point discontinuity from a spatial filtering, response from a wavelet based filtering, and local statistics of each pixel estimated by measuring the tail-ratio of a gray level histogram.
When dealing with the whole mammogram, image patterns like the boundary of a radiographic marker and curvilinear structures of the borders of the breast region are expected to produce many false positive signals. To reduce these false positive signals without affecting true positive results, a post-processing step involves a gray-level thresholding using Otsu's technique is used to eliminate false signals detected outside a glandular breast region.

The remaining sections of this chapter are organized as follows: Section 4.2 presents the details of the proposed detection method while the Experimental results and Conclusions are presented in Sections 4.3 and 4.4, respectively.

4.2 Segmentation using simplified learning Bayesian classifier

Learning machines for pattern recognition, such as artificial neural network, support vector machines and maximum a posteriori probability (MAP) classifiers consist of two phases: supervised learning and testing phases [40]-[41], [88], [90], [97], [100]. In the learning phase, a group of training samples that represents different objects or patterns to be extracted are selected manually to optimize the classifier's decision function while in the testing phase, the trained classifier is used to classify features contained in new data sets or the independent samples.

Our proposed classification approach, Figure 4.1, follows the general structure of the classical learning machines but it uses a simplified learning stage denoted here as self-learning phase. Such a process can be relatively described as an unsupervised learning since it does not require the huge number of training samples of MCs to be extracted in advance.
from different data mammograms as the case of classical supervised learning [12], [90], [40]-[41] and instead it synthesizes these samples and use them as training data.

Figure 4.1: Segmentation using Bayesian learning.

In this work, detection is modeled as a two-class pattern recognition problem where the first class, $\omega_1$, is the clustered microcalcifications group and the second class, $\omega_2$, is the healthy breast tissue. The proposed approach is described as follows:

- **Modeling of microcalcifications**: the training samples of MCs class are synthesized by blending a synthetic model of MCs with a mammogram image. More details of this process will be explained in Section 4.2.1.

- **Feature extraction**: linear and none-linear transforms are used to extract three features of each pixel of a mammogram image. These three feature images along with the graylevel mammogram image are registered spatially to form a 4-D pattern vector $x=[x_1, x_2, x_3, x_4]^T$ of each class $\omega_j$ as shown in Figure 4.2.
In Figure 4.2, each pattern vector $\mathbf{x}$ is represented by a set of four components described as follows,

- $x_1$: Graylevel or image intensity.
- $x_2$: Local maxima ranked using local histogram.
- $x_3$: Spectral feature extracted using wavelet transform.
- $x_4$: Singularity detection by detecting point discontinuity.

- **Learning phase**: The proposed learning process estimates the classifier's decision function parameters of each input mammogram, see Figure 4.1. Unlike the classical method which collects the training sets from different mammograms, the proposed approach extracts the training samples of different classes from the input mammogram itself as follows: for MCs class, it models the MCs and creates synthetic training samples of MCs class for that mammogram, the locations of these samples are identified using the binary model of the synthetic MCs. For the healthy breast tissue
class, the training data are collected from two ROIs chosen randomly within the breast region.

- **Parameters estimation**: pattern recognition using stochastic BC is based on the estimation of the probability density function of each class. Assuming that the measured features of each class have a Gaussian probability distribution, the classifier's decision function can be computed as given in equation (4) requiring the estimation of the covariance matrix and the mean vector of each class. If the training set of each class is a sufficient statistically, one can efficiently estimate the distribution parameters (i.e. covariance matrix and mean vector) of each class.

  Further discussion of the parameter estimation for Bayesian classifier used in this work is presented in Section 2.5.1.

- **Bayesian classification**: the optimized classifier is applied to perform a pixel based classification of the breast region into microcalcification and healthy tissue. In this work, the classification results are binary 0 or 1 and they are used to create a binary image by assigning a binary 1 to pixels classified as class $\omega_1$ (or MCs), while a binary 0 is assigned to pixels classified as class $\omega_2$ (or healthy breast tissue).

- **Post processing**: the purpose of this step is to reduce the false classifications and to improve the classification results through the integration of some of the physiological traits of breast tissue and clustered microcalcifications.

4.2.1 Construction of the synthetic microcalcifications

In this work, we use a method proposed in [50] as an attempt to generate a model for real MCs as illustrated in Figure 4.3. In this method, a new MC model is derived from the
standard model (StdModel), a binary model of synthetic MCs, using input image and the modeling constant $K$. That is, each gray level value from synthetic pixels is assigned initially a fraction that is proportional to the constant $K$ of its corresponding mammogram pixel and through a blending process, a hybrid image is created from the original mammogram and the modified MCs model. Such process is a pixel by pixel addition of the MCs model and mammogram followed by smoothing of the synthetic pixels using lowpass filter $H$, an example of the outcome of this scheme is shown in Figure 4.4. Our experimental results indicate that $K$ should be chosen based on the statistics of the breast tissue of the mammogram such as the mean and variance of the breast tissue intensity values.

This method of synthesis has been introduced and employed in [50], however, this proposed work has a significant difference from [50]. This method has an explicit use of a modeling constant to control synthesizing different types of MCs so that the synthetic MCs impersonate real MC as much as possible. It is also worth noting that the purpose of using synthetic MCs in [50] was to provide a testing material for the detection scheme [50], while it is employed in this work as a detection tool and a control parameter of the scheme.

4.2.2 Feature extraction and formation of a pattern vector

This work uses the general structure of pattern recognition using Bayesian classifier which stacks and spatially registers a group of feature images. Each mammogram is represented by a stack of four images; 1) gray-level feature from original image, 2) feature image extracted using local maxima ranked using their local histogram, 3) highpass filtered image extracted using discrete wavelet transform, and 4) point singularity detected using Euclidian distance $ED_8$. 
Figure 4.3: Construction of synthetic MCs.

Figure 4.4: Synthesizing clustered microcalcifications. (a) Original mammogram, (b) mammogram with synthetic microcalcifications. Red circles correspond to the training samples of the microcalcifications while yellow circles represent the training samples for the healthy breast tissue.
4.2.2.1 Highpass filtering using 2-D wavelet transforms

Highpass filtering using discrete wavelet transform has proved to be a useful tool for detecting suspicious MCs [41], [50], [78]. In [78], the authors reported that orthogonal wavelet filters such db4 are more appropriate of detecting MCs since they have higher sensitivity to the presence of microcalcifications than other wavelet filters. Also, the spike-like behavior of db4 wavelet transform justifies the successful use of this wavelet filter for detecting specious MCs in [41], [50]. Therefore, we decided to employ db4 to extract the spectral features of MCs and to use this feature as one input feature of the Bayesian classifier.

The basic idea behind this analysis is that, MCs represent highpass anomalies lay on a stationary lowpass background contributing to the detail subbands rather than to the coarse scale subbands of the wavelet multiresolution representation. In [78], the authors demonstrated that the features of MCs can be made more obvious after suppressing the background data, which is accomplished by eliminating the wavelet coefficients within coarse scale subband and reconstructing an image from detail subbands. An example of this process is illustrated in Figure 4.5.

4.2.2.2 Feature extraction using point discontinuity

Spatially, microcalcifications appear as bright spots with various and mostly irregular shapes. Microcalcifications also appear in intensities that are higher than that of the surrounding healthy tissue. Therefore, a pixel belong to a microcalcification region is expected to experience a larger gray-level difference from its local neighborhood than that of a healthy one. One approach to extract this type of singularity is by employing a point
detection kernel as shown in Figure 4.6. In this work, the point singularity feature $ED_8$ of each pixel is defined as the sum of the absolute difference of a pixel graylevel and those of its 8-neighbours. Example of this feature and other extracted features is demonstrated in Figure 4.7.

(a) ![Figure 4.6: Microcalcifications' point singularity analysis. (a) Point detection kernel and (b) 3×3 block centered around pixel $I(i,j)$.](image)
(b)
\[ ED_{\varepsilon}(i,j) = \sum_{k_2} \sum_{k_1} |I_{i+k_2,j+k_1} - I_{i,j}| \quad k_1 = -1,0,1 \quad \text{and} \quad k_2 = -1,0,1 \] (4.1)

Figure 4.7: Feature extraction of microcalcifications. (a) Original image, (b) texture features extracted using point discontinuity, and (c) spectral features extracted using wavelet based highpass filtering of image shown in (a).

4.2.3 Learning phase using synthetic microcalcifications

Pattern recognition methods are in general supervised leaning machines [12], [41], [88], [90], [97], [100], they partition the population data into training and validation sets. In such approaches, the training samples which are usually labeled manually, are employed to estimate the parameters of the classifier’s decision function [40]-[41]. Our proposed approach can be considered an unsupervised method and thus it does not require the training set of MCs class to be extracted from real mammograms as in the supervised manner. Instead, an adaptive and simple learning scheme is employed to estimate the classifier’s parameters. The advantage of the proposed training scheme over the classical one is the use of synthetic MCs as training samples for the MCs class rather than using real MCs extracted from mammograms as practiced in the supervised methods [41].
Training stage starts by extracting four features from the breast tissue; these feature images are stacked to form a multidimensional feature vector of each pixel. The learning phase of the classifier is accomplished by blind, or unsupervised, selection of training data. Such selection is done by employing the binary MCs model to identify the training samples for MCs region while two distinct regions randomly selected within the mammogram are used to locate the healthy breast tissue candidates. A drawback of this random selection of the training samples of healthy breast tissue is the possibility that these regions may lay over breast areas that have low probability of developing malignant microcalcifications such as fatty or background regions. The negative impact of this practice can be eliminated by having a preprocessing step in which, the user mark the two regions within the glandular breast area or by employing a preprocessing step to identify the glandular breast region.

The learning process we propose has many advantages over the classical one; first, is the simplicity of the process with respect to the size of learning data, second, training samples of all classes, including the MCs, were selected manually in [12],[40]-[41] while the training set of the significant class, which is the MCs, is synthetically constructed in this work. Another advantage of the proposed learning phase is that the training process of the classifier is adaptive to the breast tissue as the parameters of the classifier's decisions function are estimated using self-learning method based on the input data.

The proposed learning phase has two challenges; the first challenge occurs when the synthetic training samples are not statistically sufficient which may produce underestimation of the classifier’s parameters. This limitation can be mostly attributed to the simplicity of the proposed modeling itself, that may have add some constraints on the ability to generate MCs training set of sufficient statistics from a single mammogram. The other challenge occurs
when regions representing the training samples of healthy tissue include members of the other class (real MCs).

While this work has not investigated the first problem, left it out as future work, the second challenge was addressed by using relatively large number of training samples of healthy (or background) class extracted from two different mammogram regions. The differencing in the sample size is significant due to the fact that mammogram texture is none-stationary and many samples of non-MCs class are available compared to the number of samples representing MCs that estimated to be no more than 1% of the whole mammogram. This work used about 4300 samples to represent the healthy (or background) class obtained from two distinct regions, which is about 50 times the size sample of MCs class. We investigated the effect of the sample size on the performance of the proposed detection scheme and the results indicated that a better detection can be obtained when two different regions used to extract the training samples of the healthy class than a single region. The results also indicated that the sample size of the healthy class must be larger, three times or more, than that of MCs class for better detection performance.

4.2.4 Parameter estimation of Bayesian classifier

Assuming the two classes are equally likely and using the training pattern of both classes, our feature vectors, the decision function of the classifier is constructed by approximating the mean vector and covariance matrix [12] for each class as given by equation (2.39). The modeling constant, \( K \), plays a significant role as it controls the appearance of synthetic MCs and their blending with the surrounding breast tissue.
One approach that might be useful for selecting an appropriate value of the modeling constant prior to the training and classification stages is by measuring the difference between the corresponding components of the estimated mean vectors and the ratio of the corresponding diagonal entries (feature variances) of the estimated covariance matrices. Our investigation of both measures concluded that interpreting the mean difference is more obvious, that is easier to make a conclusion, and can be employed for better detection results.

