Mass Detection and False Positive Reduction in Mammographic Images

A Thesis submitted for the degree of

Doctor of Philosophy

By

Man To Wong

In

Faculty of Engineering and Information Technology

School of Computing and Communications

UNIVERSITY OF TECHNOLOGY SYDNEY

AUSTRALIA

Submitted DECEMBER 2015

UNIVERSITY OF TECHNOLOFY SYDNEY
SCHOOL OF COMPUTING AND COMMUNICATIONS

The undersigned hereby certifies that he has read this thesis entitled "**Mass Detection and False Positive Reduction in Mammographic Images"** by **Man To Wong** and that in his opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy.

Principal Supervisor

Prof. Xiangjian (Sean) He

CERTIFICATE OF AUTHORSHIP/ORIGINALITY

I certify that the work in this thesis has not been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Author

Abstract

Mass Detection and False Positive Reduction in Mammographic Images

Breast cancer is the most common type of cancer for women in America. Currently the most effective method for early detection of breast cancer is mammography. Mammography is the only widely accepted imaging method used for routine breast cancer screening. Masses are one of the important signs of breast cancer. However it is difficult to detect masses because masses have different size and shape and their features can be obscured or similar to the normal breast parenchyma. Reading mammograms is a demanding job for radiologists. A computer aided detection (CAD) system can provide a consistent second opinion to a radiologist and greatly improve the mass detection accuracy.

In this thesis, a computer aided detection system is developed which can segment the breast region from the background in the whole mammographic image, detect the suspicious regions from the breast region and then classify the suspicious regions to mass or normal breast tissue. The suspicious regions in the full mammographic image can be found by contrast limited adaptive histogram equalization and thresholding. These suspicious regions can be masses or normal breast tissue (false positives). To reduce the number of false positives in mass detection, a feature selection and classification approach using particle swarm optimization (PSO) and support vector machine (SVM) is proposed. Firstly, texture features are derived from the gray level co-occurrence matrix (GLCM) of each suspicious region. A PSO and SVM based feature selection is proposed to determine the significant features. The significant features found by PSO-SVM based feature selection are used by the SVM classifier to classify the suspicious region to mass or normal breast tissue.

One advantage of the proposed mass detection system is that it can detect different types of masses, including spiculated, circumscribed and ill-defined masses from the whole mammographic image. The number of false positives in mass detection can be reduced by the PSO and SVM based feature selection and mass classification method proposed

in this thesis. Experimental results show that the proposed PSO-SVM based feature selection technique can find the significant features that can improve the classification accuracy of SVM and perform better than other widely used feature selection methods. The proposed mass classification approach using PSO and SVM has better or comparable performance when compared to other state-of-the-art mass classification techniques, using sensitivity and specificity as the evaluation criteria.

In order to perform accurate image segmentation of the mass from the suspicious region, a mass segmentation method by PSO based image clustering is proposed. Two new fitness functions are proposed which can improve the performance of image clustering by generating more compact clusters and larger inter-cluster distance. The proposed PSO based image clustering, with the new fitness function, can improve the segmentation of the mass from mammographic image. It has been shown experimentally that PSO based image clustering can have better mass segmentation performance when compared to K-means, a widely used clustering technique.

Acknowledgement

This research would not have been possible without the guidance and the help of many people. My gratitude to my supervisor, Prof. Xiangjian He, for his guidance, support and encouragement. His comments and suggestions during preparation of this thesis have been valuable.

I also appreciate the following people for providing various assistance for the completion of this research work: Zhiyuan Tan, Aruna Jamdagni, Chao Zeng, Ruo Du, Muhammad Abul Hasan, Sheng Wang, Prof. Hung Nguyen, Dr Wenjing Jia, Dr Qiang Wu, Dr Min Xu and Prof. W.C Yeh.

Last but not the least, I would like to express my love and gratitude to my daughter Katherine for her understanding and tolerance so that I can concentrate on this thesis.

Author's Publication for the Ph.D

Published papers

Conference papers

1. M. T Wong, X. He, W. C Yeh, Z. Ibrahim, and Y. Y Chung. "Feature Selection and Mass Classification Using Particle Swarm Optimization and Support Vector Machine". In 21st International Conference on Neural Information Processing (ICONIP 2014), Part III, Lecture Notes in Computer Science, Vol. 8836, pages 439-446, Springer International Publishing, Kuching, Malaysia, November 3-6, 2014. (Tier A Conference)
2. M. T Wong, X. He, H. Nguyen, and W.C Yeh. "Mass Classification in Digitized Mammograms Using Texture Features and Artificial Neural Network". In 19th International Conference on Neural Information Processing (ICONIP 2012), Part V, Lecture Notes in Computer Science, Vol. 7667, pages 151-158, Springer-Verlag, Berlin, Heidelberg, Doha, Qatar, November 12-15, 2012. (Tier A Conference)
3. M. T Wong, X. He, H. Nguyen, and W. C Yeh. "Particle Swarm Optimization Based Feature Selection in Mammogram Mass Classification". In 2012 International Conference on Computerized Healthcare (ICCH), pages 152-157, Hong Kong, December, 2012.
4. M. T Wong, X. He, and W. C Yeh. "Image Clustering Using Particle Swarm Optimization". In 2011 IEEE Congress on Evolutionary Computation (CEC), pages 262-268, New Orleans, USA, June, 2011. (Tier A conference)

Table of Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

Breast cancer is the most common cancer among women in Australia [1]. According to Cancer Council Australia, the risk of being diagnosed with breast cancer by the age 85 is 1 in 8 for women [1]. A recent study by the American Cancer Society shows that between one in eight and one in twelve women in the United States will develop breast cancer during their life-time [2]. Mammography is currently the most effective method in the early detection of breast cancer [3]. Reading mammograms is a demanding job for radiologists. Even expert radiologists can miss a significant portion of abnormalities [4,5]. Computer Aided Detection (CAD) system can be used as a second reader. By using CAD detection systems as an aid, the radiologists' accuracy of detection of breast cancer has been improved [6,7,8].

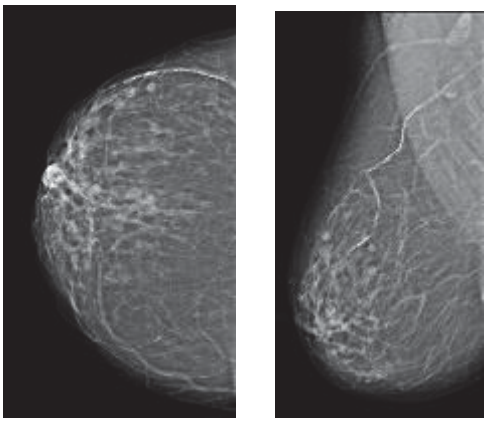
This chapter will cover the basic concept of mammography, different types of mammography, importance of computer aided detection, the objectives and outline of this thesis.

1.1 Introduction to Mammography

There are several imaging techniques for examination of the breast, including magnetic resonance imaging, ultrasound imaging, and X-ray imaging [9]. Mammography is a specific type of imaging that examines the breast by using a low-dose X-ray system. Mammography is currently the most effective method for early detection of breast

cancer. Currently it is the only widely accepted imaging method for routine breast cancer screening [9].

There are two types of mammography: screening mammography and diagnostic mammography. Screening mammography is performed to detect breast cancer in an asymptomatic population [10]. Screening mammography generally consists of four views: with two views of each breast: the craniocaudal (CC) view (top to bottom view) and the mediolateral-oblique (MLO) view (side view taken at an angle). Figure 1.1 shows the CC view and MLO view of a mammogram [11].



(a) CC view (b) MLO view

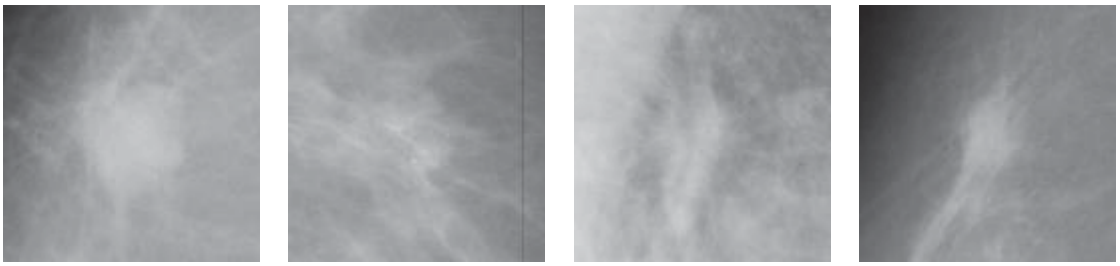
Figure 1.1: CC view and MLO view of mammogram [11].

The purpose of diagnostic mammography is to examine a patient who has already demonstrated abnormal clinical findings. Similar to screening mammography, each breast examined using diagnostic mammography may also have two views. Diagnostic mammography is often performed as a follow-up examination of an abnormal screening mammography to determine whether area of concern on the screening examination needs additional breast imaging or a biopsy to determine whether the woman has breast cancer [10].

It is a demanding task for radiologists to interpret the results of mammograms because mammograms generally have low contrast. According to research in [12], double reading of mammograms can provide effective screening and high sensitivity. Double reading means that two radiologists will read the same mammograms. However, the workload and costs for double reading are high. Computer aided detection (CAD) can be used to aid the radiologists to interpret the mammogram. With the help of CAD, only one radiologist is needed to read the mammogram and the CAD can provide the second opinion to the radiologist. Previous research by Brem et al. [13] indicates that the use of CAD significantly improved the detection of breast cancer by increasing the radiologist's sensitivity by 21.2%.

1.2 Abnormalities in Mammogram

There are many different types of mammographic abnormality which include masses, micro-calcifications, architectural distortion and asymmetry [14] (see Figure 1.2).



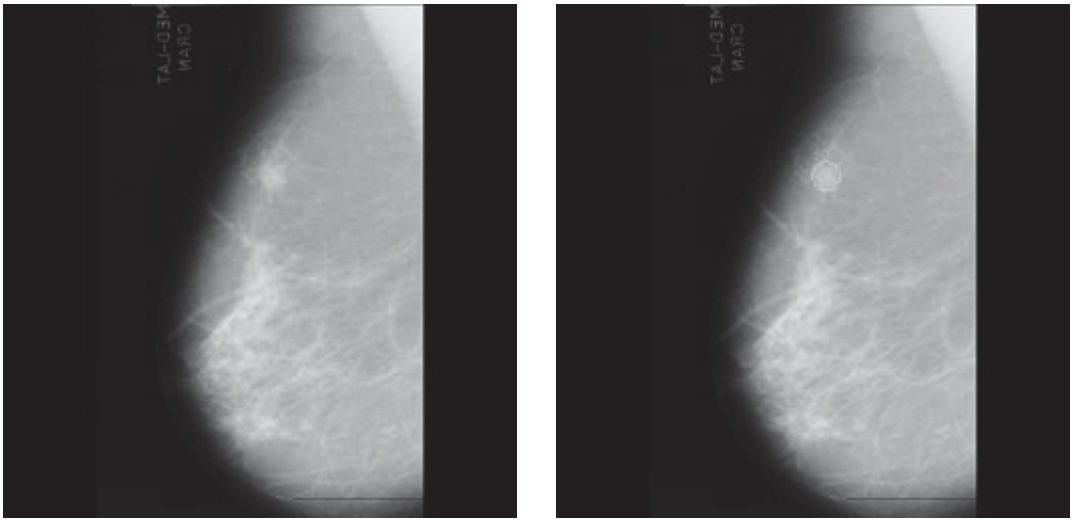
(a) Mass (b) Calcifications (c) Architectural distortion (d) Asymmetry

Figure 1.2: Different Types of mammographic abnormality.

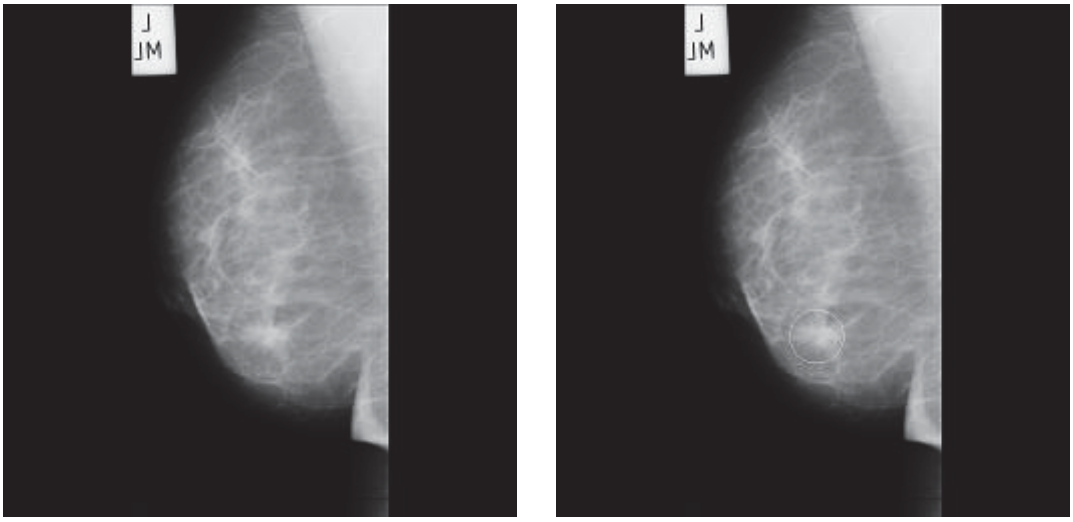
A breast mass is a generic term to indicate a localized swelling or lump in the breast. A mass is defined as a space-occupying lesion seen in at least two different projections [15]. Masses are characterized by their shape and margin characteristics [16,17]. Calcifications are tiny deposits of calcium, which appear as small bright spots on the mammogram. They are characterized by their type and distribution properties

[16]. Micro-calcifications usually form clusters and individual micro-calcifications can range from twenty to several hundred microns in diameter [18]. For architectural distortion, the normal architecture is distorted with no definite mass visible. This includes spiculations radiating from a point, and focal retraction or distortion of the edge of the parenchyma [15,16]. Although there are many types of abnormalities, in a majority of cases, the abnormalities are either micro-calcifications or masses [18]. Detection of mammographic abnormality is difficult because there is a large variability in the appearance of abnormalities and abnormalities are often occluded or hidden in dense breast tissue. Mass detection is generally more difficult than the detection of micro-calcifications because the features of mass can be obscured or similar to normal breast parenchyma [19].

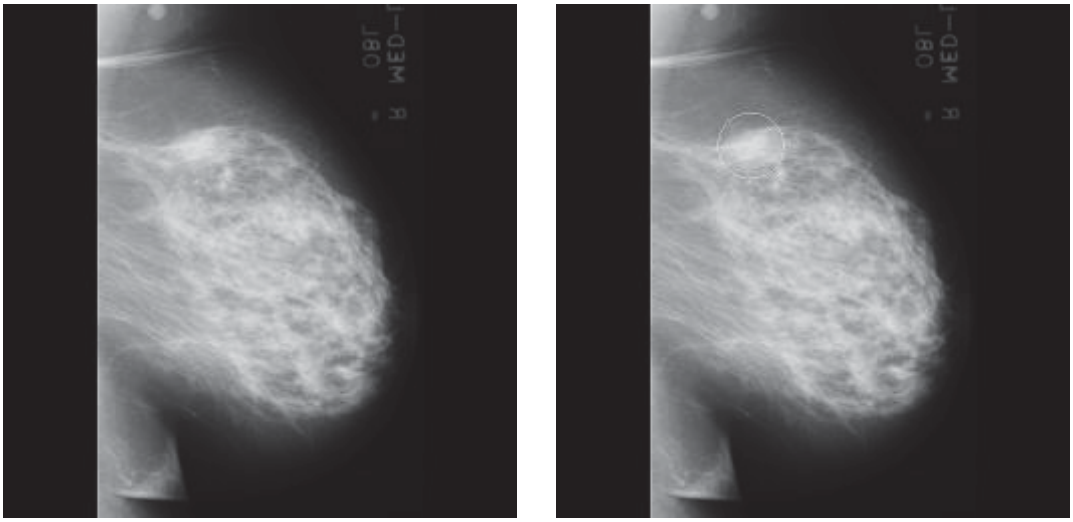
There are three types of masses in a mammogram: circumscribed, spiculated and ill-defined masses [14]. Figure 1.3 shows the three types of masses. The images are taken from the mini-MIAS mammographic database and the characters in bracket in Figure 1.3 represent the filename of the mammogram image in the mini-MIAS database. The left hand side of the figure shows the original image in the mini-MIAS mammogram database while the right hand side shows the image with the mass enclosed in a white circle. The center of the circle represents the centroid of the mass and the radius of the circle represents the size of the mass. The center and radius of the circle are provided as ground truth information in the mini-MIAS database.



(a) Circumscribed mass (mdb023)



(b) Spiculated mass (mdb181)



(c) Ill-defined mass (mdb032)

Figure 1.3: Three types of mass.

1.3 Computer Aided Detection

Computer-aided detection (CAD) systems have been used to help radiologists in detecting mammographic lesions that may indicate the presence of breast cancer. The metrics used to report the performance of detection algorithms are sensitivity and the number of false positives per image (FPI). Sensitivity is defined as the number of true positive marks divided by the number of lesions. A true-positive mark is a mark made by the CAD system that corresponds to the location of a lesion. A false-positive mark is a mark made by the CAD system that does not correspond to the location of a lesion [16]. Most detection algorithms consist of two stages. In the first stage, the aim is to detect suspicious lesions at a high sensitivity. In the second stage, the aim is to reduce the number of false positives without a significant decrease in sensitivity. The first stage is designed to have a very high sensitivity and a large number of false positives are acceptable since stage 2 can be used to reduce the number of false positives. It should be noted that regions labeled as suspicious by the detection algorithms are not necessarily malignant. The classification of detected regions into malignant or benign categories is a different problem. The purpose of a CAD system is to identify the suspicious regions which can be breast cancer lesions or normal tissue. The output of a CAD system can be marks or regions of interest [15]. Also CAD systems act only as a second reader and the final decision is made by the radiologist. Recent research has shown that CAD detection systems, when used as an aid, have improved radiologists' accuracy of detection of breast cancer [13]. As described in earlier section, there are many types of abnormalities but in a majority of cases, the abnormalities are either micro-calcifications or masses [18]. Mass detection is generally more difficult than the detection of micro-calcifications [19]. This thesis will cover mass detection, false positive reduction and mass segmentation. The proposed false positive reduction and

mass segmentation algorithms are developed by using particle swarm optimization (PSO).

1.4 Contributions

Computer-aided detection (CAD) systems have been used to help radiologists in detecting mammographic lesions that may indicate the presence of breast cancer. Research in mammographic mass detection has been done for many years. However much previous research has been done on a single type of mass. A method is proposed in this thesis which can detect different types of mass from a mammographic image. Also the proposed mass detection technique does not depend on the size, shape and boundary of the masses.

Another contribution is to reduce the number of false positives in mass detection. A particle swarm optimization (PSO) based feature selection is proposed to find the significant features that can improve classification accuracy. The proposed PSO based feature selection method has been shown to perform better than or similar to other conventional feature selection methods when it is applied in mass classification. Finally this thesis demonstrates how PSO can be used to improve false positive reduction in mass detection and how PSO based image clustering can improve mass segmentation. This thesis has shown that PSO based feature selection can find significant features that can improve mass classification accuracy. The mass segmentation approach using PSO based clustering can have better performance than mass segmentation using k-means clustering.

