

M.Tech (Computer Science) Dissertation Series

# **Analysis of Mammogram for Detection of Micro-calcification**

A dissertation submitted in partial fulfillment of the  
requirements for the M.Tech (Computer Science)  
degree of the Indian Statistical Institute

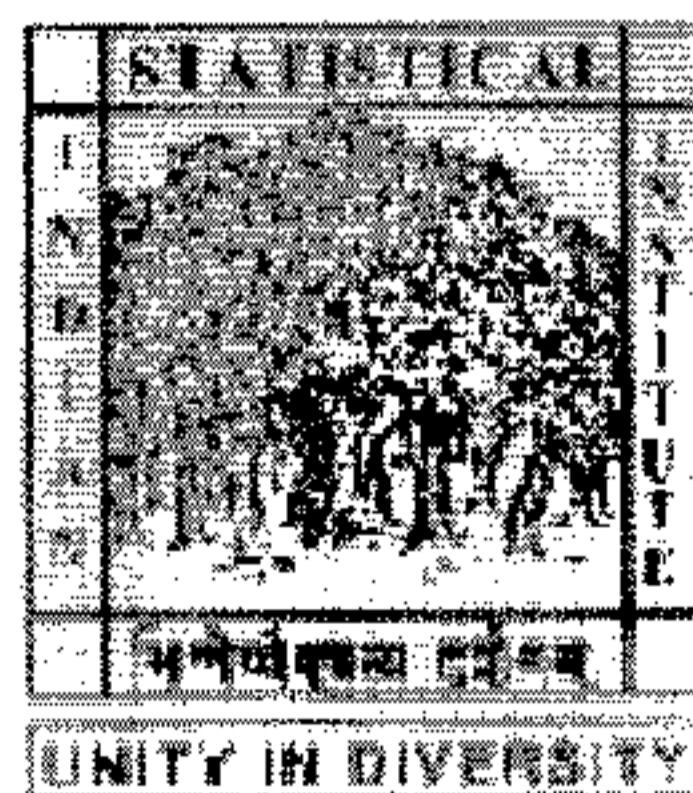
*By*

**Rudra Narayan Hota**

under the supervision of

**Prof. N. R. Pal**

*Electronics and Communication Sciences Unit.*



INDIAN STATISTICAL INSTITUTE

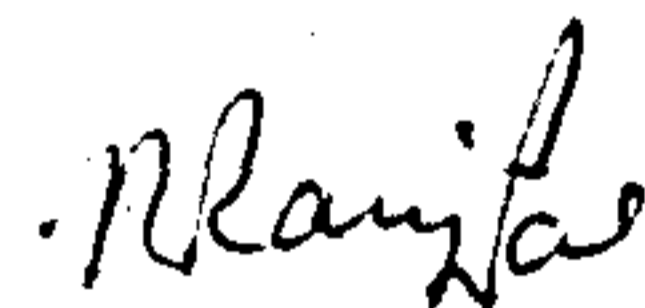
203, Barrackpore Trunk Road

Kolkata – 700 108.

July, 2005.

## **CERTIFICATE OF APPROVAL**

This is to certify that the dissertation titled "*Analysis of Mammogram for Detection of Micro-calcification*" submitted by *Rudra Narayan Hota* towards the partial fulfillment of the requirements for the award of M.Tech in Computer Science degree at the *Indian Statistical Institute, Kolkata* is a bonafide work under my supervision and it has not been submitted anywhere else for the award of any other degree.



**Prof. Nikhil Ranjan Pal**

**ECSU, ISI Kolkata**

## Acknowledgement

I would like to express my sincere gratitude to my dissertation supervisor Prof. **Nikhil Ranjan Pal**, ECSU (Electronics and Communication Sciences Unit), Indian Statistical Institute, Kolkata, for giving me a valuable opportunity to do this dissertation work under his supervision. I am fortunate to have had the opportunity to share some of his great knowledge.