Analyzing the inter-class mean difference leads to identifying two cases; the first case occurs when a large value of the modeling constant is used, one that produces a large mean difference and leads to detect a single tone detail of the image, which might fail to detect the targeted MCs. This problem can be eliminated by adjusting the modeling constant to lower values before proceeding with training and segmentation stages. The second case occurs when a very small mean difference is used that decreases the discrimination power between classes and leads to an increase in false signals.

4.2.5 Segmentation via Bayesian classifier

Testing the discrimination power of the classifier is usually accomplished by using the decision functions of the BC, equation (4), which is computed using the estimated covariance matrix and mean vector to classify an independent set of samples followed by computing the misclassification rate. The segmentation results are interpreted, from the classification results, as a target (microcalcification) and represented by a binary 1 and a none-target (healthy tissue) represented by a binary 0. Both classes are assumed to be equally likely to occur.
4.3 Experimental results

4.3.1 Mammogram test data

The proposed detection scheme is investigated using 23 mammograms from MIAS database [115], which includes 28 microcalcification clusters. Image annotations accompanied with each mammogram, which includes a ground truth of each MC cluster (i.e. the location and size) and the type of breast tissue is found to be very useful when assessing experimental results presented in this section.

4.3.2 Simulation methods and parameter settings

The proposed scheme starts by modeling of the MCs in each mammogram as explained in section 4.2. The most significant step of this process is the selection of the modeling constant $K$, a typical value of $K$ can be chosen between 0.1 and 1. Then, spatial, textural, and spectral features of all pixels are extracted and used as inputs to the Bayesian classifier. The feature vector of each pixel is composed of the following: 1. Brightness (or graylevel), 2. Local maxima ranked based on the tail ratio of their local histogram estimated within 9×9 neighborhoods, 3. Highpass filtered image obtained from suppressing coarse (or approximates) of the 2-level wavelet representation (db4 filters were used) and reconstructing an image from detail subbands, and 4. Point singularity values computed using equation (4.1) as presented in section 4.2.2. Stochastic Bayesian classifier optimized by a simplified self-learning phase is used to segment (or classify) all image pixels into MCs or healthy ones.
Experimental results demonstrate that applying this proposed approach to the whole mammogram without extraction or prior knowledge of breast region produces more false positive signals than those resulting from using breast region extracted from the whole mammogram. This result is illustrated in Figures 4.8.a and 4.8.c which also indicates that these false signals are mainly localized outside the breast region and can be significantly reduced using Otsu’s thresholding [13]. Examples of this step results are demonstrated in Figure 4.8.b and 4.8.d. In order to test the abilities of the proposed scheme for segmenting the whole mammogram and detecting the MCs, we used a simple thresholding scheme, Otsu’s method, as a postprocessing instead of employing a prior breast region extraction or using some regional context within the detection process. Another advantage of using Otsu’s thresholding as postprocessing was eliminating all misclassifications occurring along the breast border and outside the breast region. Such suppression process improved the detection performance by significantly reducing the overall number of false results, or misclassifications, while maintaining the detected MCs.

4.3.3 Experimental results analysis

Experimental results are assessed by computing the specificity and sensitivity parameters. This assessment would have been much more challenging without having the location and the size of true, real, MCs as documented by MIAS database [115].

Results indicate that synthesizing the training samples of MCs class and specifically the selection of the modeling constant $K$ plays a significant role in the performance of the proposed classifier and the detection results. They also show that $K$ values should be chosen based on statistics of the breast tissue characterized by the separation between the brightness
of the region of MCs and that of the background tissue and the variance of the breast region. That is, the optimal $K$ value is mostly correlated to the normalized mean difference (NMD) computed from the difference of the average graylevel (brightness) of the MCs and that of the of breast region.

Analyzing the breast regions of size $256 \times 256$ pixels extracted from each group indicates that a dense-glandular breast tissue has a larger intensity mean and variance than those of fatty breast tissue. Moreover, MCs that may be present in a fatty breast tissue have a larger NMD than those of MCs in dense-glandular breast tissue. Our results show that the modeling constant $K$ can be adaptively chosen based on the type of breast tissue and the statistics of the breast region. Therefore, from all experimental simulations on mammogram ROIs of size $256 \times 256$ pixels, we found that small values of $K$ such as 0.1-0.4 are suitable for detecting MCs occur within dense-glandular breast tissue while larger values of $K$ such as 0.5-1 are more appropriate for MCs in fatty breast tissue. Such results are mostly due the fact that MCs present in denser mammogram tissue have lower local contrast than MCs occurring in fatty breast tissue. These results need to be further investigated on larger set of mammograms. Figure 4.9 illustrates the effect of the parameter $K$ on the classification results, which shows that a large value of $K$ ($K > 0.2$) produces a detection of MCs with high specificity while low value of $K$ as 0.1 or less leads to detection results with many false signals (or low specificity) as shown in Figure 4.9.c. On the other hand, large values of $K$ are more appropriate (lower false signals) for detecting MCs appearing in a region that has high NMD and local variance as shown in Figure 4.10.

Results show that no optimal $K$ value produces the best detection results (lowest FP and FN) for all test data but some values such as $K=0.2$ produces the best TP and FP rates. Moreover, modeling MCs in dense breast tissue has shown more sensitivity to the value of $K$. 
Figure 4.8: Improving the FP rate of the detection results using Otsu’s thresholding. (a) and (c) are the results without postprocessing (b) and (d) are the outcomes of post processing using Otsu’s thresholding.
while the algorithm was more robust and allowed K to span a wider range for fatty breast tissue. Examples of these results are presented in Figure 4.11. Figures 4.11.c and f show that low values of K between 0.2-0.25 are suitable for detecting MCs within both fatty and dense breast tissue. Furthermore, results indicated that MCs within Fatty breast tissue can be modeled and detected using wider range of modeling values with lower FP results at larger values of K, as demonstrated in Figures 4.10.b, 4.11.h, and 4.11.i. We believe that detection results can by further improved by fine tuning the selection of K if the breast region statistics were integrated into the algorithm.

Figure 4.9: Detection results of MCs in a dense mammogram (mdh223) using different modeling constant K. (a) Original mammogram (ROI of size 256 × 256 pixels, NMD =0.05) (b) K=0.25 (c) K=0.1.

4.3.4 Performance evaluation and comparison

In fact microcalcifications occur in mammogram in form of clusters rather than standalone. According to [64], a cluster of microcalcifications is defined as a group of three or more classification within a 1 cm² area which is equivalent to a block of 50×50 pixels in
mini MIAS mammograms (digitized at 200\(\mu\)m edge resolution) [41]. Malignant calcifications also can only present in glandular breast tissue which is a fact that can also be used to eliminate any candidates (segmented as MCs) detected within dark regions. These physiological features are integrated in this work and mainly used in the computation of FP rates. Moreover, the difficulty in counting precisely the number of real calcifications within the region of true MCs forced us to count TP signal and report the sensitivity rates in a method similar to the one used in [74]. One reason for choosing TP per mammogram rather than per cluster is the nature of this proposed detection scheme, which applies a single model for segmenting all calcifications within the tested mammogram.

Before proceeding any further with the performance evaluation, the definitions of TP, FP, FN and TN as used in this chapter ought to be stated:

Figure 4.10: Detection results of MCs in fatty mammogram (mdb209) using different \(K\). (a) Original mammogram (ROI of size 256 \times 256 pixels, NMD=0.16) (b) \(K=1\) (c) \(K=0.1\).
Figure 4.11: Examples of the detection results using different $K$ values.
• TP is identified by visual inspection of the detection results at image locations corresponds to the real annotated MCs region per the mini-MIAS database.

• FN is identified when a mammogram region of size 50×50 pixels that belongs to real annotated MCs per the database is detected as a background class.

• FP is identified when a healthy or a background region of size 50 × 50 pixels included three or more image locations detected as MCs class.

• TN is identified when a healthy or background region of size 50 × 50 pixels is detected as background class or it included a maximum of two locations of isolated MCs.

Using these definitions, the total number of TN, FP of a given output binary image is calculated by dividing the segmented image into none-overlapping regions of size 50 × 50 pixels excluding the region of the real (actual) MCs as identified and labeled by the database.

Table 4.1
Detection results using self-learning BC

<table>
<thead>
<tr>
<th>K</th>
<th>0.165</th>
<th>0.2</th>
<th>0.25</th>
<th>0.33</th>
<th>0.5</th>
<th>0.75</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specificity %</td>
<td>96.9</td>
<td>98.6</td>
<td>99.2</td>
<td>99.2</td>
<td>98.9</td>
<td>98.4</td>
<td>98.1</td>
</tr>
<tr>
<td>Sensitivity %</td>
<td>95.7</td>
<td>91.3</td>
<td>78.3</td>
<td>69.5</td>
<td>56.5</td>
<td>52.2</td>
<td>56.5</td>
</tr>
</tbody>
</table>
Table 4.1 demonstrates the average specificity and sensitivity of the detection scheme obtained using BC optimized via self-learning methods. Per Table 1, the best sensitivity (or TP rate) is about 91.3% at \( K=0.2 \) and the corresponding specificity (TN rate) is at 98.6%.

Although this proposed approach performs segmentation on a mammogram on a pixel level, it does estimate FP rates using region basis by utilizing physiological characteristics of clustered MCs, which is different from the previous work reported in [41]. This fact makes it unreasonable to attempt to have a direct numerical comparison between our detection results and those obtained in [41], [100]. However, in [41], the authors reported their results using \( 87 \times 87 \) block sizes and used the total number of the true MC samples to be 25 MC regions. To ensure unbiased comparison with results reported in the literature, we decided to evaluate the performance of the detection results from this study in a similar manner.

Table 4.2

<table>
<thead>
<tr>
<th>Study</th>
<th>Database</th>
<th>Sensitivity %</th>
<th>Specificity %</th>
<th>FP/image</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed I*</td>
<td>MIAS</td>
<td>91.3</td>
<td>98.6</td>
<td>6.15</td>
</tr>
<tr>
<td>Proposed II *</td>
<td>MIAS</td>
<td>91.3</td>
<td>96.4</td>
<td>5.1</td>
</tr>
<tr>
<td>Yu et al.[41] (BC)</td>
<td>MIAS</td>
<td>92</td>
<td>97.8</td>
<td>0.75</td>
</tr>
<tr>
<td>Yu et al.[41] (BPNN)</td>
<td>MIAS</td>
<td>92</td>
<td>98.9</td>
<td>1.5</td>
</tr>
<tr>
<td>Huang et al.<a href="SVM">100</a></td>
<td>MIAS</td>
<td>76</td>
<td>88</td>
<td>NA</td>
</tr>
<tr>
<td>Huang et al.[100] (BPNN)</td>
<td>MIAS</td>
<td>72.15</td>
<td>78.4</td>
<td>NA</td>
</tr>
</tbody>
</table>

* Proposed I and II obtained using region size of \( 50 \times 50 \) and \( 87 \times 87 \), respectively.
We used the same block size and counted FP to be the case when an MC is detected. However, TP is still evaluated by visually inspecting the detection results and comparing it with the real MCs as reported by the database. Table 2 compared the specificity and the sensitivity or (TP and FP rates) of the proposed scheme with relevant works, which indicates that this proposed scheme produces lower TP and higher FP rates (or lower specificity) compared to those from [41] but better than [100].