Another main contribution of this thesis is to show how particle swarm optimization (PSO) can improve the performance of mass detection, classification and

segmentation. A PSO based feature selection method is proposed to select significant features for mass classification which can be used to improve mass classification accuracy and reduce the number of false positives in mass detection. For mass segmentation, a PSO based image clustering technique with new fitness functions is used. Experimental result shows that PSO based image clustering has better mass segmentation performance when compared to k-means clustering.

1.5 Organization of This Thesis

The organization of this thesis is as follows. This chapter describes the concepts of mammography and abnormalities in a mammogram. My contributions to this thesis are also described. Chapter 2 provides a review of previous research work in mass detection and false positive reduction. Chapter 3 describes the mass detection method used in this thesis. Chapter 4 shows how particle swarm optimization (PSO) can be used in parameter tuning, feature selection and mass classification. Chapter 5 shows how mass segmentation can be done by using PSO based image clustering. Finally Chapter 6 contains the conclusion of this thesis.

CHAPTER 2

Review of Mass Detection and False Positive Reduction Techniques

2.1 Introduction

A mass is defined as a space-occupying lesion seen in at least two different projections [15]. A breast mass is a generic term to indicate a localized swelling, protuberance, or lump in the breast [18]. Radiologists characterize masses by their shape and margin properties. A number of researchers have worked on different methods for detecting masses in mammograms. Masses with spiculated margins have a very high likelihood of malignancy and thus some methods have been developed specifically for the detection of spiculated masses. A spiculated mass is characterized by lines radiating from the margins of a mass [20]. A Computer-Aided Diagnosis (CAD) system is a set of automatic or semi-automatic tools developed to assist radiologists in the detection and/or classification of mammographic abnormalities [18,21,22]. The use of CAD in the interpretation of screening mammograms can increase the detection of early-stage malignancies [22]. However, the main drawback of CAD systems is the significant number of false positive detections [23].

The first part of this chapter is focused on the review of mass detection techniques. Some previous research publications distinguish between mass detection and mass

segmentation algorithms. Mass detection is defined as the identification of potential lesions within the parenchymal background. Usually these methods generate a marker or prompt at a suspicious region in a mammogram. Mass segmentation is defined as a method able to detect the precise outline of the potential lesion [18]. However, some algorithms have been developed that can develop mass detection and segmentation at the same time. According to research by A. Oliver et al. [18], there are three possible outputs for mass detection/segmentation algorithms:

- detection and/or segmentation of potential lesions;
- classification of detected lesion as mass or not mass (also referred as false positive reduction algorithms);
- classification of lesions as benign or malignant mass.

In this thesis, mass detection and segmentation, and false positive reduction algorithms will be covered. While the proposed mass detection approach in this thesis can perform both mass detection and segmentation, the priority in this thesis is to perform accurate mass detection. Once the potential lesions have been found, mass segmentation can be applied to a window surrounding the suspicious region to obtain accurate mass boundary. A PSO based clustering algorithm is proposed in this thesis to perform mass segmentation.

To improve the mass detection accuracy, false positive algorithms can be applied to classify the potential lesions as mass or not mass. In this thesis, a PSO based feature selection and classification approach is used for false positive reduction. By using the significant features found by PSO based feature selection instead of using all features, it can improve the classification accuracy of mass detection.

In section 2.2, a review of mass detection techniques is provided. In the first stage of mass detection, the main objective is to detect suspicious regions on the mammogram with high sensitivity and a large number of false positives can be accepted [18]. In the second stage, the suspicious region will be classified as mass or normal tissue. The purpose is to reduce the number of false positives. A review of false positive reduction techniques is given in section 2.3.

2.2 Review of Mass detection techniques

Most mass detection algorithms consist of two stages [16]:

- stage 1: detection of suspicious regions on the mammogram,
- stage 2: classification of suspicious regions as mass or normal tissue.

The objective of the first stage is detection of mass with very high sensitivity. A large number of false positives is acceptable in the first stage since they are expected to be removed by false positive reduction in stage 2. According to M. P. Sampat et al. [16], mass detection algorithms for stage 1 detection can generally be considered to be of two types, pixel based or region based. They are discussed in sections 2.2.1 and 2.2.2 respectively.

2.2.1 Pixel-based Detection Methods

In pixel-based methods, features are extracted for each pixel and they are then classified as suspicious or normal. Features are extracted from the local neighborhood of each pixel. In the next step pixels are classified as suspicious or not. This may be done by simply applying a threshold to the feature image or by using sophisticated classification techniques. Finally, suspicious pixels are grouped together into regions, generally by

collecting connected pixels [16]. It should be noted that regions labeled as suspicious by the detection algorithms are not necessarily malignant. The purpose of mass detection is to find the locations of the suspicious regions.

Some detection methods are developed for a special type of mass. As spiculated masses have high chance of malignancy, many researchers concentrated on spiculated masses. Since spiculated masses are characterized by spicules radiating in all directions, edge orientations at each pixel are often used for mass detection. Each pixel is represented by a feature vector that represents the strongest edge orientation at the pixel.

Kegelmeyer et al. [24] used the standard deviation of a local edge orientation histogram (ALOE) and the output of four spatial filters that are a subset of Laws texture features to detect speculated masses. The idea of using the ALOE feature is that a normal mammogram exhibits a tissue structure that radiates in a particular orientation. A spiculated mass would change this trend and thus normal tissue would have edge orientations in a particular direction, whereas in suspicious regions containing spiculated lesions, edges would exist in many different orientations. Karssemeijer and te Brake [25] detected stellate distortions by a statistical analysis of a map of pixel orientations. The orientation at each pixel was computed from the response of three filter kernels, which are second-order, directional derivatives of a Gaussian kernel in three different directions. These filters form a non-orthogonal basis. They used the relation that at a particular scale, the output at any orientation can be expressed as a weighted sum of the responses of the filters. This relation was used to determine the orientation at each pixel and two features for each pixel were derived by a statistical analysis of these pixel orientation maps. The pixels were then classified as suspicious or normal.

Liu et al. [26] point out that in general, it is difficult to estimate the size of the neighborhood that should be used to compute the local features of spiculated masses. To address this problem Liu et al. [26] developed a multiresolution algorithm for the detection of spiculated masses. They generated a multiresolution representation of a mammogram using the discrete wavelet transform. They extracted four features at each resolution for each pixel. The detection was carried out in a top-down manner from the coarsest resolution to the finer resolutions. If a positive detection was made and a pixel was classified as abnormal, no feature extraction and detection were needed at the corresponding pixels at all finer resolutions

Other researchers have also worked on other types of masses. Li et al. [27] developed a two-step process for detection of masses. In the first step, adaptive gray-level thresholding was used to obtain an initial segmentation of suspicious regions. The segmentation was iteratively improved using a multiresolution Markov random field (MRF)–based segmentation method. The algorithm was first applied at the coarsest resolution and the output was refined at the next finer resolution. A fuzzy binary decision tree was used to classify the segmented regions as masses or normal tissue using features based on shape, region size, and contrast.

Matsubara et al. [28] developed an adaptive thresholding technique for mass detection. Histogram analysis technique was used to divide mammograms into three categories ranging from fatty to dense tissue. Potential masses were detected using multiple threshold values based on the category of the mammogram. The number of false positives was reduced by using features such as circularity, area, and standard deviation.

Li et al. [29] developed a method for lesion site selection using stochastic model–based segmentation technique. A finite generalized Gaussian mixture distribution was

used to model histograms of mammograms. The expectation maximization algorithm [30] was used to determine the parameters of the model. The segmentation was achieved by classifying pixels using a new Bayesian relaxation labeling technique.

The advantage of using pixel-based methods is that one has a large number of samples to train a classifier. However, the disadvantage of pixel-based methods is that the spatial arrangement of the pixels is not considered and this is a very important factor to discriminate masses from normal tissue. A different set of features would be required to describe different mass types. It is computationally intensive. Most pixel-based methods must subsample images before detection [16].

2.2.2 Region-based Detection Methods

In region based detection methods, regions of interest are first extracted by a segmentation or filtering technique. Features are then extracted for each region and the region is classified as suspicious or normal. These features are designed to describe important diagnostic information like shape and texture of the extracted regions

A number of these methods are based on the idea of matched filtering. In these approaches, the filter is used as a model for a mass. The output of the filtered image will be high near the center of the tumor masses. Often the N largest outputs are selected as possible suspicious regions. This is followed by the extraction of ROIs around the N largest peaks. Features are extracted from the ROI, and the ROIs are classified as containing a mass or normal tissue. In the region-based methods, features are extracted for each region.

Brzakovic et al. [31] use a two-stage multiresolution approach for detection of masses. First they identified suspicious ROIs using Gaussian pyramids and a pyramid

linking technique based on the intensity of edge links. Edges were linked across various levels of resolution. This was followed by a classification stage, where the ROIs were classified as malignant, benign, or normal on the basis of features like shape descriptors, edge descriptors, and area.

Petrick et al. [32] developed a two-stage algorithm for the enhancement of suspicious objects. In the first stage, they proposed an adaptive density-weighted contrast-enhancement (DWCE) filter to enhance objects and suppress background structures. The central idea of this filtering technique was that it used the density value of each pixel to weight its local contrast. In the first stage, the DWCE filter and a simple edge detector (Laplacian of Gaussian) were used to extract ROIs containing potential masses. In the second stage, the DWCE was reapplied to the ROI. Finally, to reduce the number of false positives, they used a set of texture features for classifying detected objects as masses or normal. They further improved the detection algorithm by adding an object-based region growing algorithm [33].

Polakowski et al. [34] used a single difference of Gaussian (DoG) filter to detect masses. The DoG filter was designed to match masses that were approximately 1 cm in diameter. ROIs were selected from the filtered image. They used nine features based on size, contrast, circularity and Laws texture features to reduce the number of false positives and to then classify ROIs as malignant or normal.

The DoG filter, which is a band-pass filter, has been used by several researchers for the preliminary task of detection of potential masses in an image. The DoG filter must be matched to the size of the mass. Since the size of masses varies, a number of DoG filters would be required, which would increase the computational complexity. Since the size of a potential mass is not known a priori, several researchers have used multiscale region-based methods for the detection of masses [16].

Belloti et al. [35] used an edge-based segmentation algorithm to separate the suspicious regions. Second order measures obtained from the co-occurrence matrices were used to describe the texture of the region. Artificial neural network was used in the classification step.

Kobatake et al. [36] modeled masses as rounded convex regions and developed an ‘‘iris filter’’ to enhance and detect masses. The iris filter was applied to a gradient image that was generated by Perwitt-type operators. The output of the filter was computed by measuring the average convergence of the gradient over the region of support of the filter. The peaks of the output of the filter were selected as centers of tumor candidates. The filter was then re-applied locally to detect the boundaries of candidate masses. Finally, texture features were computed from the candidates and were used to reduce false-positives. One of the advantages of using this filter was that the output of the filter would be constant regardless of the contrast between a rounded convex region and the background [16].

Qian et al. [37] developed a multiresolution and multi-orientation wavelet transform for the detection of masses and spiculation analysis. They observed that traditional wavelet transforms cannot extract directional information, which is crucial for a spiculation detection task and thus, they introduced a directional wavelet transform. An input image was decomposed into two output images using the directional wavelet transform. One was a smoothed version of the original image and was used to segment the boundary of the mass. The second contained the high-frequency information and was used for directional feature extraction. The key ideas of the method were that at coarser resolutions, features such as the central mass region can be easily detected, whereas at finer resolutions, detailed directional features such as spicules can be localized.

Some researchers have developed region-based methods that are focused on the detection of masses with particular margin characteristics, such as circumscribed or spiculated masses. Lai et al. [38] developed a simple template matching algorithm to detect circumscribed masses only. They enhanced images using a modified median filtering technique to remove background noise. To cope with variations in the size of masses, templates with various radii were used. Normalized cross-correlation was used to measure the similarity between a potential mass and the template. This particular metric was chosen since it is invariant to the size of the template and the average brightness of the image.

Groshong and Kegelmeyer [39] used the circular Hough transform [40] for the detection of circumscribed lesions. A point in the three dimensional Hough domain maps to a circle in the image domain. They computed an edge image using a canny operator and selected a subset of the edges based on length and intensity. This subset of edges was the input to a circular Hough transform. Two features were extracted from the Hough domain for each pixel and ultimately these were classified as either belonging to a mass or normal tissue.

Zhang et al. [41] noted that the presence of spiculated lesions led to changes in the local mammographic texture. They proposed that such a change could be detected in the Hough domain, which is computed using the Hough transform. They partitioned an image into overlapping ROIs and computed the Hough transform for each ROI. The Hough domain of each ROI was thresholded to detect local changes in mammographic texture and to determine the presence or absence of a spiculated mass.

Region-based methods have a number of advantages. In contrast to pixel-based methods, region-based detection takes into account the spatial information. Also, the features are directly correlated to important diagnostic information like the shape and

margin of extracted regions. They are computationally less intensive than pixel-based methods. The main disadvantage is that if a classifier is used, there are fewer samples for training the classifier as compared to the pixel based methods [16].

2.3 Review of False Positive Reduction Techniques

The purpose of the second stage in mass detection is to reduce the number of false positives produced in the first stage. The suspicious regions are classified as mass or normal tissue. A review of the false positive reduction methods in mass detection is given below.

Te Brake et al. [42] defined a number of features to discriminate between lesions and normal tissue that were designed to capture image characteristics like intensity, iso-density, location, and contrast. Angelini et al. [43] had tested and compared the performance of different image representations for mass classification. Instead of extracting features from the suspicious regions, the features are embodied by the image representation used to encode the suspicious regions. The best result was given by the pixel image representation, using SVM as classifier, with 90% sensitivity and 94% specificity. Christoyianni [44] used the GLCM [45] texture features and MLP and obtained 85% sensitivity and 83% specificity. Wei et al. [46] developed a classifier using texture features and linear discriminant analysis for this task. They computed multiresolution texture features from spatial gray-level dependence matrices. Wei et al. [47] also investigated the use of global and local multiresolution texture features for this task and for reducing the number of false-positive detections on a set of manually extracted ROI. Sahiner et al. [48] proposed a convolution neural network for this task. They extracted texture features from the ROIs. Petrosian et al. [49] used the GLCM

texture features and a modified decision tree classifier and obtained 76% sensitivity and 64% specificity. Kupinski and Giger [50] studied a regularized neural network for this task. Masses were detected using the bilateral subtraction scheme. Features based on geometry intensity and the gradients of potential lesions were extracted. They also evaluated the effectiveness to minimize overtraining. Mutual information and a sub-region hotelling observer have also been tested for this classification problem. Tourassi et al. [51] had applied template matching scheme based on the mutual information and obtained 90% sensitivity and 65% specificity. Junior et al. [52] had applied Diversity Index in a spatial approach to reduce the number of false positives in mass detection. The computed measures are classified by Support Vector Machine (SVM) and the best result of true negative rate is 75.8% and true positive rate is 93.5%. Hussain et al. [53] used the Gabor filter bank to extract the most representative and discriminative textual properties of masses present at different orientations and scales. SVM with Gaussian kernel is used for classification of ROIs as masses and normal tissues. The area under the ROC curve (A_Z) is used as the performance criterion. The best result for A_Z is 0.96.

In this thesis, a mass classification method using particle swarm optimization (PSO) based feature selection is proposed. It can find the significant features in mass classification. By using these significant features instead of the full set of features, better mass classification accuracy can be obtained when compared to other existing mass classification techniques.

CHAPTER 3

Mass Detection Using CLAHE and Thresholding

3.1 Introduction

The purpose of mass detection is to identify the suspicious regions in a mammogram. Reading mammograms is a demanding job for radiologists. Computer Aided Detection (CAD) can be used as a second reader and helps the radiologists to make the final interpretation.

Threshold methods have been widely used for mass detection and/or segmentation. One approach is to obtain an initial rough detection of suspicious regions by thresholding and, in a subsequent step, the result is refined by using topological analysis [54]. Most of the thresholding algorithms are based on the grey-level value of the actual mammogram. For instance, Abdel-Dayem and El-Sakka [55] found the best threshold to detect masses based on minimizing the global fuzzy entropy of the image. In contrast, Matsubara et al. [56,57] used different threshold values depending on the type of tissue of the breast, which was analysed using histogram analysis. Özekes et al. [58] used directional thresholding to estimate the shape of the mass. The pixels of the mammogram were scanned in eight directions using various thresholds. Subsequently, a (circular) mass template was used to categorise the region as being a true mass. Sameti et al. [59,60] introduced neighboring information into the thresholding algorithm. They

first divided the image into a set of ROIs and, subsequently, a fuzzy membership was given to each pixel of the ROI. In each iteration an error value was calculated, updating also each membership value. In this process they took neighbouring values into account.

In other cases the thresholding is not applied directly to the mammographic image, but to an enhanced version of the original image. Ball et al. [61] applied thresholding to a contrast enhanced version of the mammogram. Kobatake and Murakami [62] applied an Iris filter designed to enhance rounded opacities and to be insensitive to thin anatomical structures. Using adaptive thresholding they detected round masses. A logical filter was used by Rocha et al. [63] to enhance the edges of the suspicious region. Subsequently, a thresholding value was found by using histogram information to find the edge of the lesion.

Instead of enhancing the image using filters, a different approach is to first extract some features from the image and threshold them in a posterior step. For instance, Heath and Bowyer [64] developed a new mass detection algorithm which was based on an Average Fraction Under the Minimum (AFUM) filter, which was designed to find the degree to which the surrounding region of a point radially decreases in intensity. The final step was to threshold the image to identify suspicious regions. Another example is the work of Gupta and Undrill [65] and Undrill et al. [66] who thresholded the mammograms by using Laws masks. Laine et al. [67] found the best scale to segment the lesions by using wavelet decomposition [68]. Another common approach is to threshold the result of an image subtraction. Polakowski et al. [69] found the edges in the image by subtracting two smoothed versions of the original mammogram, and subsequently thresholded this image. Two Gaussian filters with different standard deviation were used to smooth the original mammogram. Instead of subtracting two

blurred images of the mammogram, Kom et al. [70] subtracted from the original image a linearly transformed enhanced image.

Many researchers only focused on one particular kind of mass, e.g. circumscribed lesions [38,39] and spiculated masses [25,26]. This thesis will propose a method which can detect different types of masses: circumscribed, speculated and ill-defined masses. The proposed mass detection method can be used for mass with any size and shape. It consists of four stages:

- breast region extraction,
- contrast limited adaptive histogram equalization (CLAHE),
- thresholding after CLAHE,
- false positive (FP) reduction by area and shape measure.

In the fourth stage, FP reduction is done by using area and shape measure. In Chapter 4 of this thesis, a particle swarm optimization (PSO) based selection and classification technique is proposed to further reduce the number of false positives. The details of the four stages of the proposed mass detection method are given in the following sections.