I would also extend my sincere gratitude to all my teachers for their direct and indirect cooperation and support. I would also like to extend my special thanks to all Research Scholars, ECSU, Indian Statistical Institute, Kolkata, for their valuable help and all Staff of ECSU for their continuous support.

My final thanks go to my friends who helped me a lot through their valuable suggestions in completing this report.

Rudra Narayan Hota

# Contents

<b>1 Mammogram Analysis</b>	<b>3</b>
1.1 Introduction	4
1.1.1 What makes the Calcification Detection difficult?	4
1.2 Mammography CAD system	6
1.2.1 Enhancement of Microcalcifications	6
1.2.2 Segmentation of Microcalcifications	7
1.2.3 Micro-calcification detection based on feature extraction	8
1.2.4 Classification of Calcification	8
1.3 Evaluation of Calcification Detection Algorithms	8
1.4 Review of Some Existing Methods	9
<b>2 On-line Feature Selection with Regularization using Neural Networks</b>	<b>11</b>
2.1 Feature selection Method Using OFS	12
2.2 Regularization of OFS system	14
2.3 Data Set Description	15
2.4 Results and Discussion	17
<b>3 Extracting Fuzzy Rules With higher Specificity for Classification</b>	<b>21</b>
3.1 Fuzzy rule based classification method	22
3.1.1 Generation of prototypes	22
3.1.2 Fuzzy rule generation from the prototypes	22
3.1.3 Tuning of the fuzzy rules	23
3.2 Regularization of Fuzzy Rules	25
3.3 Results and Observations	26
<b>4 Detection of Micro-calcifications</b>	<b>30</b>
4.1 Detection of Micro-Calcification [2]	30
4.1.1 Mammogram Fuzzification	31
4.1.2 Image Enhancement	33

4.1.3	Irrelevant Breast Structure Removal . . . . .	33
4.2	Proposed Modifications . . . . .	33
4.2.1	Use of S-function instead of $\Pi$ -function for image fuzzi- fication . . . . .	33
4.2.2	Setting of Threshold on Local variance for Non-uniformity of calcification . . . . .	34
4.3	Calcification Detection method . . . . .	35
4.4	Results and Discussion . . . . .	37
<b>5</b>	<b>Classification of Microcalcifications.</b>	<b>38</b>
5.1	Classification of Micro-Calcification using OFS . . . . .	38
5.2	Fuzzy rules for Micro-Calcification Classification . . . . .	39
<b>6</b>	<b>Conclusion</b>	<b>43</b>

# Chapter 1

## Mammogram Analysis

With the recent westernization of our life style, the incidence rate of breast cancer is fast increasing in most of the part of the world, specifically in advanced countries. Primary prevention seems impossible since the causes of this disease still remain unknown. Early detection is the key to improving breast cancer prognosis. The presence of microcalcification clusters (MCCs) is an important sign for the detection of early breast carcinoma. Hence mammography is one of the reliable methods for early detection of breast carcinomas. In this thesis we develop methodologies for detection and classification of mammograms.

In our mammogram analysis we shall use a neural network based method for feature selection and then use both neural networks and fuzzy rule based system for classification of micro-calcification. We have made some modification of an existing feature selection algorithm and also proposed a scheme for fuzzy rule extraction that avoids an important problem of existing fuzzy rule tuning methods.

Here are the contributions made in this thesis:

1. We modified a neural network based feature selection method using regularization to improve its performance.
2. We proposed a method to extract fuzzy rules with higher *specificity* for classification.
3. We modified an existing method to detect calcification for better performance of diagnostic system.
4. We applied the feature selection method developed in (1) to select useful features for classification of calcification.
5. We applied the fuzzy rule based classification technique developed in (2) and also neural networks to classify the detected micro-calcification.

Since these the methods developed in (1) and (2) above are of general nature, nothing specific to mammogram analysis. we discuss them in Chapter 2 and Chapter 3. In Chapter 4 we discuss an existing method to detect micro-calcification and propose some modifications to get better results in detection of micro-calcification. In Chapter 5 we discuss about the classification of the detected calcification by applying the techniques proposed in Chapters (2) and (3).