4.4 Discussion and conclusions

In this chapter, we proposed and implemented a new approach using stochastic Bayesian classifier for segmenting a digital mammogram for the detection of microcalcification clusters. The proposed scheme models the image segmentation task as a two-class pattern recognition problem. This new framework accomplishes the learning phase of the classifier using a simple self-learning approach which synthesizes the training samples of MCs class in each mammogram. Each image pixel, during both the learning and testing phases, is modeled using a four-feature vector extracted using spatial, statistical, and spectral via wavelet filtering methods. The proposed scheme was tested using 23 mammograms from mini-MIAS database. Results demonstrated that synthetic patterns can be employed to simplify the supervised Bayesian learning for MCs detection, which produces moderate detection performance. The relatively high FP and low TP rates can be related to the simplicity of MCs modeling used in this study as well as applying the to the whole mammogram detection scheme as opposed to regions continuing only breast tissue.
CHAPTER V

MORPHOLOGY BASED DIAGNOSIS OF CLUSTERED MICROCALCIFICATIONS USING PSO-SVM FULL MODEL SELECTION

5.1 Introduction

The main goal of the development of a computer aided diagnosis (CADx) in mammography is to help radiologists in differentiating between malignant and benign breast abnormalities. A typical shape based CADx system, as explained in Chapter III, automates the diagnosis of a mammographic lesion (microcalcification cluster in this work) through four steps. The first step involves extraction of mammographic regions enclosing microcalcification cluster and segmentation of MCs. The second step would consist of characterizing the segmented MCs using their morphology or shape such as the work of [20]-[21],[23],[27],[45]-[46],[99],[105],[107]-[108] or analyzing texture of mammographic regions such as the work of [23], [27], [107], [112]-[113].

Feature extraction process may produce redundant or inadequate features and could produce a complex feature space and poor discrimination among its different patterns. Hence, a feature selection process, which is usually placed as a third stage to select a small subset of features that are more discriminating. Previous studies have selected best feature subset and achieved dimensionality reduction of the original feature space using exhaustive heuristic search methods such as Genetic algorithms (GAs) [23],[27],[107],[113], mathematical analysis such as linear discriminate analysis [27], sequential forward selection
method [99], [108] and dimensionality reduction using principle components analysis [110]. Other studies also presented a semi-automated feature selection method that is based on eliminating weak features using their univariate ranking method and using a rule based expert system to search for additional discriminative features [46].

In the last step, a reduced or original feature space was used to classify the MCs into benign and malignant classes. Diagnosis of MCs commonly modeled as a binary classification problem accomplished using supervised learning machine. The most popular classifiers used in previous CADx schemes are artificial neural network (ANN) [21], [45]-[46], [105], [108],[113], k-nearest neighbor (kNN) [23],[107], [112]-[113], and the state of the art kernel based SVM [45],[46],[109],[110]. Other learning machines also used in previous studies are statistical Bayesian [113], linear discriminate analysis (LDA) [20], [27], kernel fisher discriminate (KFD), relevance vector machine (RVM), and ensemble methods [45].

CADx systems that combined shape-based feature extraction and kernel-based support vector machine (SVM) have proven to be more effective than the popular ANN. Several studies have demonstrated this result by applying both SVM and ANN learning machine to classify MCs within the same experiments (i.e. similar extracted features and mammograms) [45]-[46]. However, previous SVM based MCs diagnosis methods have several shortcomings and limitations, which include employing semi-automated techniques to perform segmentation [45] of the individual MCs and feature selection [46]. The performance of SVM classifier was optimized using conventional grid search selection [46], k-fold cross validation [109] and exhaustive and computationally expansive heuristic search method using GA [110]. In addition, shape features extracted in some studies [45]-[46] were limited to the geometrical (e.g. region and distribution) descriptors and have not included
other mathematical boundary descriptors such as normalized shape moment and Fourier descriptor. Even though, mathematical descriptors have demonstrated to be very effective in discriminating benign and malignant MCs [21].

In the first stage, in our four-stage shape-based CADx scheme, I segment the individual MCs using a morphological filtering scheme with dual filtering scales. Each scale employs a modified top-hat morphological operator, an original top-hat with additional morphological closing operation before subtraction step, followed by a local thresholding process. In the second stage, morphology of MCs is used to describe each cluster and to extract 34 shape descriptors such as measures of the region (e.g. area, compactness, eccentricity, extent), distribution, and the shape boundary of individual MCs. Moreover, shape of the entire MC cluster is used to produce additional 10 shape features such as area, circularity, normalized shape moments, and Fourier descriptor. This feature extraction process leads to 44 shape features per cluster. A heuristic full model selection, or more specifically an embedded feature selection, using a PSO search method is mainly intended to integrate both processes of feature selection and the SVM classifier’s model selection.

Moreover, in this chapter I compare between two methods to achieve the feature search process. The first method is based on heuristic search using binary PSO technique to find an optimal feature subset while the second method constructs a search space and feature subsets using an outweighed univariate based nested subsets method.

In this work, I develop a new framework using PSO-SVM schemes to automate feature selection and to optimize the generalization performance of the SVM classifier. Moreover, this is the first attempt to search for features using heuristic binary PSO and nested subsets methods in MCs diagnosis via supervised learning machine. I have decided
to use heuristic PSO parameter search method since it has proven to be very competitive when compared to heuristic search GAs approaches used in previous CADx [23], [27], [110]. I further investigate the relation between the resulting features set and the robustness of the nonlinear SVM classifier against variations of its learning model. I also present the results on utilizing features cross-correlation to generate new nested set.

The remainder of this chapter is organized as follows: The proposed shape-based CADx is introduced in Section 2, Experimental results, and conclusions and discussion, are presented in sections 3 and 4, respectively.

5.2 Morphology based CADx of microcalcifications

Computer aided diagnosis for MCs, if approved for clinical use, can have a significant impact on the performance of diagnosis. This requires a careful design of the CADx scheme that produces almost a perfect diagnosis performance. Hence, one should not underestimate the impact of various components of the CADx scheme including shape feature extraction, feature selection, and classification. The shape based diagnosis scheme as proposed in this study and illustrated in Figure 5.1 segments MCs via a multiscale morphological filtering scheme. It also employs radiologist’s input (location and size of MC cluster) to automate the region selection and to improve segmentation of MCs. I also employ several groups of shape descriptors to characterize the region, distribution, and the boundary of individual MCs and their entire cluster. Our scheme also employs a PSO heuristic search technique to accomplish a full model selection of the SVM classifier to optimize the classification performance and generalization ability. This study also compares the
performance of feature selection using univariate based nested subsets methods and heuristic search using binary PSO method.

Figure 5.1: Diagnosis of MCs using shape based CADx, which uses SVM classifier generalization error estimated using leave-one-out (LOO) cross-validation method as an objective function for heuristic embedded feature selection using PSO method.

5.2.1 Morphological based segmentation

Mathematical morphology is recognized to be a very effective tool in digital image processing and is employed by many researchers for pre-filtering, enhancement, segmentation, and shape feature extraction [10], [16]. Morphological image processing is based mainly on dilation and erosion operations [16]. Several studies have reported on the effectiveness of mathematical morphology for MC segmentation and detection [26], [116]-117. In these approaches, segmentation of MCs was accomplished by combining top-hat transform with other image processing tools such as Sobel and Canny edge detectors [26], difference of Gaussian filter [116] and watershed transform [117]. Mathematical
morphology is effective because it can detect and segment bright objects and preserve their shape even when gray-level of surrounding region is inhomogeneous. This precisely what makes morphological algorithms, such as watershed and top-hat transforms, to be excellent candidate algorithms for segmenting MCs and implementing shape based CADx.

5.2.1.1 Segmentation of MCs using morphological filtering

Since it is difficult and impractical to subjectively evaluate the segmentation outcomes, I used the overall performance of the classification scheme and the discriminative power of the extracted shape-descriptors to design and evaluate the proposed segmentation scheme [21]. After several experiments and performance evaluations of the extracted shape descriptors, I have proposed a new segmentation method illustrated by Figure 5.2. This proposed scheme accomplishes MCs segmentation as logical combinations of the binary output of dual modified top-hat transform. A threshold, computed using low order statistics (first and second moments) of the filtered region, applied to the output from each morphological filter bank to produce a binary image representing the segmented MCs.

As demonstrated in Figure 5.2, the basic difference between the conventional top-hat transform and the proposed one is that the later applies additional morphological closing operation that smoothes the background image prior to its subtraction from the original image. Although, one can employ more than two scales, our experimental results indicated that the effectiveness of employing two structure elements of size 5×5 and 7×7.

An important step for an efficient supervised learning is the purity of the training examples representing each class, which requires an efficient MC segmentation and post-processing step to reduce the number false detected signals. In this study, I used a ground
truth file accompanied with each mammogram that included the location and the size of the region best fits MC cluster to generate a binary mask to eliminate all detected signal located outside a rectangular region enclosed.

Figure 5.2: Segmentation of MCs using a dual top-hat filtering scheme. (a) Illustration of a single scale modified top-hat morphological filtering stage, which smoothes an opened image via a closing operation using same structuring element. (b) Extension the filtering stage shown in (a) to a two-scale filtering scheme.

5.2.1.2 Segmentation of MC cluster

Previous studies [45], [105] have demonstrated that analyzing the shape of the entire MC cluster can also be beneficial for distinguishing malignant from benign cluster. In this work, a binary region representing an entire MC cluster is produced using successive
applications of six morphological dilation operations to merge binary regions of the individual MCs into one region, which is adapted from [105]. An illustration of the results of this process is shown in Figure 5.3. Utilizing a prior knowledge such as the size of the ROI encloses each MC cluster allows for accurate automated delineation of the cluster area. Since false detected MCs, located outside actual cluster, might change the regularity of the shape of the cluster area and so alter the computation of related shape descriptors.

![Figure 5.3](image)

Figure 5.3: Segmentation of MC cluster margin. Benign mammogram (mdb223) from MIAS. (a) Original MC cluster, (b) and (c) represent automated and manual delineation of a cluster’s margin shown in (a).

5.2.2 Shape based feature extraction

Shape features from regions represent individual MCs as well as an entire cluster are used to describe each MC cluster examined in this study. Extracted features are grouped into three subsets: region descriptors subset (e.g. area, compactness, eccentricity) [16], boundary descriptors (shape moments, Fourier descriptors) [20] subset, and features describing the distribution of MCs in a cluster (orientation and spreading of MCs in the cluster), which extracted from the binary region of each MCs and whole cluster. Other features used in this
work, which are neither region nor boundary descriptors such as the number of MCs as well as the number of MCs represented by only one pixel [107].

A typical approach for computing various shape descriptors usually started by labeling the individual objects (detected MCs) in a binary-segmented mammographic image. Such a process represents a key step for estimating most shape descriptors. In addition, this labeling of individual MCs provides us with a straightforward and significant set of descriptors each MC cluster [21]. Examples of these features are number of MCs in a cluster, the area of each individual MCs defined as the number of pixels within each connected region, and the number of single-pixel MCs. Other features are the centroid of MCs (non-single pixel) and their distances from the centroid of the entire cluster. Using pre-labeled MCs regions, each single region (representing an individual MC) is extracted to estimate the various region and boundary descriptors.

5.2.2.1 Region descriptors

Region descriptors measured in this work are perimeter, compactness, eccentricity, area of convex hole, major and minor axis length, extent, solidity, equivalent diameter, and orientation, [16]. Segmented MCs are binary regions digitally implemented, which lead to some limitations in the accuracy of the computed descriptors, and theoretically incorrect results might be obtained [23]. This limitation of the digital presentation was solved by increasing the resolution (smaller pixels size) of the region of interest via pixel upscale [23]. This study also applied the region up-scale process to a binary region represented by each single MC before computing scale invariant shape descriptors.
5.2.2.2 Boundary descriptors

Other shape features also extracted in this study are a set of boundary descriptors. These boundary descriptors are usually applied to characterize the regularity of the boundary of a given MCs region. The most popular boundary descriptors are low order normalized shape moments $F_1'$, $F_2'$, $F_3'$, and $F_3' - F_1'$ or $F_4'$ [20], and normalized Fourier descriptor $FF$ [20]. Both descriptors have been proven useful for distinguishing between benign and malignant MCs [20]. However, moments of the shape boundary are relatively more popular [20]-[21], [23], [110], and only few studies have applied Fourier descriptor to measures [20], [21].