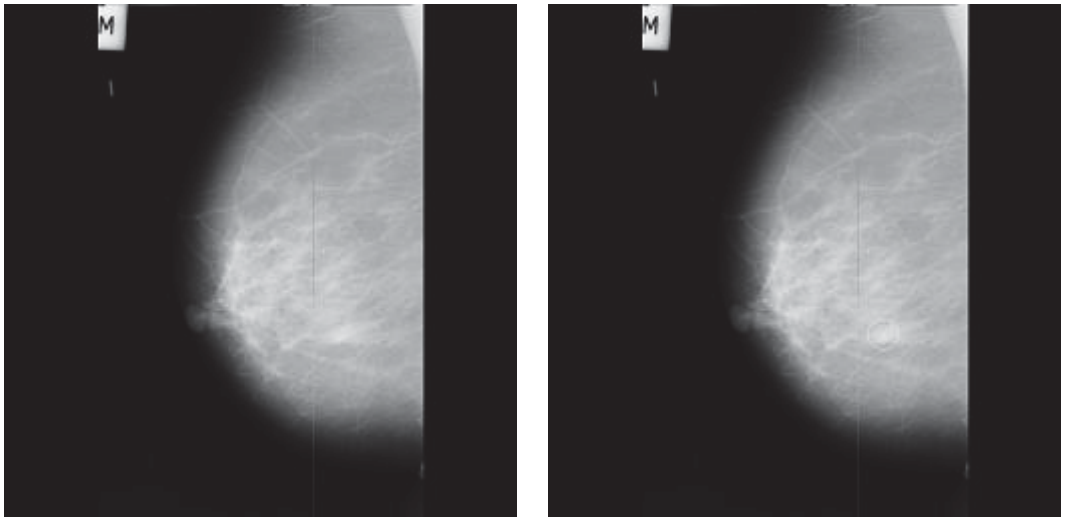
3.2 Breast Region Extraction

There are two main areas in a full mammographic image: breast region and non-breast region. The non-breast region can contain dark background and background objects which can include character labels and some artifacts. Some mammograms in mini-MIAS database contain background objects with intensity higher than the breast region, or even higher than mass region, as shown in Figure 3.1(a). They may deteriorate the results of raw ROI extraction and final segmentation if they are segmented together with the breast region. Hence these background objects have to be removed before mass detection. The intensity threshold value that separates the foreground regions (regions

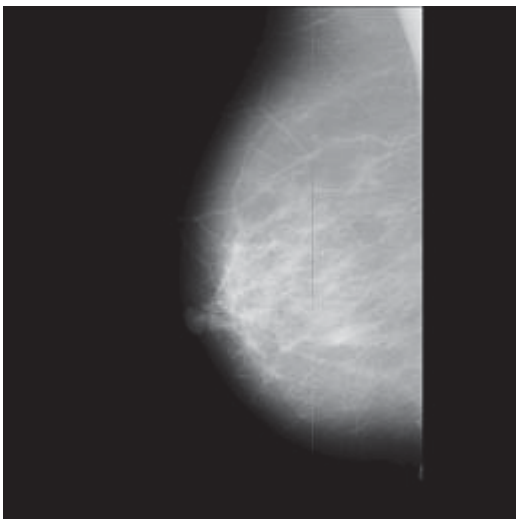
with high intensity) from the dark background has to be identified. After thresholding, the area of each segmented object is found by counting the number of pixels in each object. As the breast region is the major region in the mammogram that occupies the largest area compared with the regions of other segmented objects, the segmented region with largest area was selected as breast region.

T. S. Subashini et al. [71] have used global thresholding to convert the mammographic image to the binary image and then perform breast region extraction. In this thesis, similar to Subashini's work, a binary image is created from the mammographic image by thresholding. The threshold gray level value used is 14 in this thesis. After thresholding, the bright intensity object (with gray level value 255) with the maximum area will be the breast region and the non-breast region will contain some background objects such as character labels and artifacts. By choosing the bright object with the largest area after thresholding, the breast region can be segmented from the mammographic image.

The image in Figure 3.1(a) is taken from the mini-MIAS database [14]. In Figure 3.1 (b), the white circle represents the ground truth data provided by the mini-MIAS database. The center of the circle represents the centroid of the mass. The circle will completely enclose the mass, with some normal breast tissue inside the circle. Both the center and radius of the circle are provided in the ground truth file of the mini-MIAS database.



(a) Original mammographic image (b) Image with mass ground truth circle



(c) ROI image after breast region extraction

Figure 3.1: Breast region extraction.

After breast region extraction, it can be seen from Figure 3.1 (c) that only the breast body remains. The character labels in the original image have been removed.

3.3 Contrast Limited Adaptive Histogram Equalization

Adaptive histogram equalization (AHE) is a computer image processing technique used to improve contrast in images. Ordinary histogram equalization uses the same transformation derived from the image histogram to transform all pixels. This works well when the distribution of pixel values is similar throughout the image. However, when the image contains regions that are significantly lighter or darker than most of the image, the contrast in those regions will not be sufficiently enhanced. Adaptive histogram equalization (AHE) differs from ordinary histogram equalization in the respect that the adaptive method computes several histograms, each corresponding to a distinct section of the image, and uses them to redistribute the lightness values of the image. Adaptive histogram equalization (AHE) transforms each pixel with a transformation function derived from a neighbourhood region. It is suitable for improving the local contrast and enhancing the definitions of edges in each region of an image [72]. However, AHE has a tendency to over-amplify noise in relatively homogeneous regions of an image. A variant of adaptive histogram equalization called contrast limited adaptive histogram equalization (CLAHE) prevents this by limiting the amplification.

Contrast Limited AHE (CLAHE) differs from ordinary adaptive histogram equalization in its contrast limiting. This is achieved by limiting the contrast enhancement of AHE. For CLAHE, the contrast limiting procedure has to be applied for each neighbourhood from which a transformation function is derived. CLAHE was developed to prevent the over-amplification of noise that adaptive histogram equalization can give rise to. The contrast amplification in the vicinity of a given pixel value is given by the slope of the transformation function. This is proportional to the

slope of the neighbourhood cumulative distribution function (CDF) and therefore to the value of the histogram at that pixel value. CLAHE limits the amplification by clipping the histogram at a predefined value before computing the CDF. This limits the slope of the CDF and therefore of the transformation function. The value at which the histogram is clipped, the so-called clip limit, depends on the normalization of the histogram and thereby on the size of the neighbourhood region.

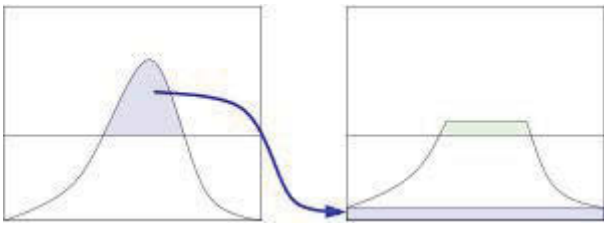


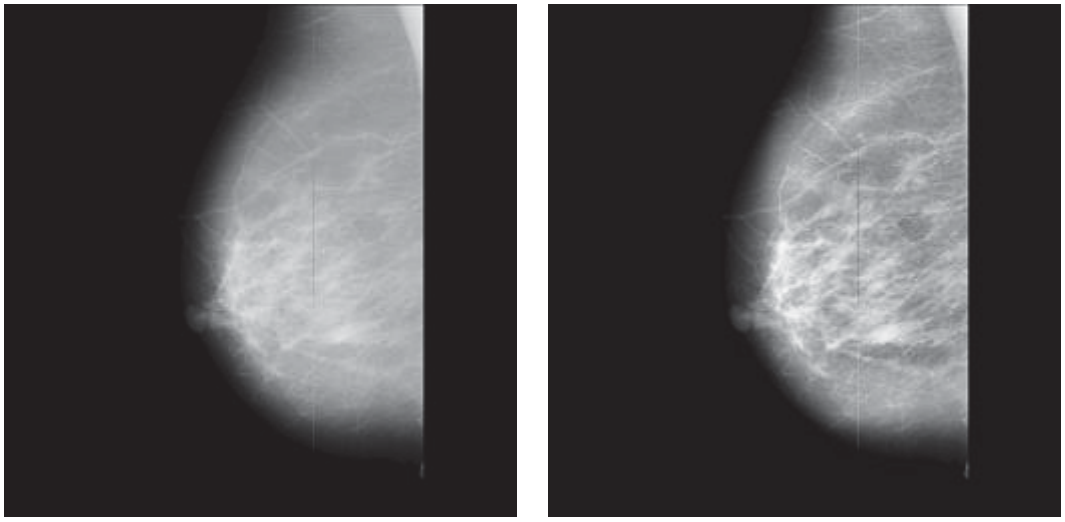
Figure 3.2: CLAHE: redistribute the part of the histogram that exceeds the clip limit equally among all histogram bins. [72]

It is advantageous not to discard the part of the histogram that exceeds the clip limit but to redistribute it equally among all histogram bins (Figure 3.2) [72]. The redistribution will push some bins over the clip limit again, resulting in an effective clip limit that is larger than the prescribed limit and the exact value of which depends on the image. If this is undesirable, the redistribution procedure can be repeated recursively until the excess is negligible [72].

There are two parameters for CLAHE: the threshold for contrast limiting (clipLimit) and size of grid for histogram equalization (tileGridSize). Input image will be divided into equally sized rectangular tiles. The grid size (tileGridSize) defines the number of tiles in row and column [73]. In openCV, the default values for clipLimit is 2 and tileGridSize is (8,8).

The advantages of using CLAHE are that it is easy to use, uses simple calculation, and gives good output in local areas of the image. CLAHE has less noise and it can prevent brightness saturation that commonly happens in histogram equalization [74,75].

A mass is brighter than the surrounding breast tissue though it is not necessary to be the brightest area in the mammogram image. CLAHE will improve the contrast of a mass (see Figure 3.3) and thresholding can be used to extract the mass after CLAHE.



(a) ROI image before CLAHE (b) ROI image after CLAHE

Figure 3.3: ROI image after CLAHE contrast enhancement.

3.4 Mass Detection by Thresholding After CLAHE

After CLAHE, the contrast of the mass will be improved. A thresholding operation is used to extract the mass from the surrounding tissues. There will be some false positives but they can be reduced in the later stage.

Figure 3.4: Image after CLAHE and thresholding.

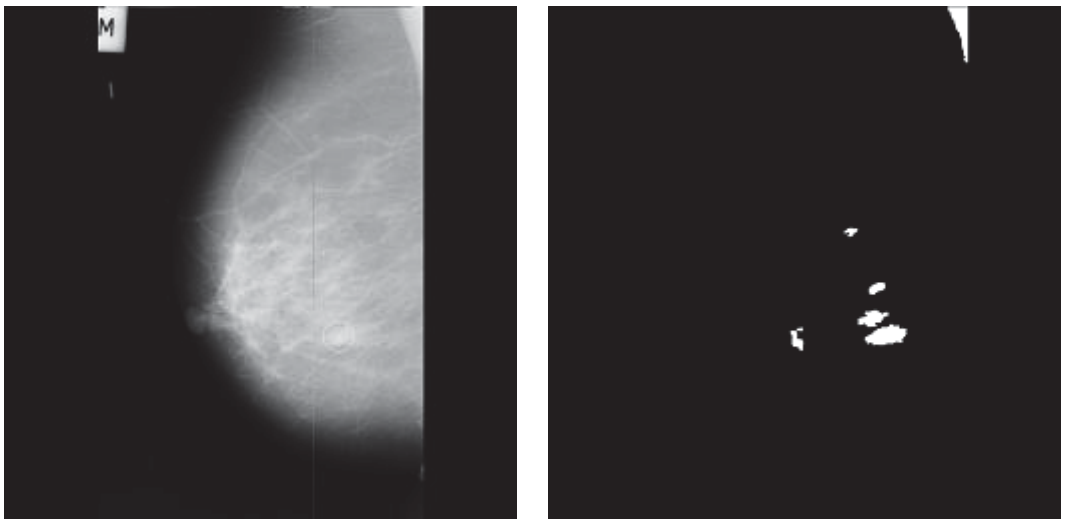
In Figure 3.4, the large white region in the top right hand corner is the pectoral muscle region after thresholding. Connected component labeling can be done by using the two openCV software library functions for C++: findContours and drawContours [76]. After connected component labeling, the perimeter and area of each region can be found by using openCV software library.

In the next stage, the number of false positives can be reduced by removing the bright areas that are too small and those regions that are very long and narrow.

3.5 False Positive Reduction by Area and Shape Measure

To reduce the number of false positives (FP) in Figure 3.4, two steps are used in false positives reduction. In the first step, the number of false positives can be reduced by removing regions with very small areas. In mini-MIAS database, according to its ground truth file, image with filename number mdb206 has the smallest radius of circle enclosing the mass. The number of pixels in the enclosing circle of mass for image mdb206 is used to determine the area threshold. Any suspicious region with area less than this area threshold is not considered as mass. Another feature to reduce the number

of false positives is the shape measure. The shape measure is used to remove very long and narrow regions. The shape measure used is defined by the perimeter squared divided by the area of a region [77]. A threshold for this shape measure has to be chosen that will reject those regions which are very long and narrow. The threshold used for shape measure is 42 and is determined by experiment for the mini-MIAS database. If the shape measure calculated for a suspicious region exceeds 42, then the suspicious region is not considered as mass. Thirty-two images are selected randomly from the Mini-MIAS mammographic database for testing. Every image used for testing has at least one mass according to the ground truth file. In Figure 3.4, after applying false positive reduction by area and shape measure, the mass detection output has one mass and five false positives (see Figure 3.5(b)). Details of the test result are given in section 3.6.



(a) Original image with white ground truth circle

(b) Image after mass detection with false positive reduction

Figure 3.5: Mass detection output after FP reduction

While only the area and shape features are used in this chapter for false positive reduction, in the next chapter (Chapter 4), the number of false positives can be further reduced by using particle swarm optimization (PSO).

3.6 Experimental Result and Discussion

Thirty-two images are randomly selected from the mini-MIAS mammographic database. Each of these selected images has at least one mass, as given by the ground truth file of the database. Of these thirty-two images, two of them contain two masses each. Each of the other thirty images has one mass. The sensitivity (true positive accuracy) of mass detection is 88.2 % and the false positives per image (FPI) is 5.66. The algorithm was implemented by C++ language using OpenCV library [76],

Table 3.1: Summary of output result of various mass detection methods

Method	**True Positive %**	**FPI**
This thesis	88.2	5.66
Li et al. [23]	90	2
Matsubarta et al [24]	82	0.65
Petrick et al. [28]	90	4.4
Belloti et al. [46]	80	4.23
Kobatake et al. [31]	90.4	1.3
Brzakovic et al. [27]	85	–
Groshong et al. [34]	80	1.34
Karsemeijer et al. [21]	90	1
Liu et al. [22]	84.2	1

The sensitivity and false positives per image (FPI) of the proposed mass detection method and other previous research work are compared in Table 3.1. In the column for method, the number in square bracket corresponds to the

reference number in the section References at the end of this thesis. The following points should be noted in Table 3.1.

- Caution should be taken to compare different mass detection techniques using Table 3.1. Different methods may use different mammogram database and different number of images in testing. Even if the same mammogram database is used, different methods may choose different images from the database for testing.
- The proposed mass detection method in this thesis has satisfactory sensitivity when compared to other methods in Table 3.1 but the FPI is a bit higher. It should be noted only simple false positive reduction techniques are used in the proposed method, namely the area and the shape feature. In order to reduce the FPI, a PSO based feature selection and classification technique is proposed in the Chapter 4 to reduce the number of false positives. The proposed PSO based feature selection and classification technique is a general approach and can also be applied to other mass detection methods to reduce the number of false positives.
- Pectoral muscle removal has not been implemented in the proposed mass detection method. The false positives per image can be further reduced if pectoral muscle removal is implemented in the future. For some mammograms, if the pectoral region is very long and narrow, the pectoral muscle region has already been removed by the shape measure used but the simple shape measure proposed in the mass detection method cannot detect and remove all pectoral muscles in the mini-MIAS database.
- The proposed mass detection method can detect three types of masses: spiculated, ill-defined and circumscribed with 88.2% sensitivity and 5.66 false

positives per image. It should be noted that some methods of mass detection in Table 3.1 are only designed to detect one type of mass. More details of mass detection algorithms which are designed to detect one type of mass only are given in Section 2.2 (section on review of mass detection techniques).

- In this thesis, the degree of overlap [78] is used to determine whether the mass detection is successful. The degree of overlap is defined as the area of intersection of segmented object and ground truth mass divided by the area of union of segmented object and ground truth mass. If the ratio is over 0.3, then mass detection is successful. Otherwise the segmented object is regarded as a false positive. The low value of 0.3 is chosen because in mini-MIAS database, the ground truth of the mass is given as a circle which surrounds the mass instead of accurate mass boundary drawn by radiologists [78]. As a mass is not always circular, much of the area enclosed by the circle is normal breast tissue. It should be noted that the value for the degree of overlap (0.3 in this chapter) is used for mini-MIAS database only. For other mammographic database, another value has to be determined.

CHAPTER 4

Mass Classification using Particle Swarm Optimization

4.1 Introduction

Many mass detection algorithms have the following two steps. In the first step, suspicious regions of interest (ROIs) are detected on the mammogram images by using some image processing techniques such as segmentation or thresholding. In the second step, one typical approach is to extract features from the suspicious regions. Classifiers can then be applied on these features to classify the regions as mass or normal tissue. This will reduce the number of false positives. A review of classification of suspicious regions as mass or normal tissue has been given in section 2.3. For the proposed mass detection method in Chapter 3, the area of mass and the shape measure are used to reduce the number of false positives in mass detection. In this chapter, particle swarm optimization (PSO) and support vector machine (SVM) are used to further reduce the number of false positives.

The regions of interests (ROIs) are extracted from the mini-MIAS mammographic database [14]. The ROIs can contain mass or normal tissue. The ROIs will be classified as mass or non-mass regions using texture features calculated from the gray level co-occurrence matrix (GLCM) [45] and statistical features from the gray level histogram.

A PSO-based feature selection technique is proposed to select a smaller subset of significant features which can provide comparable or even better performance when compared to the full set of features. It has been shown experimentally that these significant features can have better sensitivity when compared to the full set of features. Also mass classification using the significant features has better or similar accuracy when compared to other existing mass classification techniques.

Feature selection using PSO has been used widely in different areas including classification problems [79,80] for UCI Repository [81], prediction of company financial crisis [82], gene selection in cancer classification [83] and classification of micro-calcification clusters in mammography [84, 85]. However, application of PSO-SVM based feature selection and mass classification using texture features is not common. A PSO-SVM based feature selection and classification technique will be proposed in this chapter to improve the mass classification accuracy and improve false positive reduction.

4.2 Feature Selection Using PSO and SVM

4.2.1 Traditioal Classifcation Methods

4.2.1.1 Support Vector Machine

Support Vector Machine (SVM) [86,87] is a classifier that has robust and accurate classification performance in many different applications. Classifying data is a common task in machine learning. Suppose some given data points each belong to one of two classes, then the goal is to decide which class a new data point belongs to. In the case of support vector machines, a data point is viewed as a p-dimensional vector (a list of p

numbers), and we want to know whether we can separate such points with a (p-1)-dimensional hyperplane. There are many hyperplanes that might classify the data. One reasonable choice as the best hyperplane is the one that represents the largest separation, or margin, between the two classes. So we choose the hyperplane such that the distance from it to the nearest data point on each side is maximized. If such a hyperplane exists, it is known as the maximum-margin hyperplane and the linear classifier it defines is known as a maximum margin classifier [86]. (see Figure 4.1)

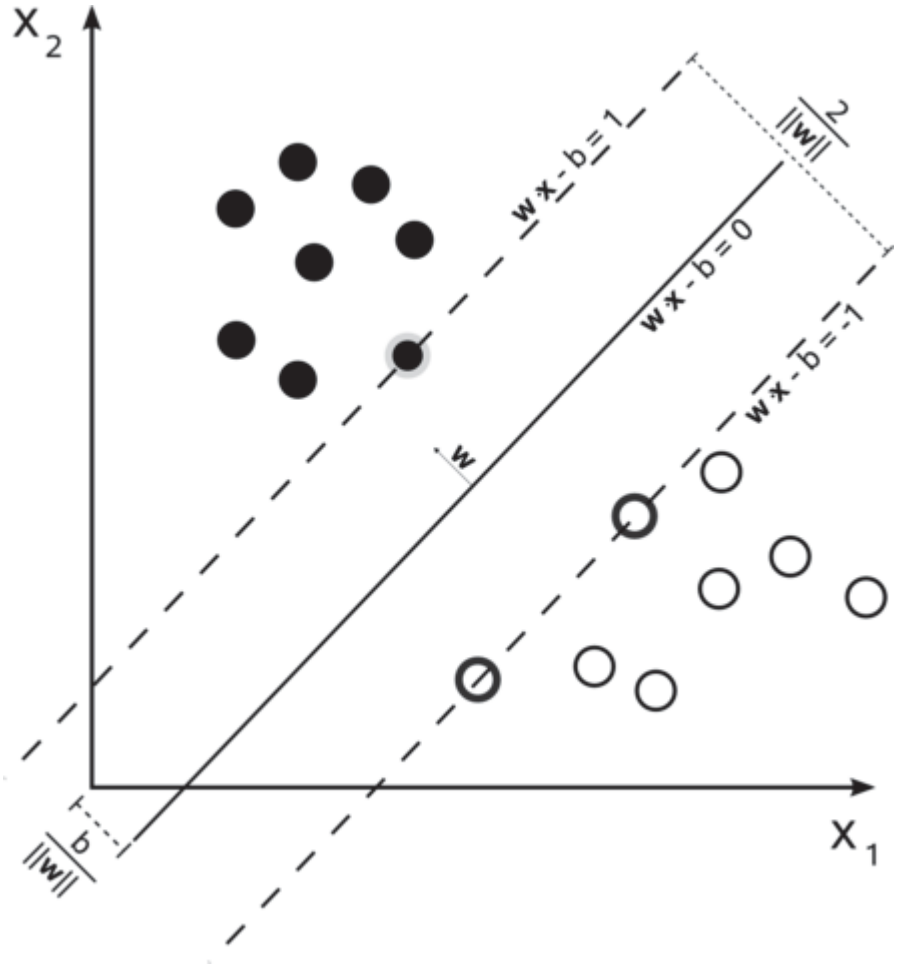


Figure 4.1: Maximum-margin hyperplane and margins for an SVM with samples from two classes [86]

In the linear case, the margin is defined by the distance of the hyperplane to the nearest of the positive and negative examples. The formula for the output of a linear SVM is

$$u = \vec{w}.\vec{x} - b, \tag{4.1}$$

where $\vec{w}$ is the normal vector to the hyperplane and $\vec{x}$ is the input vector. The separating hyperplane is the plane u=0.

If the training data are linearly separable, we can select two parallel hyperplanes that separate the two classes of data, so that the distance between them is as large as possible. The region bounded by these two hyperplanes is called the "margin", and the maximum-margin hyperplane is the hyperplane that lies halfway between them. Geometrically, the distance between these two hyperplanes is $\frac{\mathbf{2}}{||\vec{w}||}$. The parameter $\frac{\boldsymbol{b}}{||\vec{w}||}$ determines the offset of the hyperplane from the origin along the normal vector w. [86]

Maximizing margin can be expressed via the following optimization problem [87]:

$$min_{\vec{w},b}\ \frac{1}{2}||\vec{w}||^2\ subject\ to\ \boldsymbol{y_i}(\vec{w}.\overrightarrow{\boldsymbol{x_i}} - b) \geq 1, \forall i, \tag{4.2}$$

where $\overrightarrow{\boldsymbol{x_i}}$ is the ith training example, and y_i is the correct output of the SVM for the ith training example. The value y_i is +1 for the positive examples in a class and –1 for the negative examples.

Not all data sets are linearly separable. There may be no hyperplane that splits the positive examples from the negative examples. Cortes & Vapnik [88] suggested the following modification to the optimization problem:

$$\min_{\vec{w},b,\vec{\xi}} \frac{1}{2}||\vec{w}||^2 + C\sum_{i=1}^{N}\xi_i \ \text{ subject to } y_i(\vec{w}\cdot\vec{x}_i - b) \geq 1 - \xi_i, \forall i, \tag{4.3}$$

where ξ_i are slack variables that permit margin failure and C is a parameter which trades off wide margin with a small number of margin failures.

For non-linear classifiers, the output of a non-linear SVM is explicitly computed from the Lagrange multipliers:

$$u = \sum_{j=1}^{N} y_j \alpha_j K(\vec{x}_j, \vec{x}) - b, \tag{4.4}$$

where K is a kernel function that measures the similarity or distance between the input vector $\vec{x}$ and the stored training vector $\vec{x}_j$. The Lagrange multipliers α_i are computed via a quadratic program. The following dual objective function Ψ is quadratic in α [87]:

$$\begin{aligned} \min_{\vec{\alpha}} \Psi(\vec{\alpha}) = \min_{\vec{\alpha}} \tfrac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} y_i y_j K(\vec{x}_i, \vec{x}_j) \alpha_i \alpha_j - \sum_{i=1}^{N} \alpha_i, \\ 0 \le \alpha_i \le C, \forall i, \\ \sum_{i=1}^{N} y_i \alpha_i = 0. \end{aligned} \tag{4.5}$$

In Equation (4.4), K is a kernel and some common kernels include polynomial (homogeneous), polynomial (inhomogeneous), radial basis function (RBF) and hyperbolic tangent [86].

- Polynomial (homogeneous): $K(\boldsymbol{x_i}, \boldsymbol{x_j}) = (\boldsymbol{x_i} . \boldsymbol{x_j})^d$.
- Polynomial (inhomogeneous): $K(\boldsymbol{x_i}, \boldsymbol{x_j}) = (\boldsymbol{x_i} . \boldsymbol{x_j} + 1)^d$.
- RBF: $K(\boldsymbol{x_i}, \boldsymbol{x_j}) = \exp(-\gamma ||\boldsymbol{x_i} - \boldsymbol{x_j}||^2), \gamma > 0$.
- Hyperbolic tangent: $K(x_i, x_j) = \tanh(\mu x_i^T . x_j + c)$, for some (not every) $\mu > 0$ *and* $c < 0$

For the above kernels, d, μ, c and γ are kernel parameters. The performance of SVM depends on the selection of kernel, the kernel's parameters, and cost parameter C. The RBF kernel is used for testing in this chapter. This kernel nonlinearly maps samples

into a higher dimensional space and can handle the case when the relation between class labels and attributes is nonlinear. When RBF kernel is used, two parameters have to be properly chosen for good classification performance: the gamma (γ) parameter of the RBF kernel and the C parameter.

In this chapter, the SVM software implementation in OpenCV [76] software library is used. The SVM in OpenCV is based on LIBSVM [89]. The C-Support Vector Classification (C-SVC) type and the RBF kernel of LIBSVM are used. According to the recommendation of [90], the feature values are linearly scaled to the range of [0,1]. The parameters C and γ (gamma) of SVM (using RBF kernel) are chosen by using PSO to search for C and gamma (γ) that can provide the best fitness function value of PSO. The fitness function used is the classification accuracy of SVM in the training set, using leave one out (LOO) cross validation.

4.2.1.2 J48 Classifier

J48 is an efficient implementation of the C4.5 [91] tree classifier that produces decision trees. The C4.5 is a classifier using binary trees based on the concept of information entropy computed in training data. In data mining, decision tree is a predictive model which can be used to represent both classifiers and regression models. When a decision tree is used for classification tasks, it is more appropriately referred to as a classification tree. Classification tree is used to classify an instance to a predefined set classes based on their attribute values. It is a flow-chart-like tree structure, where each internal note denotes a test on an attribute, each branch represents an outcome of the test, and leaf nodes represent classes or class distributions [91]. The complex decision in this model is broken up into a group of several simpler decisions to find out

the best solution [92] for the desired classification solution. Classification tree grows in a recursive manner by partitioning the training data.

In order to classify an unknown sample, the attribute values of the sample are tested against the classification tree. The unknown sample is routed down the tree according to the values of the attributes tested in consecutive nodes. When it reaches the leaf, the instance is classified according to the class assigned to the leaf. A path is traced from the root to a leaf node that holds the class prediction for that sample.

4.2.1.3 K Nearest Neighbour Classifier

K Nearest Neighbour (kNN) [93] is one of the simplest instance based learning or lazy learning techniques that assumes all instances correspond to points in the *n*-dimensional feature space. kNN is a supervised learning algorithm where the result of new instance query is classified based on majority of kNN category. The purpose of this algorithm is to classify a new object based on attributes and training samples. The learner only needs to store the examples, while the classifier does its work by observing the most similar examples of the example to be classified. The classifiers are only based on memory and do not use any model to fit. In order to classify an instance of a test data into the corresponding categories, kNN calculates the distance between the test data and each instance of training data set [94]. For example, let an arbitrary instance x be described by the feature vector $\langle a_1(x), a_2(x),\ldots, a_n(x)\rangle$, where $a_r(x)$ is the rth attribute of instance x. The Euclidean distance between two instances x_i and x_j is defined as $d(x_i, x_j)$ where

$$d(x_i, x_j) \equiv \sqrt{\sum_{r=1}^{n} (a_r(x_i) - a_r(x_j))^2} \tag{4.6}$$

The algorithm then finds the k closest training instances to the test instances. The corresponding class with the highest frequency among k classes associated with these k training instances is the class mapped to the test data. The classification uses majority vote among the classification of the k objects. kNN algorithm used neighborhood classification as the prediction value of the new query instance.

4.2.1.4 Artificial Neural Network

Artificial Neural Networks (ANNs) have been widely used in the field of pattern recognition. The advantages of ANNs include their capability of self-learning, and their suitability to solve problems that are too complex for conventional techniques, or hard to find algorithmic solutions [19]. There are two common types of ANN classifiers for mammogram masses: the multilayer perceptron with backpropagation (MLP) and the radial basis function (RBF) network [19]. Christoyianni et al. [44] has reported that the MLP classifier performed better than RBF in mass classification accuracy but RBF networks have the advantage of fast learning rates. The main drawback of the MLP using backpropagation is the long training time

In this chapter, the implementation of MLP in the WEKA machine learning software library [95] is used. One hidden layer is used for the MLP. The number of input units of MLP is equal to the number of features. As the output of MLP is either mass or non-mass, the number of output units is one. For MLP, the validation set size in neural network training is set to 50% of the training set, and other parameter settings of MLP follow the default parameter settings in WEKA library.

4.2.2 Particle Swarm Optimization

PSO is a population based stochastic optimization technique modeled after the social behavior of bird flock [96]. In PSO, the algorithm maintains a population of particles, where each particle represents a potential solution to the optimization problem. Each particle is also assigned a randomized velocity. The particles are then flown through the problem space [96,97]. The aim of PSO is to find the particle position that results in the best evaluation of a given fitness function.

Each particle keeps track of the following information in the problem space: x_i, the current position of the particle; v_i, the current velocity of the particle, and y_i, the personal best position of the particle which is the best position that it has achieved so far. This position yields the best fitness value for that particle. The fitness value of this position, called *pbest*, is also stored.

There are two approaches to PSO, namely local best (*lbest*) and global best (*gbest*). The difference is in the neighborhood topology used to exchange information among the particles. For the *gbest* model, the best particle is determined from the entire swarm. For the *lbest* model, a swarm is divided into overlapping neighborhoods of particles. For each neighborhood, a best particle is determined. The *gbest* PSO is a special case of *lbest* when the neighbourhood is the entire swarm.

Another best value that is tracked by the global version of the PSO is the overall best value, obtained so far by any particle in the population. The location of this overall best value is called y_g. This location is also tracked by PSO. In this thesis, the *gbest* model of PSO is used.

The PSO changes the velocity of each particle at each time step so that it moves toward its personal best and global best locations. The algorithm for implementing the global version of PSO is as follows [97]:

1. Initialize a population of particles with random positions and velocities on a d-dimensional problem space.
2. For each particle, evaluate the desired optimization fitness function of *d* variables.
3. Compare particle's fitness evaluation with particle's personal best value (y_i). If current value is better than y_i, then set y_i value equal to the current value, and the personal best location equal to the current location in the d-dimensional space.
4. Compare fitness evaluation with the population's overall previous best value. If current value is better than global best value (y_g), then reset global best position to the current particle's position and y_g to current particle's value.
5. Change the velocity and position of the particle according to Equations (4.7) and (4.8), respectively.

$$v_i(t+1) = wv_i(t) + c_1 r_1(t)\big(y_i(t) - x_i(t)\big) + c_2 r_2(t)(y_g(t) - x_i(t)) \tag{4.7}$$

$$x_i(t+1) = x_i(t) + v_i(t+1) \tag{4.8}$$

where w is the inertia weight, c_1 and c_2 are the acceleration constants, and $r_1(t)$ and $r_2(t)$ are random numbers generated in the range between 0 and 1. The update strategy of the direction of motion of the individual particle in the search space is guided by the velocity which is dynamically adjusted according to its previous velocity, its personal best position found through its individual experience, and the global best position obtained among all particles

in the swarm, as shown in Figure 4.2. Velocity updates are also clamped to prevent them from exploding, thereby causing premature convergence.

6. Loop to Step 2 until a termination criterion is met. The criterion is usually a sufficiently good fitness or a maximum number of iterations. In this chapter, the maximum number of iterations used for termination is 100.

For a complete run of the PSO algorithm, the number of computations required is the sum of the computations required to calculate the cost of a candidate solution (based on current position of the particles) and the computations required to update each particle's velocity (Equation 4.7) and position (Equation 4.8). Both of these are directly proportional to the number of iterations. The computational complexity of evaluating the cost function depends on the particular cost function under consideration [98].

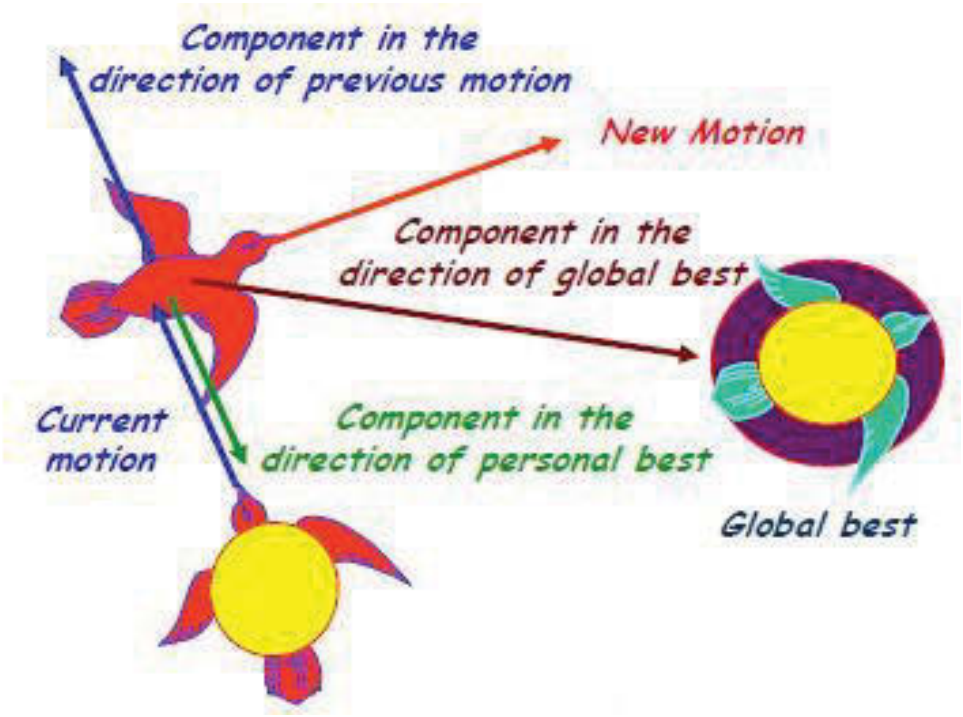


Figure 4.2 PSO: update strategy of direction of motion [99]

4.2.3 Parameters Tuning of SVM By PSO

Before feature selection, the parameters C and gamma (γ) of SVM, using the RBF kernel, are chosen by using PSO to search for values of C and γ that can provide the best fitness function value, using all the available features. The classification accuracy of SVM is used as the fitness function for PSO. In the training set, leave one out (LOO) cross validation is used. The LOO cross validation is especially suitable for small training set as it can maximize the use of training data. The two values $\log_2 C$ and $\log_2 \gamma$ are searched by PSO within the range from -10 to 10. Hence the actual range of C and γ that can be found in the search is from 2^{-10} to 2^{10}.

4.2.4 PSO Based Feature Selection

Feature selection is an important data preprocessing technique which can be used to improve the classification accuracy and minimize the number of features selected. An excessive number of features or high dimensionality of the input feature space might cause data over-fitting. A large number of features can cause the problem of "the curse of dimensionality" and in some cases may decrease the predictive accuracy. Feature selection can be used to reduce the number of features by eliminating irrelevant and redundant features, and thus improve efficiency and/or classification performance [100]. Feature selection has been applied in pattern classification, data analysis, multimedia information retrieval, medical data processing, machine learning, and data mining applications [79].

Existing feature selection approaches can be broadly classified into two categories: filter and wrapper approaches. A filter approach is conducted as a preprocessing procedure and the search process is independent of a learning algorithm. In wrapper approaches, a learning algorithm is part of the evaluation function to

determine the goodness of the selected feature subset. Wrappers can usually achieve better results than filters while filters are more general and computationally less expensive than wrappers [101,102].

A feature selection algorithm searches the attribute space of different feature combinations to reduce the number of features and simultaneously optimize the classification performance. In feature selection, for n features, the size of the search space is 2^n. In most situations, it is impractical and time-consuming to conduct an exhaustive search [101,102]. Therefore, the search strategy is important in feature selection. Different search techniques have been applied to feature selection such as greedy search. The attribute space is searched greedily in one of the two directions, top to bottom or bottom to top. At each stage, a local change is made to the current attribute subset by either adding or deleting a single attribute. For forward selection, the search is in the downward direction. It starts with no attributes and the attributes are added one at a time. For backward elimination, the search direction is upward. It starts with the full set and attributes are deleted one at a time. In forward selection, after tentatively adding each attribute that is not already in the current subset, the resulting set of attributes is evaluated. The effect of adding each attribute in turn is evaluated, the best one is chosen, and the process continues. The search will end if no attribute produces an improvement when added to the current subset. This greedy search procedure guarantees to find a locally, but not necessarily globally, optimal set of attributes. Backward elimination operates in a similar fashion. For best-first search, it does not just terminate when the performance starts to drop but keeps a list of all attribute subsets evaluated so far, sorted in order of the performance measure, so that it can revisit an earlier configuration instead. Given enough time it will explore the entire space [95]. The above mentioned searches might suffer from the problem of becoming stuck in

local optima and/or high computational cost [103,104]. Therefore, an efficient global search technique is needed to develop a good feature selection algorithm.

Evolutionary computation techniques are well-known for their global search ability, and have been applied to the feature selection problems. These include particle swarm optimization (PSO) [105,106], genetic algorithms (GAs) [107] and genetic programming (GP) [108]. Compared with GAs and GP, PSO is easier to implement, has fewer parameters, computationally less expensive, and can converge more quickly [109]. Feature selection problems have two goals, which are maximising the classification performance and minimising the number of features. These two objectives are usually conflicting and there is a trade-off between them. However, most of the existing FS approaches, including PSO based approaches, aim to maximise the classification performance only. In this thesis, PSO based feature selection is used to maximise the mass classification accuracy.

In mammography, many texture-based computer aided diagnostics systems have been proposed [110,111,112] which used a heuristic feature search based on GA approach for feature selection. In this chapter, PSO will be used in feature selection due to its well-known global search ability, less expensive computation and quicker convergence (when compared to GA and GP).