Obviously, computing these boundary descriptors requires localizing all pixels along the region contour. Hence, a border extraction method is required in order to compute normalized moments and Fourier descriptors. In this study, I used a straightforward border extraction method that is different from the method used in [23], [107]. This new method first applies a binary erosion operation using a $3 \times 3$ structure element. Then, an eroded region is subtracted from the original one to extract an object boundary. Since these boundary descriptors are invariant to scale, position, and rotation, impact of the digital representation of a binary region on these measures can be minimized by performing a prior region's up-scale [23].

Since each MC cluster contains 3 or more individual microcalcifications, it is common to use first and second moments, maximum value, and range of the measured descriptors (e.g. compactness of each MC) to form a feature vector. In this study, each MC cluster is modeled using 44 shape features as listed in Table 5.1. This feature set consists of
34 features obtained from individual MCs in each cluster, and the remaining 10 features describe the region and boundary of the entire MC cluster.

Table 5.1
Extracted shape features

<table>
<thead>
<tr>
<th>No</th>
<th>Feature Name</th>
<th>No</th>
<th>Feature Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Number of MCs in a 23</td>
<td>23</td>
<td>STD- Compactness **</td>
</tr>
<tr>
<td>2</td>
<td>Number of single-pixel 24</td>
<td>24</td>
<td>Mean- Perimeter of MCs**</td>
</tr>
<tr>
<td>3</td>
<td>Sum-Area 25</td>
<td>25</td>
<td>STD- Perimeter of MCs**</td>
</tr>
<tr>
<td>4</td>
<td>Mean-Area 26</td>
<td>26</td>
<td>Mean-distances from a cluster’s</td>
</tr>
<tr>
<td>5</td>
<td>STD-Area 27</td>
<td>27</td>
<td>STD- distances from a cluster’s</td>
</tr>
<tr>
<td>6</td>
<td>Mean- Equivalent 28</td>
<td>28</td>
<td>Mean –F2'</td>
</tr>
<tr>
<td>7</td>
<td>STD - Equivalent 29</td>
<td>29</td>
<td>STD- F2'</td>
</tr>
<tr>
<td>8</td>
<td>Mean - Solidity 30</td>
<td>30</td>
<td>Mean –F4'</td>
</tr>
<tr>
<td>9</td>
<td>STD - Solidity 31</td>
<td>31</td>
<td>STD- F4'</td>
</tr>
<tr>
<td>10</td>
<td>Mean - Eccentricity 32</td>
<td>32</td>
<td>Max-F4'</td>
</tr>
<tr>
<td>11</td>
<td>STD- Eccentricity 33</td>
<td>33</td>
<td>Mean- FF</td>
</tr>
<tr>
<td>12</td>
<td>Mean- Extent 34</td>
<td>34</td>
<td>STD- FF</td>
</tr>
<tr>
<td>13</td>
<td>STD- Extent 35</td>
<td>35</td>
<td>Area (MC)</td>
</tr>
<tr>
<td>14</td>
<td>Mean- Minor Axis 36</td>
<td>36</td>
<td>Convex Area(MC)</td>
</tr>
<tr>
<td>15</td>
<td>STD- Minor Axis length 37</td>
<td>37</td>
<td>Eccentricity (MC)</td>
</tr>
<tr>
<td>16</td>
<td>Mean- Major Axis 38</td>
<td>38</td>
<td>Circularity (MC)</td>
</tr>
<tr>
<td>17</td>
<td>STD- Major Axis length 39</td>
<td>39</td>
<td>Major Axis (MC)</td>
</tr>
<tr>
<td>18</td>
<td>Mean- Convex Area 40</td>
<td>40</td>
<td>Minor Axis (MC)</td>
</tr>
<tr>
<td>19</td>
<td>STD- Convex Area 41</td>
<td>41</td>
<td>Axis ratio (MC)</td>
</tr>
<tr>
<td>20</td>
<td>Mean- Orientation* 42</td>
<td>42</td>
<td>F2' (MC)</td>
</tr>
<tr>
<td>21</td>
<td>STD- Orientation* 43</td>
<td>43</td>
<td>F3'-F1' (MC)</td>
</tr>
<tr>
<td>22</td>
<td>Mean- Compactness ** 44</td>
<td>44</td>
<td>FF-(MC)</td>
</tr>
</tbody>
</table>

* Single pixel MCs is excluded  ** Prior region up-scale is applied

5.2.3 Feature selection methods

Embedded feature selection integrates feature selection and classifier’s model selection tasks to produce an algorithm that is more efficient for optimizing a given classification scheme than performing feature selection and classifier’s model selection tasks independently. This effectiveness becomes clearer when feature selection and SVM classifier’s model selection are accomplished using computationally expensive search
techniques such as GA or exhaustive search. Hence, I developed a heuristic based embedded feature selection scheme under a PSO-SVM framework that allows for features search during SVM learning process. I also use leave-one-out (LOO) training and testing method to minimize the risk of data over-fitting and to ensure the availability of unseen test patterns that have not been used in any training or feature selection stage.

5.2.3.1 Outweighed univariate based nested subsets method

This method as explained in Section 2.4.3 forms the nested feature subsets using two criteria: individual feature ranking using ROC analysis method and the level of the average cross-correlation among the member of each feature subset. The second criteria exploits the degree of correlation when forming nested subsets, which not only helps in discarding truly redundant features but also provides the chance to include features with some degree of redundancy. It must be noted that for a given level of correlation, that is a real constant, only N candidate feature subsets need to be examined, which offers a relatively simple feature search process.

Single variable feature evaluation using ROC analysis [29] method not only serves as a first stage of this nested subsets method, but also it is useful for investigating the discriminative power of extracted shape features that helps identifying the most important and irrelevant features. According to ROC analysis technique, a feature with stronger discrimination will produce a higher feature score \( Z_n \) that represents a larger area under ROC curve or index \( Az \). Results of a univariate feature ranking of the shape features (per Table 5.1) are presented in Table 5.2. Attempting to interpret the feature score \( Z_n \),
illustrated in Table 5.2, one can observe that some features demonstrate an excellent discrimination between malignant and benign classes.

Clearly, the strongest discriminative descriptor that achieves the highest $Z_n$ of 0.974 is $F_{13}$, which represents the standard deviation of the extent of the regions, where the extent of a binary region is defined as the ratio of the region’s area to the area of the bounding box. Individual MCs in a cluster. Other features that also show a strong discrimination between malignant and benign MCs are the standard deviation of the distance between individual MCs and their cluster’s centroid, the standard deviation of the normalized second order moments ($F_{29}$), the standard deviation of fourth order moments ($F_{31}$), and the standard deviation of normalized Fourier descriptors ($F_{34}$).

<table>
<thead>
<tr>
<th>$F_n$</th>
<th>$Z_n$</th>
<th>$F_n$</th>
<th>$Z_n$</th>
<th>$F_n$</th>
<th>$Z_n$</th>
<th>$F_n$</th>
<th>$Z_n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.897</td>
<td>12</td>
<td>0.545</td>
<td>23</td>
<td>0.936</td>
<td>34</td>
<td>0.571</td>
</tr>
<tr>
<td>2</td>
<td>0.885</td>
<td>13</td>
<td>0.974</td>
<td>24</td>
<td>0.528</td>
<td>35</td>
<td>0.878</td>
</tr>
<tr>
<td>3</td>
<td>0.859</td>
<td>14</td>
<td>0.538</td>
<td>25</td>
<td>0.705</td>
<td>36</td>
<td>0.872</td>
</tr>
<tr>
<td>4</td>
<td>0.545</td>
<td>15</td>
<td>0.769</td>
<td>26</td>
<td>0.878</td>
<td>37</td>
<td>0.529</td>
</tr>
<tr>
<td>5</td>
<td>0.538</td>
<td>16</td>
<td>0.705</td>
<td>27</td>
<td>0.91</td>
<td>38</td>
<td>0.885</td>
</tr>
<tr>
<td>6</td>
<td>0.603</td>
<td>17</td>
<td>0.564</td>
<td>28</td>
<td>0.564</td>
<td>39</td>
<td>0.859</td>
</tr>
<tr>
<td>7</td>
<td>0.59</td>
<td>18</td>
<td>0.522</td>
<td>29</td>
<td>0.91</td>
<td>40</td>
<td>0.91</td>
</tr>
<tr>
<td>8</td>
<td>0.66</td>
<td>19</td>
<td>0.673</td>
<td>30</td>
<td>0.532</td>
<td>41</td>
<td>0.529</td>
</tr>
<tr>
<td>9</td>
<td>0.91</td>
<td>20</td>
<td>0.667</td>
<td>31</td>
<td>0.859</td>
<td>41</td>
<td>0.731</td>
</tr>
<tr>
<td>10</td>
<td>0.577</td>
<td>21</td>
<td>0.534</td>
<td>32</td>
<td>0.679</td>
<td>42</td>
<td>0.718</td>
</tr>
<tr>
<td>11</td>
<td>0.59</td>
<td>22</td>
<td>0.635</td>
<td>33</td>
<td>0.583</td>
<td>44</td>
<td>0.891</td>
</tr>
</tbody>
</table>

* Strongest features are marked bold

Moreover, features that describe the region of an entire MC cluster have achieved high $Z_n$ values. Based on ranking score $Z_n$ from ROC analysis, the most effective features
are: cluster area ($F_{35}$), cluster convex area ($F_{36}$), cluster circularity ($F_{38}$), cluster's major axis ($F_{39}$), and minor length of the cluster area ($F_{40}$). Another feature found to be individually effective is the normalized FF ($F_{44}$) that describes the cluster boundary. To improve the discriminative power of features extracted from a region representing an entire cluster, an algorithm should undertake an accurate segmentation process (i.e. human based boundary tracing) to create a ground truth delineation of an entire cluster.

Univariate based feature ranking as a single feature selection criterion does not guarantee producing feature subsets with best classification performance since discards any interaction interference among features and may as well include completely redundant features. As we stated in Section 2.4, this study employs a control parameter $u$ that is a real number between 0 and 1, to control the level of cross correlation among members of a candidate feature subset. Using this outweighing process, one can significantly decrease the rank of redundant features and affecting the structure (i.e. included features) of small feature subsets rather than the large ones. Since no prior knowledge of the size of the feature subset and the level of cross-correlation will produce a better classification performance, this study integrates feature search process (i.e. selection of $u$ and $N$) into the full model selection process.

5.2.3.2 Feature selection using binary PSO method

As for heuristic feature search using binary PSO method, no prior feature evaluation is required. Candidate feature subsets are obtained from PSO as binary strings $\tilde{F} = F_0 F_1 \ldots F_N$ of $N$ bits with a binary 1 represents a feature will be used for classification and a binary 0 for a dropped feature. Then, each candidate feature subset is used for
classification using SVM classifier. The classification performance based on classification accuracy (or generalized error) is used as a fitness value of the corresponding PSO’s particle (or candidate solution) during PSO search to that achieves a predefined fitness criterion and finds an optimal feature subset.

5.2.4 PSO-SVM full model selection

Indeed, a default solution that is a learning model obtained by solving a nonlinear SVM learning problem, formulated as convex optimization with affine constraints, guarantees minimum training error but mostly will perform poorly on a test data. Hence, some free parameters would need to be adjusted to ensure a better generalization performance. Our SVM model selection process includes the selection of the kernel function and its parameters and classifier’s regularization parameter [45]-[46]. While grid search method, a straightforward method for the SVM model selection, may be feasible for searching in a 2-dimensional discontinues parameter space or when the region of the best solution is known, it can be quite challenging and computationally expensive in higher dimensional and real parameter spaces. In the later case, a heuristic search approach using GAs [110] and PSO [31] is more efficient.