The original version of PSO described above operated in continuous space. The binary version of PSO (BPSO) has been developed for discrete problems [113] which can be used in feature selection. The velocity in BPSO represents the probability of an element in the position taking value 1. Equation (4.7) is used to update the velocity while x_i, y_i and y_g are restricted to 1 or 0. A sigmoid function $s(v_i)$ is used to transform v_i to the range of (0,1). BPSO updates the position of each particle according to the following formulae:

$$x_i = 1 \; if \; rand(\,) < s(v_i), else \; 0 \;\; \& \;\; s(v_i) = \frac{1}{1+e^{-v_i}} \tag{4.9}$$

rand( ) is a random number selected from a uniform distribution in [0,1]. In this chapter, binary PSO (BPSO) is used to search for the feature subset in the training set. When x_i is 1, the feature corresponding to this bit position will be selected. When x_i is 0, the feature will not be selected. SVM classifier is used to evaluate the feature subset using leave-one-out cross validation. The fitness function used in the proposed BPSO based approach is to maximize classification accuracy. The classification accuracy is given by

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{4.10}$$

where TP is the number of true positives, FN is the number of false negatives, TN is the number of true negatives and FP is the number of false positives.

The pseudo-code of feature selection using BPSO is shown below [114].

Begin

divide dataset into a training set and a test set;

randomly initialize the position and velocity of each particle;

While the stopping criterion is not met **do**

evaluate fitness of each particle using the error rate on the training set;

For *i* = 1 to *population-size* **do**

update the *pbest* of particle *i*;

update the *gbest* of particle *i*;

For *i* = 1 to *population-size* **do**

For *d* = 1 to *number-of-available-features* **do**

update the velocity of particle *i*;

update the position of particle *i*;

calculate the classification accuracy of the selected feature subset on the test set; return the position of *gbest* (the selected feature subset)

4.3 Texture Features

In Gray Level Co-occurrence matrix (GLCM), the texture context information is specified by the matrix of relative frequencies $P(i, j, d, \theta)$ with which two neighboring pixels separated by distance d and along direction θ occur on the image; one pixel with gray level i and the other with gray level j [19,45]. After the number of neighboring pixel pairs R used in computing a particular GLCM matrix is obtained, the matrix is normalized by dividing each entry by R, the normalizing constant [45]. For each ROI, eight texture features were derived from each GLCM [45,49,115]. The notation $p(i, j)$ is used to represent the $(i, j)th$ entry in a normalized GLCM matrix and $p(i, j)$ is obtained by dividing each entry of the matrix $P(i, j)$ by R [45]. $\sum_{i,j}$ represents $\sum_{i=0}^{n-1}\sum_{j=0}^{n-1}$ where n is the number of gray levels per pixel.

$$Energy = \sum_{i,j} p(i,j)^2 . \tag{4.11}$$

$$Inertia = \sum_{i,j} (i-j)^2\, p(i,j) . \tag{4.12}$$

$$Entropy = -\sum_{i,j} p(i,j)\log(p(i,j)) . \tag{4.13}$$

$$Homogeneity = \sum_{i,j} \frac{1}{1+(i-j)^2} p(i,j) . \tag{4.14}$$

$$Max.probability = maximum\ of\ \ p(i,j) . \tag{4.15}$$

$$Cluster\ Shade = \sum_{i,j} (i+j-\mu_x-\mu_y)^3\, p(i,j) . \tag{4.16}$$

$$Cluster\ Tendency = \sum_{i,j} (i + j - \mu_x - \mu_y)^2\, p(i,j)\,. \tag{4.17}$$

$$Correlation = \frac{\sum_{i,j}(i - \mu_x)\big(j - \mu_y\big)p(i,j)}{\sigma_x \sigma_y}. \tag{4.18}$$

$$\mu_x = \sum_i i \sum_j p(i,j)\,. \tag{4.19}$$

$$\mu_y = \sum_j j \sum_i p(i,j)\,. \tag{4.20}$$

$$\sigma_x^2 = \sum_i (i - \mu_x)^2 \sum_j p(i,j)\,. \tag{4.21}$$

$$\sigma_y^2 = \sum_j (j - \mu_y)^2 \sum_i p(i,j)\,. \tag{4.22}$$

where μ_x , μ_y, σ_x and σ_y are the means and standard deviations of the marginal distributions associated with $P(i, j) / R$, and R is the normalizing constant [45,49,115].

In finding the GLCM, d is set to 1 in this paper. Four directions are used for θ : 0, 45, 90 and 135 degrees. Hence four features are generated for each feature, one for each matrix. Then the average and range of the four values of each feature are calculated. The range is defined as the difference between the maximum and minimum of the four values. Hence a total of sixteen texture features are found for each ROI. The statistics of the GLCM features are used instead of using the individual feature values of each matrix.

In addition to the 16 GLCM features, seven statistical features are derived from the gray level histogram of each ROI. The seven features are mean, standard deviation, uniformity, entropy, smoothness, skew and kurtosis [8, 32]. In the equations below, g represents the gray level value of a pixel, L is the number of gray levels in a pixel (256 for 8 bits gray level in this paper) and $P(g)$ is the histogram probability of gray level value g.

$$mean = \sum_{g=0}^{L-1} gP(g) \tag{4.23}$$

$$\sigma = \sqrt{\sum_{g=0}^{L-1} (g - mean)^2 P(g)} \tag{4.24}$$

$$skew = \frac{1}{\sigma^3} \sum_{g=0}^{L-1} (g - mean)^3 P(g) \tag{4.25}$$

$$entropy = -\sum_{g=0}^{L-1} P(g) log_2 P(g) \tag{4.26}$$

$$smoothness = 1 - \frac{1}{1 + \sigma^2} \tag{4.27}$$

$$uniformity = \sum_{g=0}^{L-1} P(g)^2 \tag{4.28}$$

$$kurtosis = \frac{1}{4} \sum_{g=0}^{L-1} (g - mean)^4 P(g) - 3 \tag{4.29}$$

A total of 23 features are used based on GLCM and gray level histogram. The BPSO-SVM based feature selection technique is then used to select the most significant features from the 23 features.

4.4 Experimental Results

4.4.1 Experimental Setup

The mini-MIAS Mammographic Database is provided by the Mammographic Image Analysis Society in UK [14]. The mammograms are digitized at 200 micron pixel edge and have a resolution 1024 x 1024. The types of abnormality in the database include calcification, masses, architectural distortion and asymmetry. Mammograms which do not contain any abnormality (classified as normal) are also provided.

One hundred and twenty ROIs were extracted from the images in the mini-MIAS database. The approach of extracting ROIs from the mammogram database is based on [43]. All the crops containing a mass are resized to a fixed size of 128 x 128 pixels. The resizing of variable size ROI to a fixed size region has been used in other research paper on mass classification [43]. 44 of the 120 ROIs contain mass and 76 of them contain normal tissue only. For ROIs which contain mass, the mass can be benign or malignant. Three types of masses were used in this chapter: circumscribed, spiculated and ill-defined masses. Five-fold stratified cross validation is used in testing. The 120 ROIs are divided into five equal sets. Four sets are used as a training set and the remaining set as a test set. Hence there are 96 ROIs in the training set and 24 ROIs in the test set. Feature selection by BPSO-SVM is done using the training set only. Then only the significant features obtained from feature selection are used to train the classifier, using the training set only. The trained classifier is then used to classify the test set using the significant features. The above process is repeated by using another set of data as a test set and the other four sets as a training set. Every ROI is used in the test set once only. The average classification accuracy of the five test sets is calculated.

In BPSO-SVM based feature selection, SVM is used to evaluate the feature subset in the training set. The classification accuracy of the feature subset on the training set is evaluated using SVM and LOO cross validation. Once the significant features have been found by the BPSO-SVM technique, only the significant features are used in the training set to train the classifier. Note that 5-fold cross validation is used to calculate the classification accuracy of the SVM on the test set while LOO cross validation is used to evaluate the feature subset found by BPSO-SVM in the training set. The PSO based parameters tuning for SVM and the BPSO-SVM feature selection method were implemented using C++ language and OpenCV software library [76]. The BPSO based

feature selection method is compared with other wrapper based feature selection methods which are all available in the WEKA machine learning workbench [95]. The wrapper subset evaluation technique used is SVM. The three different search techniques in WEKA library used to find feature subsets include stepwise forward selection, stepwise backward selection and best first search [95].

4.4.2 Experimental Results and Discussion

In Table 4.1, 4.2 and 4.3, the values of specificity, sensitivity and overall accuracy are all measured in the test set, using 5-fold cross validation. The notation "BPSO-SVM" refers to the proposed method in this chapter. Sensitivity, specificity and overall accuracy are defined as follows [95]:

$$Sensitivity = \frac{TP}{TP + FN} \quad (4.30)$$

$$Specificity = \frac{TN}{FP + TN} \quad (4.31)$$

$$Overall\ Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (4.32)$$

where TP is the number of true positives, FN is the number of false negatives, TN is the number of true negatives and FP is the number of false positives.

For Table 4.1 and 4.2, the following test condition is used:

- 120 ROIs are extracted from the images in mini-MIAS database for testing (details are given in section 4.4.1);
- 23 texture features are extracted from the ROI (details of 23 texture features are given in Section 4.3);
- five-fold cross validation is used in testing;

- for Table 4.1, SVM is used as classifier;
- for Table 4.2, feature selection is used for BPSO-SVM and other classifiers do not use feature selection.

Table 4.1: Comparison of feature selection methods using SVM as classifier

Feature Selection Method	Specificity %	Sensitivity %	Accuracy %
BPSO-SVM	97.33	97.78	97.50
All Features	96.05	88.64	93.33
Stepwise forward search	96.10	85.84	92.50
Stepwise backward search	94.76	88.34	92.50
Best first search	96.10	88.06	93.32

Table 4.1 shows that BPSO-SVM feature selection method has better specificity, sensitivity and overall accuracy when compared to stepwise forward search, stepwise backward search and best first search method.

Table 4.2: Comparison of classification methods using BPSO-SVM (with feature selection) and other classifiers without feature selection

Classifier	Specificity %	Sensitivity %	Accuracy %
BPSO-SVM + SVM	97.33	97.78	97.50
SVM (all features)	96.05	88.64	93.33
MLP	94.76	83.12	90.82
J48 (decision tree)	89.58	88.34	89.16
KNN (K=3)	97.42	86.40	93.34

In Table 4.2, except the proposed method, all the other classifiers shown were used to classify the test set without using feature selection. For the MLP, J48 and KNN classifiers, their implementations in the WEKA machine learning software library [95] are used. From Table 4.2, when all the twenty-three features are used, SVM has the best sensitivity when compared to multi-layer perceptron (MLP), J48 decision tree and KNN classifiers. In mass classification, sensitivity is usually regarded as a more important parameter than specificity. Also when all the features are used, SVM has better accuracy when compared to MLP and J48 and SVM has the same overall accuracy as KNN. From Table 4.2, by comparing the result of BPSO-SVM and SVM, it shows that instead of using all available features, a subset of significant features after feature selection can improve the performance of SVM in specificity, sensitivity and accuracy.

Table 4.3: Comparison of proposed BPSO-SVM based classification and other existing mammogram mass classification techniques

Classification method	**Specificity %**	**Sensitivity %**
BPSO-SVM + SVM	97.33	97.78
Angelini et al. [58]	94.00	90.00
Christoyianni et al. [59]	83.05	86.66
Sahiner et al. [40]	69.00	90.00
Petrosian et al. [57]	64.00	76.00
Tourassi et al. [42]	65.00	90.00

Table 4.3 compares the performance of the proposed BPSO-SVM method with other existing mammogram mass classification techniques. It should be noted that some methods in Table 4.3 used different mammographic database in testing. However, the specificity and sensitivity of the proposed method in this chapter show good and encouraging result when compared to other existing methods.

4.5 Conclusion

The objective of this chapter is to demonstrate the good performance of the proposed feature selection and mass classification approach using BPSO and SVM. PSO is used to search for the optimal parameters C and gamma of SVM, using the RBF kernel. Then BPSO-SVM feature selection technique is used to find the significant features in the training set. Finally SVM is used to classify the test set, using the significant features only. The experimental results show that the proposed BPSO-SVM feature selection method can have better result than other widely used feature selection methods when it is applied to mammogram mass classification. By using features from GLCM and gray level histogram, a small number of significant features found by BPSO-SVM can have better performance in classification accuracy than the full set of features in mass classification. Also the proposed mass classification approach has better performance when compared to other existing mass classification techniques. The proposed classification approach using PSO and SVM can achieve 97.78% sensitivity and 97.33% specificity on the test set using 5-fold cross validation.

CHAPTER 5

Image Clustering by Particle Swarm Optimization

5.1 Introduction to Clustering

Clustering refers to the process of grouping samples so that the samples are similar within each group. The groups are called clusters [77]. Clustering algorithms are used in many applications, such as pattern recognition, image analysis, data mining, machine learning and image segmentation.

Clustering methods have been widely used in image segmentation [65], and also for mass detection and/or segmentation. Clustering algorithms can be hierarchical or partitional [116]. In hierarchical clustering, the output is a tree showing a sequence of clustering, with each cluster being a partition of the data set. Partitional clustering algorithms attempt to decompose the data set directly into a set of disjoint clusters [116]. They try to optimize certain criteria such as square-error function. Hierarchical methods can be more accurate, but in general they are too slow for large datasets. For applications involving large datasets, partitional clustering algorithms are used. However there are two main disadvantages of partitional algorithms: the number of regions in the image has to be determined before clustering and partitional clustering algorithms do not use spatial information inherent to the image.

The k-means algorithm is a widely used partitional clustering method [117]. Its advantages are simple implementation and low complexity. However there are two problems of k-means algorithm, the lack of spatial constraints and the assumption of constant intensity in each cluster [20]. To overcome these two problems, an adaptive clustering algorithm for segmentation was introduced by Pappas [118]. Pappas used a generalization of K-means clustering algorithm to separate the pixels into clusters based on their intensity and their relative location. Li et al. [119] used local adaptive thresholding to segment mammographic image into parts belonging to same classes and adaptive clustering to refine the segmentation. Sahiner et al. [120] used k-means clustering algorithm followed by object selection to detect initial mass shape within the region of interest (ROI). The ROI contains the breast masses and the location of ROI is identified by radiologist. After initial mass shape detection, an active contour segmentation method is used to refine the boundaries of the segmented mass.

The Fuzzy C-Means (FCM) algorithm [121] is an extension of the k-Means algorithm which allows each pattern of the image to be associated with every cluster using a fuzzy membership function. Velthuizen [122] used FCM to group pixels with similar grey-level values in the original images. Chen and Lee [123] used it over the set of local features extracted from the application of a multi-resolution wavelet transform and Markov Random Fields (MRF) analysis [124].

According to Fu and Mui [125], threshold methods are considered as partitional clustering methods. Threshold methods have been widely used for mass detection and/or segmentation. A review of various threshold methods used in mass detection and/or segmentation has been discussed in section 3.1.

Recently particle swarm optimization (PSO) [96,126] has been applied to image clustering [127,128]. It has been shown in [127] that PSO-based image clustering can

have better performance than K-means. M. Omran et al. [127] has applied PSO image clustering to classify three types of images which include synthesized, MRI and satellite images. His experimental result showed that PSO clustering can perform better in minimizing intra-distance and maximizing inter-distance than ISO-DATA algorithm. The ISODAT algorithm is an implementation of the K-means approach, which uses Euclidean distance as the similarity measure to cluster pixels into different spectral classes [127].

In PSO-based clustering, the design of good fitness function for PSO is important to ensure the quality of clustering. Various fitness functions have been proposed for PSO based clustering [127,128,129]. In this chapter, two new fitness functions are proposed that can provide good quality of image clustering. The improved fitness functions can provide more compact clusters and larger separation between the cluster centroids when compared to k-means clustering.

In this chapter, the performance PSO based clustering will be compared with k-means clustering using mammographic images from the mini-MIAS database. The theory of the k-means and PSO clustering will be covered in section 5.2 and 5.3 respectively. Two new fitness functions for PSO clustering will be proposed in section 5.3. It can be shown in section 5.4 that the two proposed new fitness functions for PSO clustering can improve the result of mass segmentation. Also the test result in section 5.4 shows that PSO clustering has better mass segmentation performance when compared to k-means clustering.

5.2 K-MEANS Clustering

K-means clustering [130] groups data vectors into a predefined number of clusters, based on the Euclidean distance as similarity measure. Data vectors within a cluster

have small Euclidean distance from one another, and are associated with the centroid vector, which represents the mean of the data vectors that belong to the cluster.

The standard k-means algorithm is summarized as follows [131]:

1. *k* initial cluster centroids (*k*=3 in this example) are randomly generated within the data domain (the three cluster centroids are shown as circles in Figure 5.1).

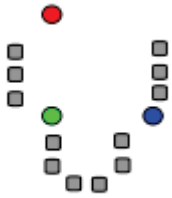


Figure 5.1: K-means algorithm: initialization of randomized centroids [131]

2. For each data vector, assign the vector to the class with the closest cluster center, using the Euclidean distance between the data vector and the centroid. *k* clusters ($k = 3$ in this example) are created by associating every observation with the nearest mean.

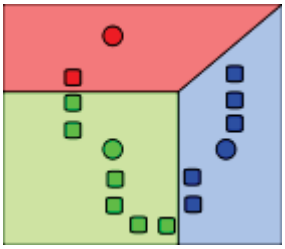


Figure 5.2: K-means algorithm: association with nearest centroid [131]

3. Re-calculate each cluster's centroid vector, which represents the mean of the data vectors that belong to the cluster.

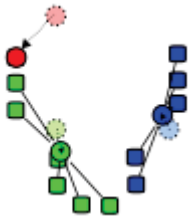


Figure 5.3: K-means algorithm: re-calculation of centroids [131]

4. Repeat steps 2 and 3 until a stopping criterion is satisfied.

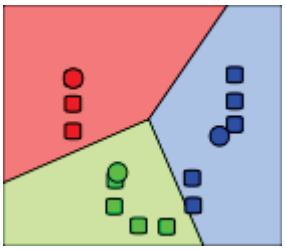


Figure 5.4: K-means algorithm: repeat until a stopping criterion is satisfied [131]

For image clustering, a data vector represents a pixel of the image. The K-means clustering has the following two main advantages [116]. It is easy to implement and the time complexity is only O(n) (where n is the number of data points), which makes it suitable for large data sets. However, its performance is heavily dependent on the initial conditions. This often causes K-means to converge to suboptimal solutions.

5.3 PSO-Based Image Clustering

The general theory of PSO has been covered in section 4.2.2 and will not be repeated here. This section will describe how PSO can be used in image clustering.

To facilitate the explanation, the following notations will be used:

- N_p denotes the number of image pixels to be clustered

- N_c denotes the number of clusters to be formed
- z_p denotes the p-th pixel
- m_j denotes the mean of cluster j
- C_j denotes the subset of pixel vectors that form cluster j
- $|C_j|$ denotes the number of pixels in cluster j
- C_{ij} denotes cluster j of particle i.

In PSO-based image clustering, a single particle represents the N_c cluster means. Each particle x_i is constructed as $x_i = (m_{i1}, \ldots, m_{ij}, \ldots, m_{iNc})$ where m_{ij} refers to the j-th cluster centroid vector of the i-th particle. The quality of each particle is measured by the fitness function.