Aiming to optimize the classification performance, defined as the classifier’s generalization error estimated using LOO method, I design and use the appropriate feature subset and the classifier’s parameters. I achieve this goal by using a PSO-SVM framework and by enforcing that each candidate to the full model selection task is a combination of two parameters’ subsets: the first subset is allocated for feature selection and the second, the parameters’ subset, is designated for optimizing the classifier’s learning model. The
Dimensionality of a candidate solution is determined by the method used in the feature selection. For the outweighed univariate based nested subsets method, each candidate solution is represented by 5 coordinates, which include two parameters for feature search (an index feature subset $n$ and average cross-correlation based penalization $u$). As for the SVM model selection, a set of three parameters consisting of a classifier hyper-parameter (i.e. kernel function $K_{Fun}$), kernel’s control parameter $\gamma$, and a classifier’s regularization constant $C$). In addition to the three parameters used for SVM model selection, binary PSO feature search requires $N$ parameters that are converted into a binary string of 44 bits, to represent a potential feature subset. This may be linked to PSO parameters’ encoding by considering each particle in the swarm as a candidate solution to the full model selection task, and to and the fitness function by modeling the corresponding generalization error as the particle’s fitness function.

The parameter’s encoding can be summarized as: let $M$ be the number of particles (solutions to a given objective function) in the swarm. Each particle $X$ can be characterized in the parameter space by a 5-dimensional vector $X = [K_{Fun}, \gamma, C, N, u]$ for the nested subset method and 47-dimensional vector $X = [K_{Fun}, \gamma, C, \bar{F}]$ the binary PSO feature search method. $\bar{F} = [F_0 F_1 \ldots F_{44}]$ with the $ith$ feature $F_i, i = 1, \ldots, 44$ taking 0 or 1 binary values. $K_{Fun}$ is an integer that takes two values choosing between RBF and polynomial kernels. $\gamma$ is a kernel’s control parameter $\gamma$ that can be $\sigma$ or $p$ for the Gaussian and polynomial kernel, respectively.
5.3 Experimental results

5.3.1 Mammographic test data

Proposed CADx scheme and feature selection methods have been tested using 25 MC clusters, illustrated in Figures 3.4 and 3.5, of which 13 are benign and 12 are malignant clusters from [115]. This study also used the ground truth file (the size and centroid of the region that best fits each microcalcification cluster) as a radiologists’ input to extract $128 \times 128$ region centered at each cluster’s centroid.

5.3.2 Experimental setup

PSO algorithm requires two sets of parameters to be determined during the search process. First parameters’ set selected once during the initialization stage and kept fixed for whole search process. This set includes the size of the swarm or number of particles, the boundaries of the search space, and the maximum and minimum velocities for each dimension, the termination criterion, which might be selected as number of iteration or desired fitness level (i.e. average generalization error). The second parameters’ set controls the movement of the particles and PSO search process and includes $c_1, c_2, W, r_1, and r_2$. The learning rate parameters $c_1$ and $c_2$ selected with typical values of $c_1 = 2$, $c_2 = 2$, and a constraint of $c_1 + c_2 = 4$. I also use an adaptive inertia weight $W$ between 0.9 and 0.4 that decreases as the number of iteration increases [32]. The last two parameters $r_1$ and $r_2$ model the random contribution of the social and personal best fitness to the velocity of the particles, respectively.
All experimental results presented in this chapter are obtained by using PSO heuristic search with a swarm of size 100 particles and termination criteria of either maximum iterations of 50 or a zero generalization error. Additionally, I choose the search space limits to be individually selected for each coordinate. For example, the classifier's regularization constant $C$ is real valued between 1 and $10^5$, and a kernel parameter $\sigma$ is real valued between 0.5 and 35 for RBF kernels and an integer $P$ between 1 and 5 for the polynomial kernel case. We also used the classifier's generalization error, the ratio of the number of falsely classified test patterns to the overall number of test patterns, as a primary criterion for model selection and feature selection processes while the corresponding area under ROC or $Az$ index were used as a secondary performance metric to evaluate obtained models.

5.3.3 Impact of MC segmentation

In this study, MCs segmentation with reduced false detected signals and discriminative shape features has been accomplished by considering several design factors. These factors included design of the filtering scheme, utilizing ground truth data (location and size of MC cluster provided by MIAS) to improve the segmentation process, and employing different shape descriptors to characterize the region and boundary of the individual MCs and the entire cluster, and the distribution of MCs in the cluster. As for filtering method, I used a modified top-hat transform that applies additional morphological closing (or smoothing) to the background before performing an image subtraction. Experimental results, as illustrated in Fig 5.4.a, indicated that the proposed morphological filtering achieved classification performance better than segmentation using a standard top-
hat transform. The impact of the selection of the threshold level, used to produce a binary representation of MCs, is presented in Figure 5.4.b. I have tested several threshold levels and used classification performance to select a threshold level that led to shape features with best discrimination between malignant and benign classes. Aiming to eliminate false calcifications outside the actual cluster region, I constructed a binary mask using radiologist’s input, mammogram annotations specifying the size and centroid of the region that best fits each MC cluster. I also found this mask to be useful for segmenting the margin of the entire MC cluster.

Figure 5.4: Impact of the segmentation of the MCs on the classification performance. (a) Comparison of the performance of MC segmentation using the original top-hat and the modified top-hat transforms, (b) the impact of different threshold’s levels on the performance.

5.3.4 Results on feature selection

Embedded feature selection, which integrates feature selection task with parameter’s adjustment of SVM classifier, presented in this study used two feature search strategies:
outweighed univariate based nested subsets method and heuristic search using Binary PSO technique. These methods are mainly different with respect to the complexity of the search process and the size of search space. Embedded feature selection using conventional univariate nested subset method (search is guided by a univariate ranking only) requires N evaluation of the classification performance, which is more computationally efficient than other sequential feature selection (SFS) and heuristic search methods. However, the simplicity of the search space using conventional univariate based nested subset technique mostly will miss an optimal feature subset and lead to sub-optimal feature selection process. Using cross-correlation based outweighing scheme as an additional criterion can improve such method. Experimental results, presented in Figure 5.5, indicate that feature search using outweighed univariate-nested subsets method \((u = 0.33, 0.66 \text{ and } 1.0)\) generates more predictive feature subsets than the univariate ranking-nested subsets method \((u = 0)\).

![Figure 5.5: Classification performance of feature subsets constructed using conventional (\(u=0\)) and outweighed (\(u \geq 0\)) nested subsets.](image)

The correlation level becomes more influential when the feature subset has a small size \((n\) is less than 20). Since no prior knowledge of the best size of feature subset and correlation
level is available, it was essential to optimize this process as part of the model selection process.

When using the PSO-SVM with an outweighed nested subsets approach as a feature selection method, several learning models (feature subsets, kernel and regularization parameters) have achieved the best classification performance. An example of this is presented in Table 5.3, which indicates feature subsets (N=4, 10, 14, and 17) produced similar classification accuracy of 96% and approximately Az of 0.98.

Table 5.3

Full model selection using univariate nested subsets and PSO-SVM methods

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>$C$</th>
<th>$\mu$</th>
<th>$N$</th>
<th>Features</th>
<th>TP/FN</th>
<th>TN/FP</th>
<th>Accuracy</th>
<th>Az</th>
</tr>
</thead>
<tbody>
<tr>
<td>18.3</td>
<td>439</td>
<td>0.18</td>
<td>17</td>
<td>$F_1 F_2 F_3 F_9 F_{13} F_{23} F_{26} F_{27} F_{29} F_{31} F_{35} F_{36} F_{38} F_{39}$</td>
<td>1.0/0</td>
<td>0.92/1</td>
<td>0.96</td>
<td>0.98</td>
</tr>
<tr>
<td>5.5</td>
<td>55</td>
<td>0.71</td>
<td>14</td>
<td>$F_1 F_2 F_3 F_{13} F_{23} F_{26} F_{27} F_{29} F_{31} F_{35} F_{36} F_{40} F_{44}$</td>
<td>1.0/0</td>
<td>0.92/1</td>
<td>0.96</td>
<td>0.98</td>
</tr>
<tr>
<td>6.9</td>
<td>55</td>
<td>0.88</td>
<td>10</td>
<td>$F_2 F_9 F_{13} F_{23} F_{35} F_{36} F_{37} F_{29} F_{31} F_{40}$</td>
<td>1.0/0</td>
<td>0.92/1</td>
<td>0.96</td>
<td>0.98</td>
</tr>
<tr>
<td>11.3</td>
<td>75</td>
<td>0.84</td>
<td>4</td>
<td>$F_9 F_{13} F_{31} F_{40}$</td>
<td>0.91/1</td>
<td>1.0/0</td>
<td>0.96</td>
<td>0.98</td>
</tr>
</tbody>
</table>

Clearly, binary PSO algorithm is relatively more complex than nested subset methods because the former requires $N$-dimensions of PSO parameter space to accomplish feature selection task while only 2-dimensions is needed for the nested subsets method. This relatively complex feature search using binary PSO provides a larger search space with a higher possibility of finding an optimal feature subset. Results presented in Table 5.4 demonstrate the superiority of feature selection using binary PSO method, which achieved an optimal classification performance, 100% classification accuracy, using several learning
models. I found that the higher the classification accuracy using binary PSO was at the expense of the relatively higher complexity in the search method and the size of the best feature subset. The smallest size of the optimal feature subset was 9 compared to a subset of size 4 from the nested subset method.

Table 5.4

Full model selection using binary PSO-SVM method

<table>
<thead>
<tr>
<th>$C$</th>
<th>$\sigma$</th>
<th>$N$</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.74</td>
<td>2.62</td>
<td>9</td>
<td>F$<em>{13}$F$</em>{14}$F$<em>{23}$F$</em>{26}$F$<em>{30}$F$</em>{31}$F$<em>{33}$F$</em>{35}$F$_{37}$</td>
</tr>
<tr>
<td>378.0</td>
<td>11.85</td>
<td>13</td>
<td>F$<em>7$F$</em>{13}$F$<em>{18}$F$</em>{22}$F$<em>{25}$F$</em>{32}$F$<em>{36}$F$</em>{37}$F$<em>{38}$F$</em>{40}$F$_{42}$</td>
</tr>
<tr>
<td>235.0</td>
<td>1.28</td>
<td>14</td>
<td>F$<em>3$F$</em>{11}$F$<em>{13}$F$</em>{15}$F$<em>{16}$F$</em>{17}$F$<em>{18}$F$</em>{20}$F$<em>{22}$F$</em>{31}$F$<em>{32}$F$</em>{34}$F$<em>{38}$F$</em>{40}$</td>
</tr>
<tr>
<td>167.0</td>
<td>8.20</td>
<td>16</td>
<td>F$<em>3$F$</em>{10}$F$<em>{12}$F$</em>{16}$F$<em>{18}$F$</em>{23}$F$<em>{24}$F$</em>{29}$F$<em>{30}$F$</em>{31}$F$<em>{33}$F$</em>{37}$F$<em>{40}$F$</em>{44}$</td>
</tr>
<tr>
<td>384.0</td>
<td>9.87</td>
<td>18</td>
<td>F$<em>3$F$<em>8$F$</em>{11}$F$</em>{13}$F$<em>{17}$F$</em>{18}$F$<em>{21}$F$</em>{22}$F$<em>{30}$F$</em>{31}$F$<em>{32}$F$</em>{33}$F$<em>{34}$F$</em>{37}$F$<em>{38}$F$</em>{39}$F$<em>{40}$F$</em>{44}$</td>
</tr>
<tr>
<td>102.0</td>
<td>6.77</td>
<td>21</td>
<td>F$<em>3$F$<em>7$F$<em>8$F$</em>{12}$F$</em>{14}$F$</em>{15}$F$<em>{17}$F$</em>{18}$F$<em>{22}$F$</em>{25}$F$<em>{29}$F$</em>{30}$F$<em>{31}$F$</em>{32}$F$<em>{34}$F$</em>{37}$F$<em>{39}$F$</em>{41}$F$<em>{42}$F$</em>{43}$F$_{44}$</td>
</tr>
<tr>
<td>315.0</td>
<td>2.88</td>
<td>24</td>
<td>F$<em>1$F$<em>2$F$<em>3$F$</em>{10}$F$</em>{11}$F$</em>{14}$F$<em>{15}$F$</em>{17}$F$<em>{18}$F$</em>{20}$F$<em>{23}$F$</em>{24}$F$<em>{26}$F$</em>{27}$F$<em>{30}$F$</em>{31}$F$<em>{32}$F$</em>{33}$F$<em>{35}$F$</em>{36}$F$<em>{40}$F$</em>{44}$</td>
</tr>
</tbody>
</table>

*Indices $F_i$'s represent shape features presented in Table 5.1

Results presented in Table 5.3 and 5.4 indicated that the best classification performance was achieved using different learning models. Hence, one may ask, which feature subset one should select as the final classification model? A very well acceptable answer can be formulated using Occam's razor principle that suggests the following: a simple solution is a correct one. In other words, selecting a solution or model with lower number of features mostly leads to a classifier with better generalization ability.