In this chapter, the PSO based clustering algorithm proposed in [127] is used. It can be summarized below:

1. Initialize each particle to contain N_c randomly selected cluster means.
2. For $t = 1$ to t_{max} (maximum number of iterations)

   (a) For each particle i

   - For each pixel z_p

     Calculate $d(z_p, m_{ij})$ for all clusters C_{ij}

     Assign z_p to C_{ij} where

$$d(z_p, m_{ij}) = min_{\forall c=1,\ldots,N_c}\{d(z_p, m_{ic})\}$$

$d(z_p, m_{ij})$ represents the Euclidean distance between the p-th pixel z_p and the centroid of j-th cluster of particle i.

- Calculate the fitness function $f(x_i(t),Z)$ where Z is a matrix representing the assignment of pixels to clusters of particle i.

(b) Update the personal best and the global best positions.

(c) Update the cluster centroids using Equations (5.1) and (5.2).

$$v_i(t+1) = wv_i(t) + c_1r_1(t)\big(y_i(t) - x_i(t)\big) + c_2r_2(t)(y_g(t) - x_i(t)) \quad (5.1)$$

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (5.2)$$

Equations (5.1) and (5.2) are the same as Equations (4.7) and (4.8) and are repeated here for convenience of reference. Details of the equations are found in Chapter 4.

The fitness function proposed in [127,128] uses the following three evaluation criteria: quantization error, intra-cluster distance and inter-cluster separation.

The quantization error J_e is defined below:

$$J_e = \frac{\sum_{j=1}^{N_c}[\sum_{\forall z_p \in C_j} d(Z_p, m_j)]/|C_j|}{N_c} \quad (5.3)$$

where $d(z_p,m_j)$ represents the Euclidean distance between the p-th pixel z_p and the centroid of j-th cluster m_j.

The intra-cluster distance is measured by $\bar{d}_{\max}$ which is defined in [127,128] as

$$\bar{d}_{\max}(Z, x_i) = max_{j=1,\ldots,N_c}\{\sum_{\forall z_p \in C_{ij}} d(z_p, m_{ij})/|C_{ij}|\} \quad (5.4)$$

Z is a matrix representing the assignment of pixels to clusters of particle i. A smaller value of $\bar{d}_{\max}$ means the clusters are more compact.

Another measure of quality is the inter-cluster separation. It is measured by the minimum Euclidean distance between any pair of clusters and is defined below:

$$d_{min}(x_i) = min_{\forall j1, j2, j1 \neq j2}\{d(m_{ij_1}, m_{ij_2})\} \tag{5.5}$$

The above three criteria have been used by [128] to form the fitness function as shown by Equation (5.6).

$$f_1(x_i, \boldsymbol{Z}) = w_1 \bar{d}_{max}(\boldsymbol{Z}, x_i) + w_2(z_{max} - d_{min}(x_i)) + w_3 J_e \tag{5.6}$$

where w_1, w_2 and w_3 are user defined constants and determine the relative weights of intra-cluster distance ($\bar{d}_{max}$), inter-cluster separation (d_{min}) and quantization error (J_e) in the fitness function. z_{max} is the maximum pixel value in the image set, which is 255 for 8-bit grayscale image used in this chapter. One objective of the fitness function in Equation (5.6) is to minimize the intra-cluster distance ($\bar{d}_{max}$) and the quantization error (J_e). The two objectives are to make the clusters compact and maximize the inter-cluster separation (d_{min}) so that the clusters are well separated.

However, a recent research paper [132] has pointed out a problem with the use of quantization error J_e as defined by Equation (5.3) in data clustering. In Equation (5.3), for every cluster, it first calculates the average distance of the pixels of a cluster to its cluster centroid. Then it takes the average distances of all clusters and calculate another average, which is denoted by J_e. In [132], Esmin *et al.* pointed out that a cluster with just one data vector will influence the final result as much as another cluster with many data vectors. For example, suppose that one of the particle's clusters has one data vector that is very close to the centroid, and another cluster has many data vectors that are not so close to the centroid. This is not a very good solution, but giving the same weight to the cluster with one data vector as the cluster with many data vectors can make it seem

to be a good solution [132]. To solve this problem, [132] proposed another equation which gives a higher weighting to the cluster with many data vectors in the calculation of the fitness function. In this thesis, the modified quantization error proposed by [132] is called the weighted quantization error J_{e2}.

$$J_{e2} = \{\sum_{j=1}^{N_c}[(\sum_{\forall z_p \in C_{ij}} d(z_p, m_{ij})/|C_{ij}|).(|C_{ij}|/N_o)]\} \quad (5.7)$$

where N_o is the number of data vectors to be clustered.

The paper in [132] used the weighted quantization error J_{e2} alone to cluster three benchmark data sets from the UCI repository of Machine Learning Databases. It reported that the use of J_{e2} in clustering improved the performance when compared to data clustering using J_e alone. In this chapter, J_{e2} is first used alone in clustering of natural images and its performance is compared to k-means using the three evaluation criteria: intra-cluster distance ($\bar{d}_{max}$), quantization error (J_e) and inter-cluster separation (d_{min}) . When J_{e2} is used alone in the fitness function, it gives better result in quantization error when compared to k-means. However, k-means method provides better result in inter-cluster separation. To solve this problem and enhance the performance in inter-cluster separation of PSO-based clustering, this thesis proposes that J_{e2} should not be used alone in the fitness function for PSO-based image clustering. This thesis proposes a new fitness function similar to Equation (5.6) used by Omran *et al.* in [128] but replaces J_e by J_{e2}, as given by the equation below.

$$f_2(x_i, \boldsymbol{Z}) = w_1 \bar{d}_{max}(\boldsymbol{Z}, x_i) + w_2\big(z_{max} - d_{min}(x_i)\big) + w_3 J_{e2} \quad (5.8)$$

where J_{e2} is given by Equation (5.7).

The proposed new fitness function in Equation (5.8) will improve the fitness function used by Esmin *et al.* in [132] (which uses weighted quantization error only) as shown by the experimental result. Equation (5.8) will also improve the fitness function used by Omran *et al.* in [128] (which uses Equation (5.6)) as [132] has shown that J_{e2} solves the problem of J_e in clustering. This paper shows that J_{e2} should be used together with $\bar{d}_{max}$ and d_{min} to obtain both compact clusters and large inter-cluster separation.

The second new fitness function proposed in this chapter uses the mean square-error (*MSE*) defined by

$$MSE = \frac{1}{n}\sum_{j=1}^{K}\sum_{z_p \in C_j}(z_p - m_j)^2, \tag{5.9}$$

where n is the total number of pixels in the image, z_p is the *p-th* pixel, K is the number of clusters, m_j is the centroid of the *j-th* cluster C_j. *MSE* is a measure of the compactness of the clusters [133] and represents the mean squared distance of the pixels from its associated cluster centroid.

It should be noted that *MSE* defined by Equation (5.9) does not have the problem of J_e as described above. A cluster with one data vector will not influence the result as much as another cluster with many data vectors. For example, if one of the particle's clusters has one pixel that is very close to the centroid, and another cluster has many pixels that are not so close to the centroid, the *MSE* in Equation (5.9) will correctly give a large error value.

Using *MSE* alone in PSO clustering will generally give good performance in $\bar{d}_{max}$ and J_e but slightly worse performance in d_{min} when compared to k-means, as shown by experimental results in Section 5.4. To improve the performance in inter-cluster separation, *MSE* is used together with $\bar{d}_{max}$ and d_{min} in the fitness function f_3 below.

$$f_3(x_i, Z) = w_1 \bar{d}_{max}(Z, x_i) + w_2\left(z_{max} - d_{min}(x_i)\right) + w_3 . MSE \quad (5.10)$$

5.4 Results and Discussion

5.4.1 Image Clustering Using Standard Test Images

The two new fitness functions f_2 and f_3 in Equations (5.8) and (5.10) are used in the PSO-based image clustering algorithm. The PSO based image clustering algorithm has been applied to three grayscale images: Lena, Pepper and Airplane. The performance is measured by the following three criteria: intra-cluster distance ($\bar{d}_{max}$), quantization error (J_e) and inter-cluster distance (d_{min}). These three criteria have been used in [127,128]. The performance of PSO-based clustering is then compared to the k-means algorithm.

For all the experiments, the choice of the parameters for PSO-based clustering is based on the recommendation of previous research publications. The following parameters are used for PSO-based clustering:

- Number of particles = 20
- Number of iterations for termination = 150
- Number of clusters = 5
- Acceleration constants c_1 and c_2 = 2

The number of particles used is problem-dependent. The common choice of number of particles varies from 20 to 50 [97, 134]. In all experiments of this chapter, 20 particles are used for PSO clustering as smaller number of particles can reduce

computation time and 20 particles can provide good clustering performance in this paper when compared with k-means. The number of clusters is chosen to be 5 for both k-means and PSO clustering to allow a fair comparison of their performance.

For the inertia weight w, the initial weight value is 0.9 and w decreases linearly with the number of iterations. The final value is 0.4 when the termination condition (150 iterations) is reached. By linearly decreasing the inertia weight from a relatively large value to a small value through the course of the PSO run, the PSO tends to have more global search ability at the beginning of run while having more local search ability near the end of the run [135,136]. The acceleration constants c_1 and c_2 are both set to 2. The settings of acceleration constants and the inertia weight are based on the recommendation by [136].

For k-means algorithm, the number of iterations is 3000. This is chosen to equal the number of fitness function evaluations in PSO based clustering (20 particles and 150 iterations will give 3000 fitness function evaluations).

For fitness functions f_2 and f_3 in PSO clustering, each fitness function consists of three sub-objectives. The weighting of each sub-objective (w_1, w_2 and w_3) that provides best performance is determined empirically in this chapter. The research work of Omran et al. [128] is used to guide the choice of the weights of the three sub-objectives. The new fitness function proposed in Equation 5.8 is similar to the fitness function in Equation 5.6 used by Omran et al. but replaces the quantization error in Equation 5.6 by weighted quantization error in Equation 5.8. Omran et al. had tried different combinations of the three weights values for the sub-objectives empirically and his result is that weights of $w_1 = 0.1$, $w_2 = 0.1$ and $w_3 = 0.8$ result in the smallest quantization error, shortest intra-distances and largest inter-distances for MRI image [128]. These

sub-objectives weights values are also used in the testing of all images in this chapter. To eliminate the tuning of these weight values, multi-objective optimization approach can be used [128,137].

All data in Table 5.1 to 5.5 below are the averages of 25 program runs. The numbers after the ± symbols in the tables represent the standard deviation.

By comparing Table 5.1 and 5.2, PSO clustering using $\bar{d}_{max}$, d_{min}, and J_{e2} together can give more compact clusters (smaller $\bar{d}_{max}$) and larger inter-cluster separation for all the three images while the performance with respect to J_e is comparable.

By comparing Tables 5.2 and 5.5, PSO clustering using $\bar{d}_{max}$, d_{min} and J_{e2} together can give better performance than k-means for image Pepper for all the three evaluation criteria. For images Lena and Airplane, PSO clustering has better performance than k-means with respect to $\bar{d}_{max}$ and d_{min} while PSO has slightly higher quantization error J_e than k-means. However, their performance with respect to J_e is still comparable. The big improvement of PSO clustering in $\bar{d}_{max}$ and d_{min} shows that by using $\bar{d}_{max}$, d_{min} and J_{e2} together, more compact clusters and larger inter-cluster separation can be achieved.

Table 5.1: PSO-based clustering using weighted quantization error only

Images	***Intra-cluster distance*** $\bar{d}_{max}$	***Quantization error*** J_e	***Inter-cluster distance*** d_{min}
Pepper	10.9006 ±1.0487	9.6045±0.1473	30.2707±0.9067
Lena	10.2972±0.1532	8.4037±0.0047	27.0092±0.0307
Airplane	15.9973±0.0183	9.1213±0.0110	13.2569±0.3834

Table 5.2: PSO-based clustering using intra-cluster distance, inter-cluster distance and weighted quantization error with w_1=0.1, w_2=0.1,w_3=0.8

Images	*Intra-cluster distance* $\bar{d}_{max}$	*Quantization error* J_e	*Inter-cluster distance* d_{min}
Pepper	10.9006 ±1.0487	9.6045±0.1473	30.2707±0.9067
Lena	10.2972±0.1532	8.4037±0.0047	27.0092±0.0307
Airplane	15.9973±0.0183	9.1213±0.0110	13.2569±0.3834

TABLE 5.3: PSO-BASED CLUSTERING USING MSE ONLY

Images	*Intra-cluster distance* $\bar{d}_{max}$	*Quantization error* J_e	*Inter-cluster distance* d_{min}
Pepper	10.2567 ±0.2182	9.7898±0.0521	31.7022±0.5417
Lena	9.6387±0.1311	8.4393±0.0036	29.1630±0.1793
Airplane	15.8544±0.0388	9.7643±0.0114	18.7229±0.0941

TABLE 5.4: PSO-BASED CLUSTERING USING INTRA-CLUSTER DISTANCE, INTER-CLUSTER DISTANCE & MSE

Images	*Intra-cluster distance* $\bar{d}_{max}$	*Quantization error* J_e	*Inter-cluster distance* d_{min}
Pepper	10.2110 ±0.0009	9.7692±0.0007	33.8947±0.0226
Lena	9.3937±0.0555	8.4455±0.0009	30.8199±0.0210
Airplane	10.8355±0.0893	9.8273±0.0373	39.5477±0.4057

TABLE 5.5: K-MEANS CLUSTERING

Images	*Intra-cluster distance* $\bar{d}_{max}$	*Quantization error* J_e	*Inter-cluster distance* d_{min}
Pepper	13.0798 ±1.9485	10.0068±0.2643	32.4692±0.1867
Lena	9.7053±0.4377	8.4432±0.0042	29.3819±0.1248
Airplane	15.6241±0.2511	9.8564±0.0986	20.0633±1.5214

By comparing Table 5.3 and 5.5, for the three images under test, PSO clustering using *MSE* alone generally can provide more compact clusters as the values for $\bar{d}_{max}$ and J_e are generally smaller. However, PSO clustering using *MSE* alone have smaller inter-cluster separation when compared to k-means. To improve the performance in inter-cluster separation, *MSE* will be used together with $\bar{d}_{max}$ and d_{min} in the fitness function f_3 (Equation (5.10)).

By comparing Tables 5.3 and 5.4, PSO clustering using $\bar{d}_{max}$, d_{min} and *MSE* together will give better performance in intra-cluster distance and inter-cluster separation when compared to PSO clustering using *MSE* alone. The performance in J_e is comparable for both methods.

By comparing Tables 5.4 and 5.5, PSO clustering using $\bar{d}_{max}$, d_{min} and *MSE* together have better performance than k-means for all the three images in all evaluation criteria except for Lena image, where the quantization error for the PSO clustering is only slightly worse than J_e for k-means . Hence, it can be concluded that PSO clustering using $\bar{d}_{max}$, d_{min} and *MSE* together perform better than k-means by giving more compact clusters and larger inter-cluster separation.

By comparing Tables 5.2 and 5.4, PSO clustering using f_3 is more robust than PSO clustering using f_2. PSO clustering using f_3 can produce nearly same results over repeated runs when compared to clustering using f_2. In Table 5.4 (using f_3), all data have very small standard deviation when compared to results in Table 5.2 (using f_2).

The conclusion from the experiments is that both PSO clustering methods, using either fitness functions f_2 or f_3, can give more compact clusters and larger inter-cluster separation when compared to k-means.

5.4.2 Mass Segmentation Using PSO Based Image Clustering

There are different types of image segmentation techniques which include amplitude thresholding, component labelling, boundary-based approaches, temmplate approaches, texture segmentation, region based approaches and clustering [138]. It should be noted that the objective of mass detection is to find the location of suspicious regions. For some masses, the mass boundary found by mass detection may not be very accurate. Once the suspicious region for the mass is found, more accurate mass segmentation can be applied to a window surrounding the suspicious region. Clustering is one common approach of image segmentation. In this section, PSO based image clustering is used for mass segmentation and its performance is compared to k-means, a widely used clustering method.

The parameters of PSO based clustering used in this section are as follows:

- Number of particles = 20
- Number of iterations for termination = 100
- Number of clusters = 4
- Acceleration constants c_1 and c_2 = 2

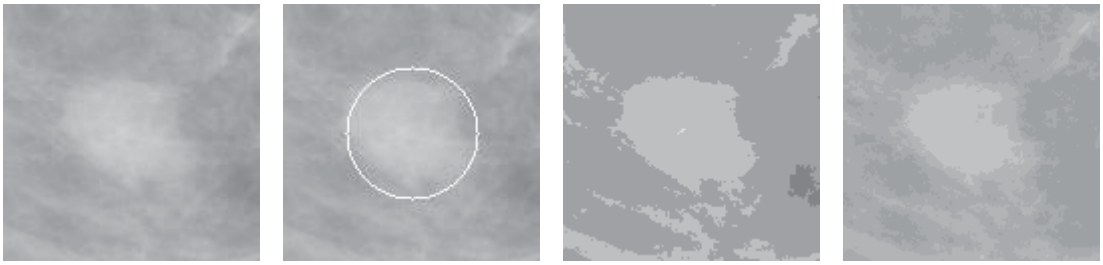
For the inertia weight w, the initial weight value is 0.9 and w decreases linearly with the number of iterations. The final value is 0.4 when the termination condition (100 iterations) is reached.

For k-means clustering, the same number of clusters (four) is used in order to compare the performance of PSO based clustering with k-means. All software algorithms are implemented in C++ using the OpenCV software library.

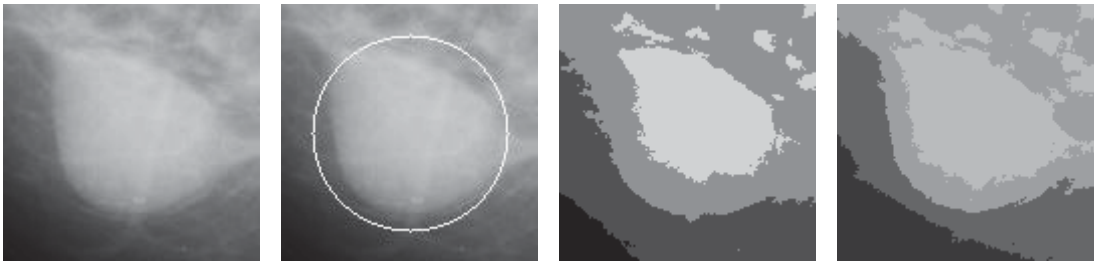
To perform mass segmentation, a square window is used to enclose the mass. The center of the square window is the centroid of the mass given by the ground truth data of the mini-MIAS mammogram database. The window size is 128 x 128 pixels. All the mass images are extracted from the mini-MIAS mammogram database [14].

The objective of this section is to show that PSO based clustering can have better mass segmentation performance when compared to k-means clustering. The ground truth data of the mini-MIAS database is used to extract the windows of mass images for image segmentation. It is also possible to use the output of mass detection method in Chapter 3 to generate the window of image for segmentation. The centroid of the suspicious region can be used as the center of the window. The size of the suspicious region can be used to determine the square window size. The square window size should be big enough to enclose the whole suspicious region, together with some normal tissue surrounding the suspicious region.