Considering various models obtained from outweighed univariate nested subsets method, presented in Table 5.3, I observe that a model with a feature subset of size 4 is a
possible candidate. However, this subset produced 0 FP and 1 FN results while all other subsets produced 1 FP and 1 FN. Since FN results (i.e. missing a cancer) is relatively considered of a higher risk than a FP (false breast biopsy), selection of a subset 4 should be avoided by radiologist. Following Occam’s razor again, the candidate feature subset of size 10 is expected to be more suitable for the final model and for classifying new test patterns.

This process of determining best feature subset was also applied to optimal feature subsets (using binary PSO), results presented in Table 5.4, and allowed for the selection of a feature subset of size 9 for the final classification model.

Even though a model selection using a theoretical principle might provide a general guideline, empirical evidence is still necessary to validate any selection. Hence, in the next subsection, I examine the impact of the feature selection process on the robustness of the SVM classifier to variations of the regularization and kernel’s parameters.

5.3.5 Results on classifier model selection

Although the generalization performance of SVM classifier has been demonstrated to be sensitive to the model selection process, only few studies [44]-[45] have examined the robustness of their SVM based classification scheme to parameter values. This study not only has examined the robustness of SVM classifier to the selection of the kernel function and regularization constant but also it investigated how the feature selection process is impacting performance.

For instance, the RBF kernel outperformed polynomial kernel in all experiments of the classifier’s hyper-parameter selection. Results obtained using univariate based feature selection method indicate higher classification accuracy of 96 % (0 FN and 1 FP) and $Az$ of
0.98 from RBF kernel compared to a 92% classification's accuracy (1 FN and 1 FP ) and Az of 0.95 using a polynomial kernel, as shown in Figure 5.6. While both kernel functions produced perfect classification using binary PSO feature selection, the RBF kernel can be considered to be the more effective since it used a feature subset of size 9 compared to 16 needed by the polynomial kernel.

![Graph showing classification performance](image)

Figure 5.6: Effect of the selection of the kernel function on the classification performance using SVM.

The classification performance of all learning models, Tables 5.3 and 5.4, is a function of both RBF kernel’s parameter σ and regularization constant C. Investigating the effect of these parameters on the classifier’s performance (i.e. generalization error) indicated sensitivity of classifier’s performance to the values σ and C. I present our analysis of the models in Table 5.3 and Table 5.4, which are illustrated in Figure 5.7 and 5.8 as follow:

- The generalization error of SVM descends as the C value deviates from its optimal value that is given in Table 5.3. As shown in Figure 5.8.a, learning models with feature subsets of size 4, 14, and 10 achieved best generalization error using C values of less
than 60, while a learning model with a subset of size 17 required a C value of about 400.

- Using all learning models, results indicate that increasing the value of the regularization parameter C increases the generalization error. This is may be due to the fact the SVM classifier tends to over-fit the training data when using larger values of the regularization parameter.

- Results on varying the RBF kernel parameter $\sigma$, as illustrated in Figure 8.b, indicated that the performance of the SVM classifier is more sensitive to small values of the kernel parameter $\sigma$. This can be justified by observing that a small value of $\sigma$ leads to a highly nonlinear decision boundary that produces a poor generalization performance. As the $\sigma$ value increases, the generalization ability of most learning models becomes more robust. This trend is mostly because the larger kernel’s width tends to improve the linearity of the decision function and thus leading to the attainment of a better generalization performance.

The feature selection process also indicated that there is a significant effect on the robustness of the SVM classifier to variations of $\sigma$ and C and consequently, on the generalization performance.

- Using feature subsets of size 17, SVM shows superior robustness over a wide range of C values.

- For small values of the parameter C ($C < 40$), other feature subsets such as subset of size 10 seem to provide better robustness than a subset of size 17.
Similarly, a feature subset of size 17 indicates better robustness over most of the range of $\sigma$. However, for small values of $\sigma$, small features subsets such as a subset of size 4 provides better performance.

Using the same procedure, I also examined the learning models presented in Table 5.4. Results also pointed to the importance of selecting appropriate values of the parameters C and $\sigma$. For instance, a learning model with the feature subset of size 9 produced the best generalization error when the values of the parameters C and $\sigma$ are chosen between 211 to $10^5$, and 3 to 11, respectively. In addition, the learning model with a feature subset of size 18 outperformed all other models (excluding a model with a subset of size 9) when C values are set between 60 and 500. As for how robust is our feature selection especially in response to variations of C and $\sigma$, our results as shown by Figure 5.8 demonstrate that a feature subset of size 9 consistently provided the best robustness and outperformed all other models regardless of the value of the regularization constant C as well the parameters $\sigma$.

5.3.6 Impact of a human based interpretation

The main goal of developing CADx technology is to aid radiologists in interpreting mammograms, as opposed to replacing them, to decrease the rate of false invasive breast biopsies. Moreover, a CADx system that might be adopted for clinical use should have a perfect performance. Hence, I believe that such a system can be improved by incorporating non-image features data such as patient’s age, weight, family history, and prior radiologists’ interpretations such as BI-RAD™ rating obtained from previous readings of the patient’s mammograms.
Figure 5.7: Impact of the model selection on the classifier generalization performance with feature subsets obtained using modified nested subsets method. Robustness of SVM classifier to variations of the a) regularization parameter, and (b) the RBF kernel parameter $\sigma$. Generalization error in figures a and b was computed by averaging over several values of the parameter $\sigma$ and $C$, respectively.
Figure 5.8: Impact of the model selection on the classifier generalization performance with feature subsets obtained using a heuristic binary PSO method. (a) Robustness of SVM classifier to variations of the a) regularization parameter, and (b) the RBF kernel parameter $\sigma$. Generalization error in figures a and b was computed by averaging over several values of the parameter $\sigma$ and $C$, respectively.
Another approach for improving CADx technology performance is by incorporating the results of the interactions between a human and computer system. Such interactions may include establishing a ground truth that describes the size and location of the breast lesion and manually delineating the margin of the lesion. For instance, including non-image features such as patient's age has shown to be useful when used with other shape and distribution features in the work presented in [21].

In this work, I have incorporated the human based interpretation of mammograms in two different scenarios. The first approach used a ground truth file of each lesion that describes the size and location of each MC cluster to improve their segmentation while in the second approach; an accurate delineation of the cluster margin performed by an expert radiologist is used to replace the segmentation results from the automatic method, which is illustrated by Figure 5.9. Similar to the feature extraction process that was presented in Section 5.2.2, the region and boundary of the entire microcalcification cluster is described using 10 features along with 34 shape descriptors that are obtained from the individual MCs for charactering the malignancy of any cluster.

The results of this hybrid segmentation method that is a combination of the manual tracing of the cluster margin and the automatic segmentation of MCs has led to an optimal classification performance of 100 % accuracy using both nested subsets and binary PSO feature search methods. Moreover, results highlighted the importance of characterizing the cluster margin in the discrimination process between malignant and benign MC clusters. This result was demonstrated by discarding the 34 features that are extracted from the shape of individual MCs and by accomplishing the classification of 25 MC clusters using only 10 features that represent the entire MC cluster.
Figure 5.9: Examples of the manual (i.e. done by an expert radiologist) and the automatic delineation of the margin of the microcalcification cluster. Figures (a), (c), and (e) demonstrate results of manual delineation while (b), (d), and (f) are examples of the automatic process.
Using shape features of entire MC cluster, I have classified 25 MC clusters and obtained a classification accuracy of 96% (0 FP and 1 FN) and 92% (0 FP and 2 FN) for a manually and automated delineated cluster, respectively. As for the most discriminative features from different methods that have been used for delineating the cluster margin, the best classification result was obtained using two features (two normalized shape boundary moments $F_2$ and $F_4$) in the case of manual delineation while it required three features (the area, length of the minor axis of the cluster, and normalized Fourier descriptor $FF$) in the case of automated delineation of the cluster margin.

5.3.7 Comparison with other CADx

The fact that there is no one common dataset is used by different CADx approaches makes any comparison attempts difficult. However, comparing results of CADx scheme proposed in this work with other CADx algorithms that used the same dataset, indicate that our scheme achieved better classification results than others such as [45], [110]. Papadopoulos et al. [45] used SVM and ANN and obtained $Az$ of 0.81 and 0.78. In addition, Wang et al. [110] used mixed texture and shape features and GA for SVM model selection and dimensionality reduction of the feature space and achieved $Az$ of 0.86.

5.4 Discussion and conclusions

In this chapter, an integrated framework composed of MC morphological-based segmentation techniques, region and boundary descriptors of the individual MCs and the MC cluster shapes, PSO-based embedded feature selection method, and supervised learning using nonlinear SVM is developed to classify mammographic MCs into malignant and
benign classes. The proposed approach employed a PSO heuristic search to accomplish a full model selection of our SVM-based classification scheme.

Experimental results demonstrated that several design factors have significant effect on automating the diagnosis of MCs and thus they must be carefully selected. These factors include the choice of the segmentation technique, the shape features used to characterize individual MCs and the MC cluster, the feature selection method and the level of redundancy within selected features, and the elected learning model for kernel based SVM classifier. I, also, would like to reiterate the importance of appropriate feature selection and its effect on the generalization capability of the SVM classifiers and its robustness to the variations of kernel’s and regularization parameters.

Comparing the performance of feature search using outweighed-nested subsets method and heuristic binary search methods indicated the effectiveness of both methods. However, the feature search using binary PSO method is more efficient because of the limitations of the nested subsets methods when the input features individually do not show strong discriminative power.

Results of our work illustrated the effectiveness of heuristic search using PSO for accomplishing both feature selection and classifier’s model selection. This work can be further extended to optimize the feature extraction process and MCs segmentation using a larger mammographic dataset.
CHAPTER VI

CHARACTERIZATION OF CLUSTERED MICROCALCIFICATIONS
USING MULTISCALE HESSIAN BASED FILTERING

6.1 Introduction

Using normalized second derivative and an image Hessian are standard techniques for enhancement, detection, and extraction of curvilinear structures such as blood vessels and blobs in medical images [19],[118],[120]. Expansion of these methods for the analysis of mammographic abnormalities has been limited to enhancement and detection applications [83]-[84], [119]. Hessian filtering is based on using two eigenvalues and their corresponding eigenvectors obtained from the Hessian matrix for each pixel. Both the sign and the magnitude of the eigenvalues, that is the directional second derivatives, can be used to characterize intensity, shape, and orientation of different 2D/3D image structures [17]. For example, a bright blob like structure in a 2-D image identified by having negative and equal eigenvalues in all directions. Moreover, computing the ratio of the two eigenvalues can also distinguish between a line and blob like structures.