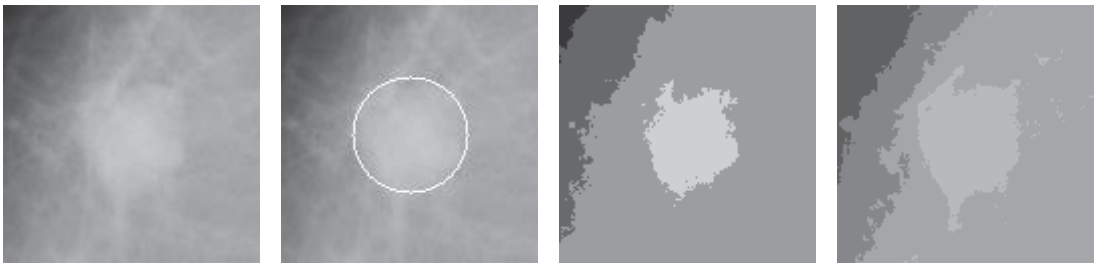
In Figure 5.5, the third column from the left represents the output image after PSO based image clustering while the fourth column represents the output image after k-means clustering. The cluster with the brightest intensity represents the mass region. From Figure 5.5 above, it can be observed that for k-means clustering, the mass merges with neighbouring normal tissue for mdb021 (Figure 5.1 (b)), mdb023 (Figure 5.1 (c)) and mdb204 (Figure 5.1 (e)). For PSO clustering, there is a small amount of merging for mdb010 (Figure 5.1 (a)). K-means has merging problem of mass and normal tissue in three images while PSO clustering has merging problem in one image (Figure 5.1 (a)). Moreover, the degree of merging problem for K-means is more serious in the three images, especially for mdb204.



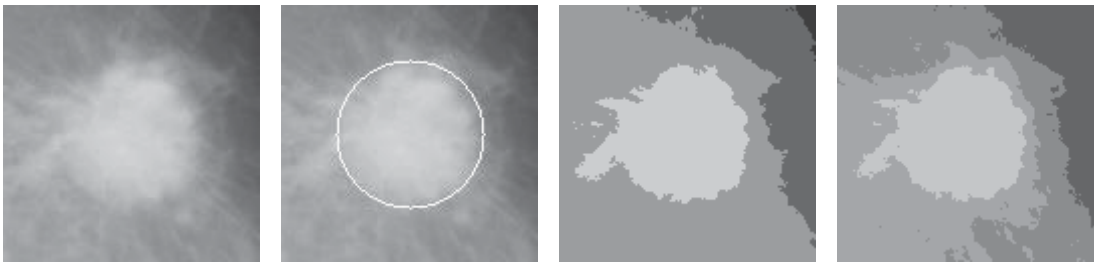
(a) Mass image of mdb010, ground truth circle, outputs of PSO and K-means clustering



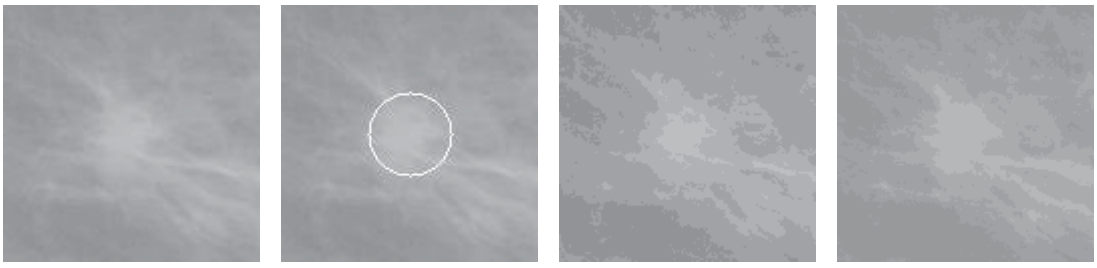
(b) Mass image of mdb021, ground truth circle, outputs of PSO and K-means clustering



(c) Mass image of mdb023, ground truth circle, outputs of PSO and K-means clustering



(d) Mass image of mdb202, ground truth circle, outputs of PSO and K-means clustering



(e) Mass image of mdb204, ground truth circle, outputs of PSO and K-means clustering

Figure 5.5: Mass segmentation using PSO clustering and K-means

Details of locations of merging of mass and neighbouring normal tisuue are given below for each image:

- mdb010 (Figure 5.1 (a)): Minor merging of mass and normal tissue near bottom right hand corner of the mass for PSO clustering.
- mdb021 (Figure 5.1 (b)): merging of mass and normal tissue at the top left corner and bottom right corner of the mass for k-means.
- mdb023 (Figure 5.1 (c)): merging of mass and normal tissue at the bottom of the mass for k-means. It can be seen as a long tail at the bottom.
- mdb204 (Figure 5.1 (e)): merging of mass and normal tissue at the bottom right hand corner of the mass for k-means. The merging problem is serious for this image.

For the five images tested in Figure 5.1, for k-means clustering, three images have the problem of merging of mass and neighbouring normal tissue. For PSO clustering, only one image has a small amount of merging. From Figure 5.1, it can be seen that PSO based image clustering has better performance in mass segmentation when compared to k-means as far as the merging of mass and neighbouring normal tissue is concerned. For mini-MIAS database, detailed mass boundary is not given in the ground truth file. Instead, the radius of the circle which encloses the mass is given. The center of the circle is the centroid of the mass. It should be noted that there is normal breast tissue inside the circle because the mass is usually not circular, especially for spiculated and ill-defined masses.

While PSO clustering has better performance in mass segmentation, k-means is more efficient and the time complexity of k-means is only O(n) (where n is the number of data points) [117]. For a complete run of the PSO algorithm, the number of computations required is the sum of the computations required to calculate the cost of a candidate solution (based on current position of the particles) and the computations

required to update each particle's velocity and position. Both of these are directly proportional to the number of iterations [98]. However, in mammographic mass classification, it is more important to have good mass segmentation performance in order to assist the radiologist to determine if a suspicious region is mass or normal tissue.

5.5 Conclusion

This chapter has proposed two fitness functions that can improve PSO-based image clustering. This thesis shows that when the modified quantization error proposed in [132], called weighted quantization error (J_{e2}) in this thesis, is used alone in image clustering, its performance in inter-cluster distance is worse than k-means though it can give smaller quantization error. To solve this problem, the first proposed fitness function uses J_{e2} together with intra-cluster distance ($\bar{d}_{max}$) and inter-cluster separation (d_{min}) to improve clustering quality. In the second proposed fitness function, the mean square-error is used together with $\bar{d}_{max}$ and d_{min}. Experimental results show that PSO-based image clustering, using the two proposed new fitness functions, can have more compact clusters and larger inter-cluster separation when compared to k-means clustering.

The PSO based image clustering can be used to perform mass segmentation. Using the proposed new fitness function which include the weighted quantization error, intra-cluster distance and inter-cluster separation, it has been shown that PSO clustering has better mass segmentation performance for the mini-MIAS mammogram database when compared to k-means clustering.

CHAPTER 6

Conclusion and Future Work

Masses and micro-calcifications are two major types of mammographic abnormalities [18]. The objective of mass detection is to find locations of suspicious regions with high sensitivity and small number of false positives per image. Mass detection is generally more difficult than the detection of micro-calcifications because mass can have various size and shape. Also the features of mass can be obscured or similar to normal breast parenchyma [19]. In this thesis, a mass detection method is proposed to detect three different types of masses: circumscribed, spiculated and ill-defined masses. Then particle swarm optimization (PSO) based mass classification technique is proposed to classify the suspicious regions into masses or normal tissue. The objective is to reduce the number of false positives in mass detection. Finally a PSO based image clustering technique is proposed for mass segmentation which can have better segmentation performance when compared to k-means clustering.

6.1 Mass Detection

The proposed mass detection method can detect three difference types of masses: circumscribed, spiculated and ill-defined. Masses of different size and shape can be

detected. The method is based on contrast enhancement by contrast limited adaptive histogram equalization, thresholding, false positives reduction by area and shape measure. It has comparable performance with other existing mass detection techniques. The proposed method has a sensitivity of 88.2% with the number of false positives per image (FPI) equal to 5.66. The FPI can be reduced by using the PSO based mass classification technique proposed in this thesis.

6.2 Mass Classification by PSO and SVM

This thesis proposes an effective method to classify the suspicious regions (ROI) of mammograms into mass and normal breast tissue regions by using PSO based feature selection and support vector machine (SVM). PSO is used to search for the optimal parameters C and gamma of SVM, using the RBF kernel. Twenty-three texture features were derived from the gray level co-occurrence matrix (GLCM) and gray level histogram of each ROI. Feature selection based on binary PSO (BPSO) and SVM is used to find the significant texture features in the training set. Then SVM is used to classify the test set, using the significant features only. Experimental results show that the proposed BPSO-SVM feature selection method can have better result when compared to other widely used feature selection techniques in mass classification. Also a small number of significant features found by the BPSO-SVM feature selection can have better mass classification accuracy than the full set of features. The proposed BPSO-SVM mass classification approach has better or comparable sensitivity & specificity when compared to other existing mass classification techniques. The proposed PSO based mass classification method can achieve 97.78% sensitivity and 97.33% specificity on the test set selected from the mini-MIAS database using 5-fold cross validation.

6.3 PSO Based Image Clustering and Mass Segmentation

This thesis proposes two new fitness functions which can improve the performance of PSO based image clustering. The first fitness function is based on weighted quantization error, intra-cluster distance and inter-cluster separation. The second fitness function is based on mean square error, intra-cluster distance and inter-cluster separation. Two experiments are performed to demonstrate that PSO based image clustering performs better than k-means, a widely used clustering algorithm. For the first experiment, three standard test images are used which include Lena, Pepper and Airplane. The performance of clustering is evaluated by the following three criteria: quantization error, intra-cluster distance and inter-cluster distance. It can be shown experimentally that PSO based image clustering, using the proposed fitness functions, can perform better than k-means by generating more compact clusters and larger inter-cluster separation.

In the second experiment, mammographic images from the mini-MIAS mammogram database are used. The test images are generated by manually cropping a square window which completely surrounds the mass in the mammogram image. The centroid and the size of the mass are provided by the ground truth file in the mini-MIAS database. The new fitness function used for PSO image clustering is based on weighted quantization error, intra-cluster distance and inter-cluster distance. Experimental results show that PSO based image clustering can give better performance in mass segmentation when compared to k-means clustering. For k-means clustering, there is more merging of the mass and its neighboring normal breast tissue.

6.4 Future Work

For mass detection, pectoral muscle removal can be implemented in the future. S. M. Kwok et al. [139] had developed a method for automatic segmentation of the pectoral muscle on the mediolateral oblique views of mammograms. The pectoral boundary is approximated with a straight line. The line is then iteratively refined to a curve which can accurately delineate an enclosed pectoral region. The removal of pectoral muscle will reduce the number of false positives per image.

For particle swarm optimization based image clustering and k-means clustering, the number of clusters is chosen to be four for mammographic mass segmentation in this thesis. S. Ray and R. H. Turi [140] had proposed an automatic method to determine the number of clusters in image clustering. They used a simple validity measure based on the inter-cluster and intra-cluster distance to determine the number of clusters. The procedure involves producing all the segmented images from two clusters up to k_{max} clusters, where k_{max} is an upper limit on the number of clusters. The validity measure is then calculated to determine which is the best clustering by finding the minimum value for the measure. The validity measure has been tested for synthetic images and some standard natural images. This approach can be tested for mammographic images in the future.

For mammographic mass segmentation, in this thesis, it has been shown that PSO based image clustering can have better performance than k-means clustering. The Fuzzy C-Means (FCM) algorithm [121] is another popular clustering method which has been used in image segmentation. FCM is an extension of the k-means algorithm which allows each pattern of the image to be associated with every cluster using a fuzzy

membership function. FCM will be applied in mass segmentation and its performance will be compared with PSO based image clustering.

References

1. Cancer Council Australia.

   URL http://www.cancer.org.au/about-cancer/types-of-cancer/breast-cancer.html Web page was updated on 16 Oct 2015.

2. American Cancer Society, 2007. Breast cancer: facts and figures. 2007-08. ACS.

3. K. Bovis, S. Singh, J. Fieldsend, and C. Pinder. Identification of masses in digital mammograms with MLP and RBF nets. In Proc. of the IEEE-INNS-ENNS International Joint Conference in Neural Networks, pages 342-347, 2000.

4. R.E. Bird, T. W. Wallace, and B. C. Yankaskas. Analysis of cancers missed at screening mammography. Radiology, 184(3):613-617, 1992.

5. R.L. Birdwell, D. M. Ikeda, K. D. O'Shaughnessy, and E. A. Sickles. Mammographic characteristics of 115 missed cancers later detected when screening mammography and the potential utility of computer-aided detection. Radiology, 219(1):192-202, 2001.

6. M.L. Giger, N. Karssemeijer, and S.G. Armato. Computer-aided diagnosis in medical imaging. IEEE Transactions on Med. Imag, 20:1205-1208, 2001.

7. M. L. Giger. Computer-aided diagnosis of breast lesions in medical images. Comput. Science Engin, 2:39-45, 2000.

8. C. J. Vyborny, M. L. Giger, and R. M. Nishikawa. Computer aided detection and diagnosis of breast cancer. Radiol. Clin. North Am., 38:725-740, 2000.

9. J. Tang, R. M. Rangayyan, J. Xu, and Y. Yang. Computer-aided detection and diagnosis of breast cancer with mammography: recent advances. IEEE Trans. on information technology in biomedicine, 13(2):236-251, 2009.

10. NCI Cancer Fact Sheets (2007). URL http://www.cancer.gov/cancertopics/faqctsheet/Detection/screening-mammograms.

11. Jelena Bozek, Mario Mustra, Kresimir Delac, and Mislav Grgic. A Survey of Image Processing Algorithms in Digital Mammography. In M. Grgic et al., editors, Rec. Advan. in Mult. Sig. Process. and Commun., SCI 231, Springer-Verlag Berlin Heidelberg, pages 631–657, 2009.

12. R. Warren and S. Duffy. Comparison of single and double reading of mammograms, and change in effectiveness with experience. Br. J. Radiol., 68(813):958–962, 1995.

13. R. Brem, J. Baum, M. Lechner, S. Kaplan, S. Souders, L. Naul, and J. Hoffmeister. Improvement in sensitivity of screening mammography with computer-aided detection: A multiinstitutional trial. Amer. J. Roentgenol., 181(3):687–693, 2003.

14. The mini-MIAS Database of Mammograms. URL http://peipa.essex.ac.uk

15. American College of Radiology. ACR BI-RADS—Mammography, Ultrasound & Magnetic Resonance Imaging, 4th ed. Reston, VA. Amer. Coll. Radiol., 2003.

16. M. P. Sampat, M. K. Markey, and A. C. Bovik. Computer-aided detection and diagnosis in mammography. In Handbook of Image and Video Processing, A.C. Bovik, Ed., 2nd ed., New York: Academic, pages 1195–1217, 2005.

17. S. Timp and N. Karssemeijer. A new 2D segmentation method based on dynamic programming applied to computer aided detection in mammography. Med. Phys., 31(5):958–971, 2004.

18. A. Oliver, J. Freixenet, J. Marti, E. Perez, J. Pont, E. R. E. Denton, and R. Zwiggelaar. A review of automatic mass detection and segmentation in mammographic images. Medical Image Analysis, 14:87-110, 2010.

19. H. D. Cheng, X. J. Shi, R. Min, L. M. Hu, X. P. Cai, and H. N. Du. Approaches for automated detection and classification of masses in mammograms. Pattern Recog., 39: 646-668, 2006.

20. D. B. Kopans: Breast Imaging. *Lippincott Williams, New York*, 1998.

21. R. L. Birdwell, D. M. Ikeda. K. D. O'Shaughnessy, and E. A. Sickles. Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection.

Radiology 219 (1):192–202, 2001.

22. T. W. Freer and M. J. Ulissey. Screening mammography with computer-aided detection: prospective study of 12860 patients in a community breast center. Radiology, 220:781–786, 2001.

23. R. M. Nishikawa and M. Kallergi. Computer-aided detection, in its present form, is not an effective aid for screening mammography. Med. Phys., 33(4):811–814, 2006.

24. W. P. Kegelmeyer, Jr., J. M. Pruneda, and P. D. Bourland, et al. Computer-aided mammographic screening for spiculated lesions. Radiology, 191:331–337, 1994.

25. N. Karssemeijer and G. M. te Brake. Detection of stellate distortions in mammograms. IEEE Trans. Med. Imag., 15, 1996.

26. S. L. Liu, C. F. Babbs, and E. J. Delp. Multiresolution detection of spiculated lesions in digital mammograms. IEEE Trans. Image Process., 10:874–884, 2001.

27. H. D. Li, M. Kallergi, L. P. Clarke, V. K. Jain, and R. A. Clark. Markov random field for tumor detection in digital mammography. IEEE Trans. Med. Imag., 14:565–576, 1995.

28. T. Matsubara, H. Fujita, T. Endo, et al. Development of mass detection algorithm based on adaptive thresholding technique in digital mammograms. K. Doi, M. L. Giger et al. eds., 391–396 Elsevier, Amsterdam, The Netherlands, 1996.

29. H. Li, Y. Wang, K. J. Liu, et al. Computerized radiographic mass detection - part I: Lesion site selection by morphological enhancement and contextual segmentation. IEEE Trans. Med. Imag., 20:289-301, 2001.

30. R. O. Duda, P. E. Hart, and D. G. Stork. Pattern Classification, 2nd ed. Wiley-Interscience, New York, 2000.

31. D. Brzakovic, X. M. Luo, and P. Brzakovic. An approach to automated detection of tumors in mammograms. IEEE Trans. Med. Imag., 9:233–241, 1990.

32. N. Petrick, H. P. Chan, B. Sahiner, et al. An adaptive density-weighted contrast enhancement filter for mammographic breast mass detection. IEEE Trans. Medical Imag., 15:59–67, 1996.

33. N. Petrick, H. P. Chan, B. Sahiner, et al. Combined adaptive enhancement and regiongrowing segmentation of breast masses on digitized mammograms. Medical Physics, 26:1642–1654, 1999.

34. W. E. Polakowski, D. A. Cournoyer, S. K. Rogers, et al. Computer-aided breast cancer detection and diagnosis of masses using difference of Gaussians and derivative-based feature saliency. IEEE Trans. Med. Imag., 16:811–819, 1997.

35. R. Bellotti. A completely automated CAD system for mass detection in a large mammographic database. Medical Physics, 33, 2006.

36. H. Kobatake, M. Murakami, H. Takeo, et al., ‘‘Computerized detection of malignant Tumors on digital mammograms,’’ IEEE Trans. Med. Imag.,18:369-378, 1999.

37. W. Qian, L. Li, L. Clarke, et al. Comparison of adaptive and non adaptive CAD methods for mass detection. Academic Radiol., 6: 471–480, 1999.

38. S. M. Lai, X. Li, and W. F. Bischof. On techniques for detecting circumscribed masses in mammograms. IEEE Trans. Med. Imag., 8:377–386, 1989.

39. B. R. Groshong and W. P. Kegelmeyer. Evaluation of a Hough transform method for circumscribed lesion detection. K. Doi, M. L. Giger et al. eds. 361-366 Elsevier, Amsterdam, The Netherlands, 1996.

40. R. C. Gonzalez and R. E. Woods. Digital Image Processing, 2nd ed. Prentice Hall, Upper Saddle River, NJ, 2001.

41. W. Zhang, K. Doi, M. L. Giger, et al. An improved shift invariant artificial neural network for computerized detection of clustered microcalcifications in digital mammograms. Med. Physics, 23:595–601, 1996.

42. G. M. te Brake, N. Karssemeijer, and J. H. Hendriks. An automatic method to discriminate malignant masses from normal tissue in digital mammograms. Physics Med. Biol., 45:2843–2857, 2000.

43. E. Angelini, R. Campanini, E. Iampieri, N. Lanconelli, and M. Masotti. Testing the performances of Different Image Representations for Mass Classification in Digital mammograms. Int'l Journal of Modern Phys. C, 17(1): 113-131, 2006.