In this dissertation, characterization of MCs is accomplished by first performing multiscale representation of mammographic regions (containing MCs) using multiscale Hessian based filtering. Then, a set of spectral measures such as spectral entropy and energy are extracted from each scale. The performance of the extracted features is evaluated by classifying a set of microcalcification (MC) clusters as malignant and benign using simple k-nearest neighbor (kNN) classifier and area under ROC.
The remaining sections of this chapter are organized as follows: Section 6.2 presents pattern analysis using image Hessian. Classification using Hessian based feature extraction is presented in Section 6.3. Experimental results and performance evaluation are presented in Section 6.4 while conclusions are included in Section 6.5.

6.2 Analysis using image Hessian

Characterizing various image structures using image Hessian is usually accomplished by interpreting the directional second derivative or equivalently the eigenvalues from a Hessian matrix [17] as shown in Table 6.1.

<table>
<thead>
<tr>
<th>Eigenvalue</th>
<th>2-D structure patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda_1 &gt; 0 ) and ( \lambda_2 &gt; 0 )</td>
<td>Dark, nodular</td>
</tr>
<tr>
<td>( \lambda_1 &gt; 0 ) or ( \lambda_2 &gt; 0 )</td>
<td>Dark, linear</td>
</tr>
<tr>
<td>( \lambda_1 &lt; 0 ) and ( \lambda_2 &lt; 0 )</td>
<td>Bright, nodular</td>
</tr>
<tr>
<td>( \lambda_1 &lt; 0 ) or ( \lambda_2 &lt; 0 )</td>
<td>Bright, linear</td>
</tr>
</tbody>
</table>

For a bright nodular structure, second derivatives \( f_{xy} \) and \( f_{xx} \) are both negative and significantly larger than \( f_{yx} \) and \( f_{yx} \), which leads to negative and non-zero \( \lambda_1 \) and \( \lambda_2 \) [17]. These results can be also extended to model a dark pattern located on a relatively brighter background by simply considering the complement of the model given in equation (2.17).
Filtering curvilinear image structures (nodular or linear) using image Hessian is usually characterized by using two measures; a function of the magnitude of the two eigenvalues called structureness and the ratio of the two eigenvalues [17], [120].

\[ C_{NC}(\lambda_1, \lambda_2) = |\lambda_1| + |\lambda_2| \quad (6.1) \]

Other methods for constructing a structureness measure \( C_{NC} \) have been presented in [83]-[84], [120]. Truc et al. [120] computed \( C_{NC} \) as the norm of the two eigenvalues while a maximum absolute value of \( \lambda_1 \) and \( \lambda_2 \) was used in [83]-[84]. The impact of various \( C_{NC} \) will be discussed in Section 6.4.

Since a nodular image component has a non-zero and negative \( \lambda_1 \) and \( \lambda_2 \). Then, one can to enhance nodular mammographic features and suppress undesired linear structures by using the ratio of the two eigenvalues \( R_{NC}(\lambda_1, \lambda_2) \), expected to be close to one for disk-like structures, which is defined as follows

\[ R_{NC}(\lambda_1, \lambda_2) = \eta(\lambda_1, \lambda_2) \exp(-\beta(\lambda_{\text{max}} / \lambda_{\text{min}})^2) \quad (6.2) \]

where \( \eta(\lambda_1, \lambda_2) \) is an indicator function that is unity if both \( \lambda_1 \) and \( \lambda_2 \) are negative and zero otherwise, and \( \beta \) is a real constant that has been selected empirically to 0.5 in this study.

Using structureness and ratio measures, overall response of Hessian filter can be defined as a product of two terms

\[ H_{NC} = R_{NC}(\lambda_1, \lambda_2) C_{NC}(\lambda_1, \lambda_2) \quad (6.3) \]
Figure 6.1: Filtering of mammographic microcalcifications using a Hessian based filter. (a) A mammographic region, b) the structureness of the Hessian filter and b) the filter response computed as defined in equations (6.1) and (6.3), respectively. Mammogram is from MIAS database.

In case of a nodular structure, two-eigenvalues ratio $R_{NC}(\lambda_1, \lambda_2)$ mostly will have a nonzero value that is close to one that can be used to enhance nodular structure. On the other hand, the value of $R_{NC}(\lambda_1, \lambda_2)$ is very small value for linear structures, which can be used to suppress a signal from such structures. Figure 6.1 illustrates an example of mammographic microcalcifications filtering using Hessian image analysis.

6.3 Diagnosis of MCs using Hessian based feature extraction

The classification methodology as illustrated by Figure 6.2 starts by extracting a mammographic region enclosing a microcalcification cluster in the center. The process of a multiscale Hessian based analysis process, as explained in Section 2.2.3, is achieved by first computing multiscale directional second derivatives of the Gaussian kernel at different
scales for the identified mammogram region. Then, four directional derivatives from each scale are used to create a Hessian matrix and to compute two eigenvalues for each pixel.

Multiscale analysis and transform domain methods such as multi-wavelet transforms, wavelet packets [23], and a discrete cosine transform (DCT) [26], commonly perform a texture characterization of clustered MCs by measuring two spectral features: normalized energy and entropy from each subband of the multiscale representation. In this work, in addition to measuring energy and entropy spectral values used usually to analyze the texture, I characterize the texture of each scale by computing two spectral measures from each scale, which represent the average of the two-eigenvalue ratio and the filter response as defined in equations (6.2) and (6.3), respectively.

Univariate feature ranking based on Fisher-score criterion [34], explained in Section 2.4.2, is used to identify the significance of the extracted features by estimating the discriminative power of each feature. This ranking process can be further used to guide the feature search and formation of candidate feature subsets using multivariate feature selection methods such as nested subset and sequential forward feature selection methods. This dissertation mainly focuses on investigating whether a Hessian based feature extraction can produce a satisfactory discrimination between benign and malignant MCs or not.

Hence, the proposed classification methodology adopts relatively simple classifiers such as kNN instead of a more complex supervised learning scheme such as SVM, also it does not model or assumes any prior knowledge of the distribution of the classified data or integrates other popular feature extraction techniques.
Figure 6.2: Classification of MC cluster using Hessian image analysis and a kNN classifier. *ROI depends on a cluster ground truth from MIAS.

6.4 Experimental results

6.4.1 Mammogram test data

The proposed feature extraction scheme has been evaluated using 33 microcalcification (MC) clusters, of which 20 clusters are malignant and 13 are benign. These MC clusters have been extracted from 23 mammograms from MIAS mini-database [115]. More details on this database and illustration of these MC clusters have been presented in Section 3.6.
6.4.2 Parameters and feature selection

The individual MCs sizes and the entire MC cluster were employed in several studies [23], [107] to distinguish a malignant MC cluster from a benign one. This has lead to the following questions: 1) what is the size of the filtering kernel that can maximize a response from MCs of different sizes and orientations? 2) What is the size of the region depicting a microcalcification cluster that might be suitable for characterizing a given MC cluster?

Generally, it is very difficult to have a prior knowledge of the size of each MCs in a given MC cluster. However, filtering the MCs using a multiscale filtering kernel that is also a rotation invariant kernel can provide an efficient method to maximize signals from existing MCs of different sizes. On the other hand, the size of the whole MC cluster is easier to be estimated, which can be accomplished either through a manual delineation of the cluster margin manually by radiologists or by using a morphological image processing technique.

Being tiny deposits of calcium, clustered MCs can be modeled as small scale bright blobs that contribute more to the highpass frequency subbands. Hence, Gaussian kernels of small scales are more suitable and can produce a stronger second derivative and a larger magnitude of the eigenvalues. I have empirically selected a set of three Gaussian kernels with \( \sigma = 0.25, 0.5, \) and \( 0.75 \) and size in pixels of \( 3 \times 3, 5 \times 5, \) and \( 7 \times 7, \) respectively. These three kernels were applied to all mammographic regions to produce a multiscale Hessian based filtering and used to extract different spectral measures.

To answer the second question, I have performed two experiments. In the first one, a region of size \( 128 \times 128 \) pixels that contained a true MC cluster in its center was used to compute normalized entropy and energy features. The second experiment used a region that best fits a given MC cluster for feature extraction. Our results indicated that the second test
is more appropriate since it produced more discriminating features and better classification results.

Analysis of mammographic regions using a three-scale image Hessian presented in this chapter leads to a set of 12 spectral measures (4 measures per scale) per mammographic region. Examining the cross-correlation among the features indicated the high correlation between corresponding measures at different scales. Thus, I used the first statistical moments (i.e. mean over scales) of each spectral measure that produced a smaller, and less correlated, feature subset that consists of four features: mean of a scale-normalized energy, mean of a scale-normalized entropy, mean of a scale-two eigenvalue ratio, and mean of a scale-total filter response.

The results of evaluating these four features using a Fisher criterion or F-score method are given in Table 6.2. Clearly, the feature representing the mean of the normalized entropy achieves the highest F-score, which is expected to provide the strongest discrimination between classes. This result is further illustrated by Figure 6.3, which demonstrates the superior predictive power of the normalized entropy feature compared to the normalized energy measure. Another feature that produces the second highest score is the average of the total filter response, which is better than both the normalized energy and the two-eigenvalue ratio features. Although the feature extracted from the two-eigenvalue ratio achieves slightly a higher score than normalized energy, the latter is found more effective when combined with other features.
Table 6.2

Univariate feature ranking using F-score method

<table>
<thead>
<tr>
<th>NO</th>
<th>Feature description</th>
<th>F-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mean-scale normalized entropy</td>
<td>0.742</td>
</tr>
<tr>
<td>2</td>
<td>Mean-scale normalized energy</td>
<td>0.145</td>
</tr>
<tr>
<td>3</td>
<td>Mean-scale two-eigenvalue ratio</td>
<td>0.166</td>
</tr>
<tr>
<td>4</td>
<td>Mean-scale Hessian filter response</td>
<td>0.2527</td>
</tr>
</tbody>
</table>

![Graph showing the discriminative power of energy and entropy features](image)

Figure 6.3: The discriminative power of the energy and entropy features.
6.4.3 Performance evaluation

The predictive power of the proposed feature extraction scheme was evaluated using the kNN classifier and the leave-one-out (LOO) cross-validation method to classify MC clusters into malignant and benign cases. Again, a cross-validation using LOO strategy as explained in Chapter II simply trains a classifier using all data samples excluding one sample that is kept for a testing purpose, which is more appropriate for a small data set like this study. The small size of the feature set extracted in this chapter enables us to search for an optimal subset by evaluating all possible feature subsets. Our experiments indicated that the best classification performance can be obtained using three feature subsets. The first subset included only one feature represented by mean of normalized entropy, the second subset included both the mean of total filter response and the normalized entropy. In addition to the feature included in the second subset, the third subset included the energy feature.

Classification results are evaluated using a set of popular evaluation metrics; specificity sensitivity, classification’s accuracy, and an area under receiver operating characteristic curve. Table 6.3 presents the classification results for several \( k \) values and two feature subsets, which indicates that the best classification performance that could be achieved is of accuracy of 85\%, which corresponds to a sensitivity of 90\% (or 2 FN) and a specificity of 77\% (or 3 FP).

Further evaluation of the different feature subsets was accomplished by constructing a computerized ROC curve of the classifier and by computing the area under ROC curve or \( Az \) index for each model (feature subset, \( k \)). Experimental ROC curves and the corresponding \( Az \) values for each feature subset and \( k = 7 \) are demonstrated in Figure 6.4, which demonstrates that texture based classification of MIAS mammograms using the
The proposed feature extraction scheme could achieve a classification’s performance up to $Az = 0.83$. This result is obtained using a feature subset of three features $[F_1, F_2, F_3]$.

Table 6.3

<table>
<thead>
<tr>
<th>$k$</th>
<th>$[F_1, F_4]$</th>
<th>$[F_1, F_2, F_4]$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensitivity</td>
<td>Specificity</td>
</tr>
<tr>
<td>1</td>
<td>0.85</td>
<td>0.615</td>
</tr>
<tr>
<td>3</td>
<td>0.9</td>
<td>0.767</td>
</tr>
<tr>
<td>5</td>
<td>0.9</td>
<td>0.767</td>
</tr>
<tr>
<td>7</td>
<td><strong>0.9</strong></td>
<td><strong>0.767</strong></td>
</tr>
<tr>
<td>9</td>
<td>0.9</td>
<td>0.767</td>
</tr>
<tr>
<td>11</td>
<td>0.75</td>
<td>0.767</td>
</tr>
</tbody>
</table>

Although feature subsets of size 1 and 2 produce classification accuracy that is similar to three-feature subset, their performance measured using the area under ROC curve is lower than performance obtained using a three-feature subset.

Other factors that might affect the discriminative power of the extracted features are the design of the Hessian filter (i.e. the method used to compute the structureness measure) and the size of the mammographic region used to extract spectral measures of each cluster. I have examined the classification performance using three different definitions of the structureness $C_{nc}$: a two-eigenvalue norm as given in equation (6.1), a two-eigenvalues maxima, and a two-eigenvalues sum. Our results, as demonstrated in Figure 6.5, indicated that the diagnosis performance of the features significantly degraded when the structureness measure was defined as norm and maxima of the two eigenvalues while a superior
Figure 6.4: ROC analyses of the extracted features.

Figure 6.5: Effect of the structureness of the Hessian filter.
performance was obtained when $C_{NC}$ of the Hessian filter was computed as a two-eigenvalue sum.

As for the size of the mammographic region used to characterize MC clusters, results indicated that using a region, which best fits each microcalcification cluster is more suitable and can lead to more discriminating features than using one region of a fixed size (e.g. 128×128) for all MC clusters.

Several approaches accomplished the diagnosis of MCs using different texture feature extraction techniques such as wavelet packets [23], which achieved classification performance $Az$ between 0.74 and 0.79, multi-wavelets produced $Az$ of 0.89 and GLCM achieved $Az$ of 0.75 [23]. Other studies [112] also accomplished the diagnosis of MCs via analyzing the texture of the breast tissue surrounds MCs using Law texture measures, gray-level co-occurrence matrices (GLCM), and gray-level run-length matrix, which produced classification accuracies of 0.89, 0.82, and 0.63, respectively. However, direct comparison of the results of our study with other existing method is difficult because of the difference in the datasets used in these studies.

6.5 Discussion and conclusions

In this chapter, a spectral feature extraction based on a multiscale Hessian image analysis has been used to characterize mammographic microcalcification clusters as malignant and benign classes. The proposed features were tested using 33 MC clusters, extracted from MIAS database of which 13 are benign and 20 are malignant cases. Experimental results indicated that the proposed feature extraction scheme achieved a satisfactory characterization of the malignancy of microcalcification clusters. Evaluating the
proposed feature extraction approach using a k-nearest neighbor classifier and the ROC performance measure produced a classification accuracy of 0.85 (2 FN and 3 FP) and Az of 0.83.

Results obtained in this study indicated that MCs could be effectively characterized by analyzing the Hessian. However, empirical evidence needs further investigation using a larger database. In addition, a better classification performance could be achieved if Hessian analysis combined with other feature extraction techniques (morphological and other texture methods). Another direction of investigation could include testing other filtering kernels such as truncated Gaussian kernel, and perfect reconstruction filter banks. Moreover, this Hessian based filtering method could be a more effective and fruitful tool because it can provide a unified scheme for extracting both spectral and shape features of MCs that can be extended to characterize other types of abnormality in mammogram such curvilinear structures of spiculated masses, and circumscribed masses that can be modeled as large nodular structures.
CHAPTER VII

CONCLUSIONS AND FUTURE WORK

Clustered microcalcifications (MCs) appear frequently on screen mammograms and mostly represent an exclusive early sign of breast cancer. However, about 80% of MCs found in the female breast are benign breast disease. Detection of clustered MCs in digital mammograms is relatively easier than characterizing malignancy of a specific microcalcification cluster. Additionally, detection and segmentation tasks become more challenging and error-prone processes when MCs are surrounded by dense breast tissue. Hence, computer aided detection and diagnosis systems are developed to serve as a second opinion that can assist radiologists in interpreting screen mammography and hopefully they can be a substitute of a double reading stage that is not feasible in many situations.

7.1 Summary and conclusions

Computer aided detection (CADe) and diagnosis (CADx) methods presented in this work present a new machine learning based CADe, a four-stage shape based CADx, and a new approach for characterizing the malignancy of microcalcification cluster using texture analysis.

In Chapter 4 of this dissertation, I have developed a new model based framework for segmentation and detection of microcalcification clusters. This new approach was mainly based on modeling anonymous real MCs in a given mammogram using a synthetic model of mammographic MCs that was blended in the glandular breast region. I was able to detect
MC clusters via a pixel based feature extraction and accomplish a self-learning of the Bayesian classifier by using synthetic MCs patterns and real examples of the glandular breast tissue that were extracted from the same mammogram. Comparing the performance of the proposed CADe approach with other methods from the literature indicated an adequate detection accuracy of the proposed CADe.

To discriminate a malignant from a benign microcalcification cluster, I have developed a four-stage morphology based CADx, which is based on the segmentation of MCs using morphological image processing, shape feature extraction, a heuristic PSO-SVM feature selection, and a binary classification using SVM. Diagnosis of MCs using their shape significantly depends on the accuracy and robustness of the methods used for segmentation and shape feature extraction. To attain this goal, I have proposed a new segmentation method based on using a morphological top-hat operator and a multiscale structuring element.

Bearing in mind the main purpose of CADx that is to help radiologists to differentiate between malignant and benign microcalcification clusters, I have utilized some image annotations that was a cluster “ground truth” to improve the segmentation process. In this dissertation, I examined the impact of the human based interpretation by presenting a hybrid-segmentation scheme that automatically segmented individual MCs in a given MC cluster while a human based delineation of the cluster margin was used to segment the entire cluster. This process improved the discriminative power of the shape features representing the entire cluster (e.g. compactness, area, eccentricity) and the overall classification performance of the diagnosis scheme.

An efficient representation of the morphology of a microcalcification cluster is based on finding a small number of features with the best discrimination and generalization ability,
which can be accomplished via appropriate selection of the shape features and the classification scheme. In this dissertation, I developed a heuristic PSO-SVM full model selection framework to incorporate a feature selection task and SVM learning process. This dissertation also developed and examined the performance of feature selection based on nested subsets methods and a heuristic and random search using a binary PSO method. Results of this study concluded that a feature search using an outweighed based nested subsets method is very comparable to the heuristic approach using a binary PSO method if the extracted features are individually discriminative.

Compared to the size of the training set, feature space was relatively large and some degree of correlation was presented among the shape features. Several learning models (feature subset, kernel’s parameter $\sigma$, and classifier’s regularization constant $C$) with the same classification performance were obtained using PSO-SVM full model selection process. To select a final learning model that provide the best generalization performance, I have developed and applied a selection criterion based on the robustness of the classification performance to the variation of some parameters such as ($\sigma$ and $C$).

Texture analysis of mammographic regions enclosing microcalcification clusters can improve the performance of an overall CADx when a shape based CADx fails or its performance is severely degraded due to the prior poor segmentation process results. In Chapter 6, I have tackled this problem by characterizing the texture of the MCs region by analyzing multiscale image Hessian and by measuring spectral descriptors such as energy and entropy. Although the proposed texture analysis has not competed with the shape features of Chapter 5, this method has shown to be more efficient than other existing texture based approaches tested on MIAS dataset. Moreover, the inferior performance of the texture
analysis using Hessian can be attributed to the fewer number of features and the sup-optimaly and limitations of a kNN classifier compared to the SVM method.

7.2 Dissertation contributions

This dissertation presents several contributions by addressing and responding to several of the limitations and challenges of the current state of computer aided detection and diagnosis systems for mammographic MCs.

- This dissertation proposed a new framework that integrates a statistical Bayesian classifier and a pattern-synthesizing scheme for detecting clustered MCs. This new detection approach provide a self-learning scheme by which it synthesizes the training set of MCs class and accomplishes the learning phase in an efficient and simplified method that does not require large training data set, which is usually created via a supervised process.

- A heuristic PSO-SVM framework is presented in this dissertation that integrates feature selection and SVM learning processes, which provides a unified and efficient framework for selecting the best feature subset, reducing the dimensionality of the feature space, and optimizing classification performance and the generalization ability of a supervised learning machine.

- This dissertation also presented two different feature search techniques. The first approach used a univariate feature ranking and outweighed cross-correlation criteria to form potential feature subsets. In the second approach, a binary heuristic search using PSO algorithm was used to construct the feature search space.
• A novel criterion for evaluation of learning models and selection of the best features subset, which used the sensitivity of generalization performance of the SVM classifier to the variation of the kernel’s control parameter and the classifier’s regularization constant as a final criterion for selecting the best feature subset.

• Diagnosis of MCs using their shape features is sensitive to the segmentation of MCs. This segmentation can be challenging and difficult when MCs are present in a dense breast tissue of young women. An alternative method to the shape analysis of MCs is the use of texture features, which can be obtained by analyzing the gray-level histogram and spectral representation of MCs.

Also, in our work on the feature extraction, and performance evaluation I had come to the following conclusions:

• A CADx scheme proposed in this dissertation utilizes a cluster’s ground truth file that represents the location and the size of MC cluster to automate the region selection and to improve the segmentation of MCs.

• A new morphological top-hat transform is proposed, which was used to construct a dual morphological filtering scheme to segment MCs.

7.3 Future work

For future work, our ongoing plan is to validate the proposed feature extraction, feature selection, and the PSO-SVM full model methods presented in Chapter V of this dissertation using a larger mammogram dataset. Other improvements and extensions of the work presented in this dissertation are summarized as follows:
For the detection approach, a simple thresholding using Otsu's method has been used to reduce the high number of false positive results obtained from applying the proposed CADe approach to a full mammogram. This process can be further improved by employing other techniques such as fuzzy K-mean clustering, to extract the glandular breast regions and thus eliminate many of the misclassifications that are resulting at the breast borders and a radiographic marker. Another improvement would be is to use a more sophisticated modeling of MCs and background structures, which should be useful to overcome the impact-on the results-from estimating the parameters of the Bayesian classifier.

For the shape based diagnosis scheme, suggested future works may include testing the segmentation and shape analysis methods on higher resolution digital mammograms and integrating the design of segmentation and feature extraction stages within the full model selection process. Extension of the proposed PSO-SVM embedded feature selection method for other feature extraction and classification techniques is a potential future work.

The combination of different feature extraction techniques (e.g. shape, texture, and spectral) to provide a final and better characterization of the MC malignancy is an interesting extension of the work presented in this dissertation. Such combinations can be accomplished via a decision fusion strategy and a committee of learning machines.

I believe that via a suitable design of the parameters of the Hessian filter and by using a new filtering scheme to construct the Hessian matrix, I can improve the discriminative power of the extracted texture features. Moreover, the classification performance can be significantly improved by using a non-linear SVM.
BIBLIOGRAPHY


Lindeberg, T. (1994), In scale space theory in computer vision, Kluware academic publishers.


Piscataway, NJ, USA.


Artificial Intelligence in Medicine, 34(2),141-150.