44. I. Christoyianni, E. Dermatas, and G. Kokkinakis. Neural classification of abnormal tissue in digital mammography using statistical features of the texture. In Proc. of the 6th IEEE Int'l Conf. on Electronics, Circuits & Systems, Vol. 1, pages 117-120, 1999

45. R. M. Haralick, K. Shanmugam, and I. Dinstein. Texture features for image classification. IEEE Trans. Syst. Man Cybernet, SMC-3(6): 610-621, 1973.

46. D. Wei, H. P. Chan, M. A. Helvie et al. Classification of mass and normal breast tissue on digital mammograms: multiresolution texture analysis. Med. Physics, 22:1501–1513, 1995.

47. D. Wei, H. P. Chan, N. Petrick, et al. False positive reduction technique for detection of masses on digital mammograms: global and local multiresolution texture analysis. Med. Physics, 24:903–914, 1997.

48. B. Sahiner, H.-P. Chan, N. Petrick, et al. Classification of mass and normal breast tissue: a convolution neural network classifier with spatial domain and texture images. IEEE Trans. Med. Imag., 15:598–610, 1996.

49. A. Petrosian, H. P. Chan, M. A. Helvie, M. M. Goodsitt, and D. D. Adler. Computer-aided diagnosis in mammography: classification of mass and normal

tissue by texture analysis. Physics in Medicine and Biology, 39(12):2273-2288, 1994.

50. M. A. Kupinski and M. L. Giger. Investigation of regularized neural networks for the computerized detection of mass lesions in digital mammograms. In Proceedings of the 19th Annual International Conference of the IEEE, 1997.

51. G. D. Tourassi, R. Vargas-Voracek, D. M. Catarious, Jr., and C. E. Floyd, Jr. Computer-assisted detection of mammographic masses: a template matching scheme based on mutual information. Med. Physics, 30:2123–2130, 2003.

52. G. B. Junior, S. V. da Rocha, A. Silva and A. Paiva. A false positive reduction in mass detection in mass detection approach using spatial diversity analysis. In the Fifth International conference on eHealth, Telemedicine, and Social Medicine, eTELEMED, 2013

53. M. Hussain, S. Khan, G. Muhammad, I. Ahmad, and G. Bebis. Effective extraction of Gabor features for positive reduction and mass classification in mammography. Appl. Math. Inf. Sci., 6(1):29-33, 2012.

54. Victor G. Martinez, Daniel M. Gamo, Juan Rios, and Amparo Vilarrasa. Iterative method for automatic detection of masses in digital mammograms for computer-aided diagnosis. In Proceedings of SPIE, volume 3661, pages 1086-1093, 1999.

55. A. R. Abdel-Dayem, M. R. El-Sakka. Fuzzy entropy based detection of suspicious masses in digital mammogram images. In IEEE Conference on Engineering in Medicine and Biology Society, pages 4017–4022, 2005.

56. T. Matsubara, H. Fujita, T. Hara, S. Kasai, O. Otsuka, Y. Hatanaka, and T. Endo. Development of a new algorithm for detection of mammographic masses. In International Workshop on Digital Mammography, pages 139-142, 1998.

57. T. Matsubara, H. Fujita, S. Kasai, M. Goto, Y. Tani, T. Hara, T. Endo. 1997. Development of new schemes for detection and analysis of mammographic masses. In International Conference on Information Systems, pages 63-66, 1997.

58. S. Özekes, O. Osman, and A. Y. Çamurcu. 2005. Mammographic mass detection using a mass template. Korean J. Radiol. 6 (4):221–228, 2005.

59. M. Sameti, and R. K. Ward. A fuzzy segmentation algorithm for mammogram partitioning. In International Workshop on Digital Mammography, pages 471-474, 1996.

60. M. Sameti, R. K. Ward, J. Morgan-Parkes, B. Palcic. A method for detection of malignant masses in digitized mammograms using a fuzzy segmentation algorithm. In IEEE Conference on Engineering in Medicine and Biology Society, pages 513–516, 1997.

61. J. E. Ball, T. W. Butler, L. M. Bruce. Towards automated segmentation and classification of masses in digital mammograms. In IEEE Conference on Engineering in Medicine and Biology Society, pages 1814-1817, 2004.

62 H. Kobatake, M. Murakami. Adaptive filter to detect rounded convex regions: Iris filter. In IAPR International Conference on Pattern Recognition, volume 2, pages 340-345, 1996.

63. A. Rocha, T. Fu, and Y. Zhuangzhi. A logic filter for tumor detection on mammograms. J. Comput. Sci. Technol., 15(6):629–632, 2000.

64. M. D. Heath, K. W. Bowyer. Mass detection by relative image intensity. In International Workshop on Digital Mammography, pages 219–225, 2000.

65. R. Gupta, P. E. Undrill. 1995. The use of texture analysis to identify suspicious masses in mammography. Phys. Med. Biol., 40 (5):835–855, 1995.

66. P. Undrill, R. Gupta, S. Henry, and M. Downing. Texture analysis and boundary refinement to outline mammography masses. In IEE Colloquium Digest Mammography, pages 511-516, 1996.

67. A. Laine, W. Huda, D. Chen, J. Harris. Segmentation of masses using continuous scale representations. In International Workshop on Digital Mammography, pages 447–450, 1996.

68. A. Laine, W. Huda, B. G. Steinbach, and J. C. Honeyman. Mammographic image processing using wavelet processing techniques. Epidemiol. Rev., 5 (5):518–523, 1995.

69. W. E. Polakowski, D. A. Cournoyer, S. K. Rogers, M. P. DeSimio, D. W. Ruck, J. W. Hoffmeister, R. A. Raines. Computer-aided breast cancer detection and diagnosis of masses using difference of Gaussian and derivative-based feature saliency. IEEE Trans. Med. Imag., 16 (6):811–819, 1997.

70. G. Kom, A. Tiedeu, M. Kom. Automated detection of masses in mammograms by local adaptive thresholding. Comput. Biol. Med., 37 (1):37–48, 2007.

71. T. S Subashini, V. Ramalingam, and S. Palanivel. Pectoral muscle removal and detection of masses in digital mammogram using CCL. International Journal of Computer Applications (0975-8887), 1(6), 2010.

72. Wikipedia: Adaptive histogram equalization URL https://en.wikipedia.org/wiki/Adaptive_histogram_equalization

73. OpenCV 3.0.0-dev documentation: Histogram Calculation URL http://docs.opencv.org/3.0-beta/modules/cudaimgproc/doc/histogram.html

74. Shefali Gupta and Yadwinder Kaur. Review of Different Local and Global Contrast Enhancement Techniques for a Digital Image. International Journal of Computer Applications (0975 – 8887), 100(18), August 2014.

75. K. Rajesh, S. Harish, and Suman. Comparative Study of CLAHE, DSIHE & DHE Schemes. International journal of research in management, science & technology, 1(1).

76. G. Bradski and A. Kaehler. Learning OpenCV, first edition. O'Reilly, 2008.

77. E. Gose, R. Johnsonbaugh, and S. Jost. Pattern Recognition and Image Analysis. Prentice Hall PTR, 1996.

78. H. Zhang, S. Foo, S. M. Krishnan, and C. H. Thng. Automated breast masses segmentation in digitized mammograms. In IEEE International Workshop on BioMedical Circuits & Systems, pages 1-4, 2004.

79. Chung-Jui Tu, Li-Yeh Chuang, Jun-Yang Chang, and Cheng-Hong Yang. Feature Selection using PSO-SVM. IAENG International Journal of Computer Science, 33:1, IJCS_33_1_18, 2007.

80. B. Xue, M. Zhang, and W. N. Browne. Single Feature Ranking and Binary Particle Swarm Optimisation Based Feature Subset Ranking for Feature Selection. In Proceedings of the Thirty-Fifth Australasian Computer Science Conference (ACSC 2012), Melbourne, Australia, 2012.

81. P. M. Murphy and D. W. Aha. UCI Repository of Machine Learning Databases. Technical report, Department of Information and Computer Science, University of California, Irvine, Calif., 1994. URL http://www.ics.uci.edu/~mlearn/MLRepository.html.

82. Mu-Yen Chen and Bo-Tsuen Chen. Applying particles swarm optimization for support vector machines on predicting company financial crisis. In 2010 International Conference on Business and Economics Research, volume 1, IACSIT Press, Kuala Lumpur, Malaysia, 2011.

83. Enrique Alba, Jos´e Garcia-Nieto, Laetitia Jourdan, El-Ghazali Talbi. Gene Selection in Cancer Classification using PSO/SVM and GA/SVM Hybrid Algorithms. In Congress on Evolutionary Computation, CEC 2007, Singapore, IEEE, Sept. 2007.

84. Imad Zyout, Ikhlas Abdel-Qader, and Christina Jacobs. Embedded Feature Selection using PSO-kNN: Shape-Based Diagnosis of Microcalcification Clusters. Mammography Journal of Ubiquitous Systems & Pervasive Networks, 3(1):7-11, 2011.

85. Imad Zyout and Ikhlas Abdel-Qader. Classification of Microcalcification Clusters via PSO-KNN Heuristic Parameter Selection and GLCM Features.

International Journal of Computer Applications (0975 – 8887), 31(2), October 2011.

86. Wikipedia: Support Vector Machine URL https://en.wikipedia.org/wiki/Support_vector_machine#/media/File:Svm_max_sep_hyperplane_with_margin.png

87. John C. Platt. Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines. Microsoft Research, Technical Report MSR-TR-98-14, April 21, 1998. URL www.research.microsoft.com/pubs/69644/tr-98-14.doc

88. C. Cortes and V. Vapnik. Support Vector Networks. Machine Learning, 20:273-297, 1995.

89. C. C. Chang and C. J. Lin. LIBSVM : a library for support vector machines. ACM Trans. on Intelligent Systems and Technology, 2(3), Article No. 27, 2011.

90. C. W. Hsu, C. C. Chang, and C. J. Lin. A practical guide to support vector classification. URL http://www.csie.ntu.edu.tw/~cjlin, last updated on April 15 2010.

91. J.R. Quinlan. C4.5: Programs for Machine Learning. San Mateo: Morgan Kaufmann, 1993.

92. S.R. Safavian and D. Landgrebe, A survey of decision tree classifier methodology. IEEE Transactions on Systems, Man and Cybernetics, 21(3):660-674, May-Jun 1991.

93. D.W. Aha, D. Kibler, and M.K. Albert. Instance-Based Learning Algorithms. *Machine Learning,* 6: 37-66, 1991.

94. T.M. Mitchell. Machine Learning, McGraw-Hill, 1997.

95. I. H. Witten, E. Frank. Data mining: practical machine learning tools and techniques, second edition. Morgan Kaufman, 2005.

96. J. Kennedy and R. Eberhart. Particle swarm optimization. In Proc. of the IEEE International Joint Conf. on Neural Networks, Australia, 4:1942-1948, 1995.

97. R. Eberhart and Y. Shi. Particle swarm optimization: developments, applications and resources. In Proc. of the Congress on Evolutionary Computation, pages 81-86, 2001.

98. M. S. Sohail, M. O. B. Saeed, S. Z. Rizvi, M. Shoaib, and A. U. H. Sheikh. Low-Complexity Particle Swarm Optimization for Time-Critical Applications. arXiv:1401.0546v1 [cs.NE], 2014.

99. Mohamed A. El-Sharkawi. Particle Swarm Optimization, 2008 URL https://class.ee.washington.edu/555/el-sharkawi/lectures/pso.pdf

100. M. Dash and H. Liu. Feature selection for classification. Intelligent Data Analysis, 1(4):131–156, 1997.

101. R. Kohavi and G. H. John. Wrappers for feature subset selection. Artificial Intelligence, 97:273–324, 1997.

102. B. Xue, M. Zhang, and Will N. Browne. Multi-Objective Particle Swarm Optimisation (PSO) for Feature Selection. In GECCO'12, Philadelphia, Pennsylvania, USA, July 7–11, 2012.

103. T. Marill and D. Green. On the effectiveness of receptors in recognition systems. IEEE Transactions on Information Theory, 9(1):11–17, 1963.

104. A. Whitney. A direct method of nonparametric measurement selection. IEEE Transactions on Computers, C-20(9):1100–1103, 1971.

105. A. Unler and A. Murat. A discrete particle swarm optimization method for feature selection in binary classification problems. European Journal of Operational Research, 206(3):528–539, 2010.

106. C. S. Yang, L. Y. Chuang, C. H. Ke, and C. H. Yang. Boolean binary particle swarm optimization for feature selection. In IEEE Congress on Evolutionary Computation (CEC'08), pages 2093–2098, 2008.

107. H. Yuan, S. S. Tseng, and W. Gangshan. A two-phase feature selection method using both filter and wrapper. In IEEE International Conference on Systems, Man, and Cybernetics (SMC'99), volume 2, pages 132–136, 1999.

108 K. Neshatian and M. Zhang. Dimensionality reduction in face detection: A enetic programming approach. In 24th International Conference Image and Vision Computing New Zealand (IVCNZ'09), pages 391–396, 2009.

109. J. Kennedy and W. Spears. Matching algorithms to problems: an experimental test of the particle swarm and some genetic algorithms on the multimodal problem generator. In IEEE Congress on Evolutionary Computation (CEC'98), pages 78–83, 1998.

110. S. Singh, V. Kumar, H. K Verma, and D. Singh. SVM Based System for classification of Microcalcifications in Digital Mammograms. In proceeding of the 28th IEEE EMBS Annual International Conference, New York City, USA, 2006.

111. H. S Zadeh, P. S Nezhad, and F. R Rad. Shape based and texture-based feature extraction for classification of microcalcifications in mammograms. In Proceedings of SPIE Medical Imaging, 4322, 3010-310, 2001.

112. H. S Zadeh, F. R Rad, and S. P Nejad. Comparison of multiwavelet, wavelet, Haralick, and shape features for microcalcification classification in mammograms. Pattern Recognition, 37, 1973-1986, 2004.

113. J. Kennedy and R. Eberhart. A discrete binary version of the particle swarm algorithm. In IEEE Int'l Conf. on Syst., Man, and Cybernetics, 5:4104-4108, 1997.

114. B. Xue, M. Zhang, and W. N. Browne. New fitness functions in binary particle swarm optimization for feature selection. In WCCI 2012 IEEE World Congress on Computational Intelligence, Brisbane, Australia, June 2012.

115. M. M. Eisa, A. A. Ewees, M. M. Refaat and A. F. Elgamal. Effective medical image retrieval technique based on texture features. International Journal of Intelligent Computing and Information Science, 13(2): 19-33, 2013.

116. A. K. Jain, M. N Murty, and P. J. Flynn. Data clustering: a review. ACM Computer Surv., 31(3):264-323, 1999.

117. J. B. MacQueen. Some methods of classification and analysis of multivariate observations. In Berkeley Symposium on Mathematical Statistics and Probability, volume 1, pages 281–297, 1967.

118. T. N. Pappas. An Adaptive Clustering Algorithm for Image Segmentation. IEEE Transactions on Signal Processing, 40(4):901–914, 1992.

119. L. H. Li, W. Qian, L. P. Clarke, R. A. Clark, and J. Thomas. Improving Mass Detection by Adaptive and Multi-Scale Processing in Digitized Mammograms. In Proceedings of SPIE - The International Society for Optical Engineering, 3661(1):490–498, 1999.

120. B. Sahiner, L. M. Hadjiiski, H. P. Chan, C. Paramagul, A. Nees, M. Helvie, J. Shi. Concordance of Computer-Extracted Image Features with BI-RADS Descriptors for Mammographic Mass Margin. In M. L. Giger, N. Karssemeijer, editors, Proc. of SPIE Medical Imaging 2008: Computer-Aided Diagnosis, volume 6915, 2008

121. J. C. Bezdek. Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York, 1981.

122. R. P. Velthuizen. Computer diagnosis of mammographic masses. In: Workshop on Applied Imagery Pattern Recognition, pages 166–172, 2000.

123. C. H. Chen, G. G. Lee. On digital mammogram segmentation and microcalcification detection using multiresolution wavelet analysis. Graph.Models Image Process., 59 (5):349–364, 1997.

124. C. M. Bishop. Pattern Recognition and Machine Learning. Springer, New York, 2006.

125. K. S. Fu and J. K. Mui. A survey on image segmentation. Pattern Recogn., 13:3–16, 1981.

126. R. Eberhart and J. Kennedy. A new optimizer using particle swarm theory. In 6th International Symposium on Micro Machine and Human Science, 1995.

127. M. Omran, A. Salman, and A. Engelbrecht. Image classification using particle swarm optimization. In Proceedings of the 4th Asia-Pacific Conference on Simulated Evolution and Learning, Singapore, 2002.

128. M. Omran, A. Engelbrecht, and A. Salman. Particle swarm optimization method for image clustering. International journal of Pattern Recognition and Artificial Intelligence, 19(3):297-322, 2005.

129. D. Van der Merwe and A. Engelbrecht. Data clustering and particle swarm optimization. In Proceedings of IEEE Congress on Evolutionary Computation (CEC 2003), Caribella, vol. 4, pages 215-220, 2003.

130. P. Tan, M. Steinbach, and V. Kumar. Introduction to data mining. Pearson Education, 2006.

131. Wikipedia: k-means algorithm. URL http://en.wikipedia.org/wiki/K-means_algorithm

132. A. A. A. Esmin, D. L. Pereira, and F. P. A. de Araújo. Study of different approach to clustering data by using particle swarm optimization algorithm. In Proceedings of the IEEE Congress on Evolutionary Computation, CEC 2008, Hong Kong, China, 2008.

133. S. Ouadfel, M. Batouche, and A. taleb-Ahmed. A modified particle swarm optimization algorithm for automatic image clustering. In Proceedings of the Int'l Symposium on Modelling and Implementation of Complex Systems, MISC 2010, pages 49-57, May 2010.

134. R. Poli, J. Kennedy, and T. Blackwell. Particle swarm optimization: an overview. Swarm Intelligence, 1(1):33-57, 2007.

135. Y. Shi and R. Eberhart. A modified particle swarm optimizer. In Proceedings of IEEE International Conference on Evolutionary Computation, World Congress on Computational Intelligence, Anchorage, Alaska, 1998.

136. Y. Shi and R. Eberhart. Empirical study of particle swarm optimization. In Proceedings of the 1999 Congress on Evolutionary Computation, Piscataway, NJ, pages 1945-1950, 1999.

137. C. A. Coello-Coello and M. S. Lechuga. MOPSO: a proposal for multiple objective particle swarm optimization. In Proceedings of the Congress on Evolutionary Computation, CEC 2002, volume 2, pages 1051-1056, 2002.

138. A. K. Jain. Fundamentals of Digital Image Processing. Prentice Hall International Editions, 1989.

139. S. M. Kwok, R. Chandrasekhar, Y. Attikiouzel, and M. T. Rickard. Automatic pectoral muscle segmentation on mediolateral oblique view mammograms. IEEE Transactions on Medical Imaging, 23(9):1129-1140, Sept. 2004.

140. S. Ray, and R. H. Turi. Determination of number of clusters in K-means clustering and application in colour image segmentation. (invited paper) In N. R. Pal, A. K. De and J. Das, editors, Proceedings of the 4th International Conference on Advances in Pattern Recognition and Digital Techniques (ICAPRDT'99), Calcutta, India, 27-29, Narosa Publishing House, New Delhi, India, ISBN: 81-7319-347-9, pages 137-143, December 1999. URL http://www.csse.monash.edu.au/~roset/papers/cal99.pdf