In the remaining part of this chapter we present basics of the mammogram analysis, followed by different types of calcification. We also discuss how mammogram results are validate. The chapter is concluded with a very brief survey of literature.

## 1.1 Introduction

In the past several years there has been tremendous interest in image processing and analysis techniques in mammography. This is because the tactile examination is insufficient for the correct diagnose of the breast cancer. The high correlation between the appearance of the microcalcification clusters and the diseases shows that the CAD (computer aided diagnosis) systems for automated detection of MCCs will be very useful and helpful for breast cancer control.

The mammogram analysis for identification breast cancer needs two steps:  
(I) Detection of calcification with low false positive rate.  
(II) Classification of the calcification into Normal, Benign or Malignant class.

Basically breast cancer abnormalities are characterized into three different types:

- [a] Circumscribed Lesions.
- [b] Microcalcifications
- [c] Speculated Lesions.

These three types of abnormalities are shown in Fig 1.1.

### 1.1.1 What makes the Calcification Detection difficult?

Although computer-aided mammography has been studied over two decades, automated interpretation of microcalcifications remains very difficult. It is mainly due to their fuzzy nature, low contrast and low distinguishability from their surroundings:

- (a) Usually it involves *small object* area of interest, and hence leads to potential misidentification.
- (b) Due to *Different sizes, various shapes, and variable distributions* of mi-



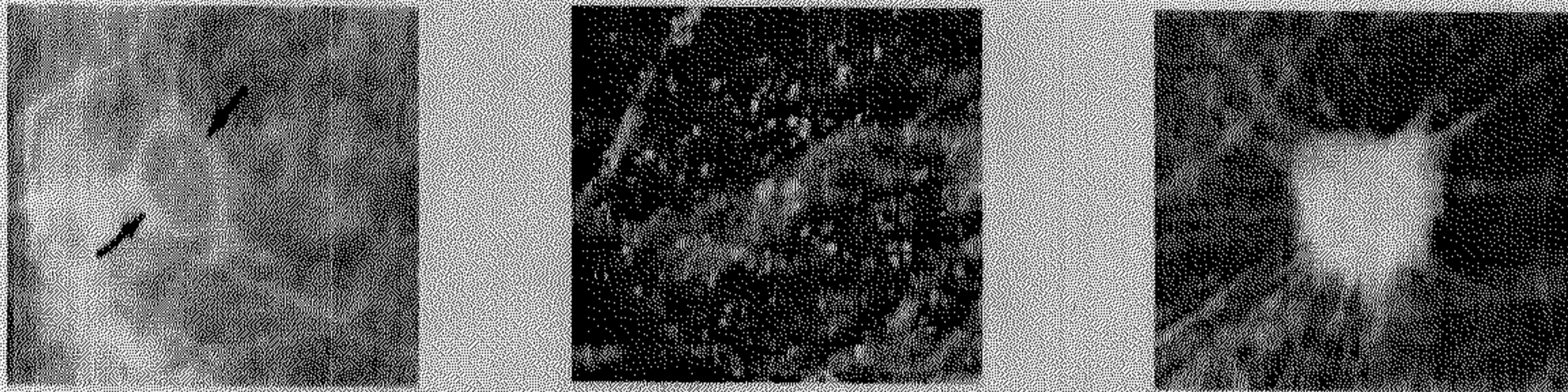


Figure 1.1: *Circumscribed Lesions, Microcalcifications and Speculated Lesions.*

microcalcifications, sample matching seems to be impossible.

(c) *Low contrast* (intensity difference between suspicious areas and their surrounding tissues) makes it difficult to detect.

(d) The *dense tissues* and skin thickening cause suspicious areas to be almost invisible, (misinterpreted as calcifications ) yielding a high false-positive (FP) rate.

One of the difficult cases is shown in Fig 1.2 and its calcified areas are shown in Fig 1.3. There are two calcified regions enclosed by circles. The smaller calcification cluster is very difficult to detect.

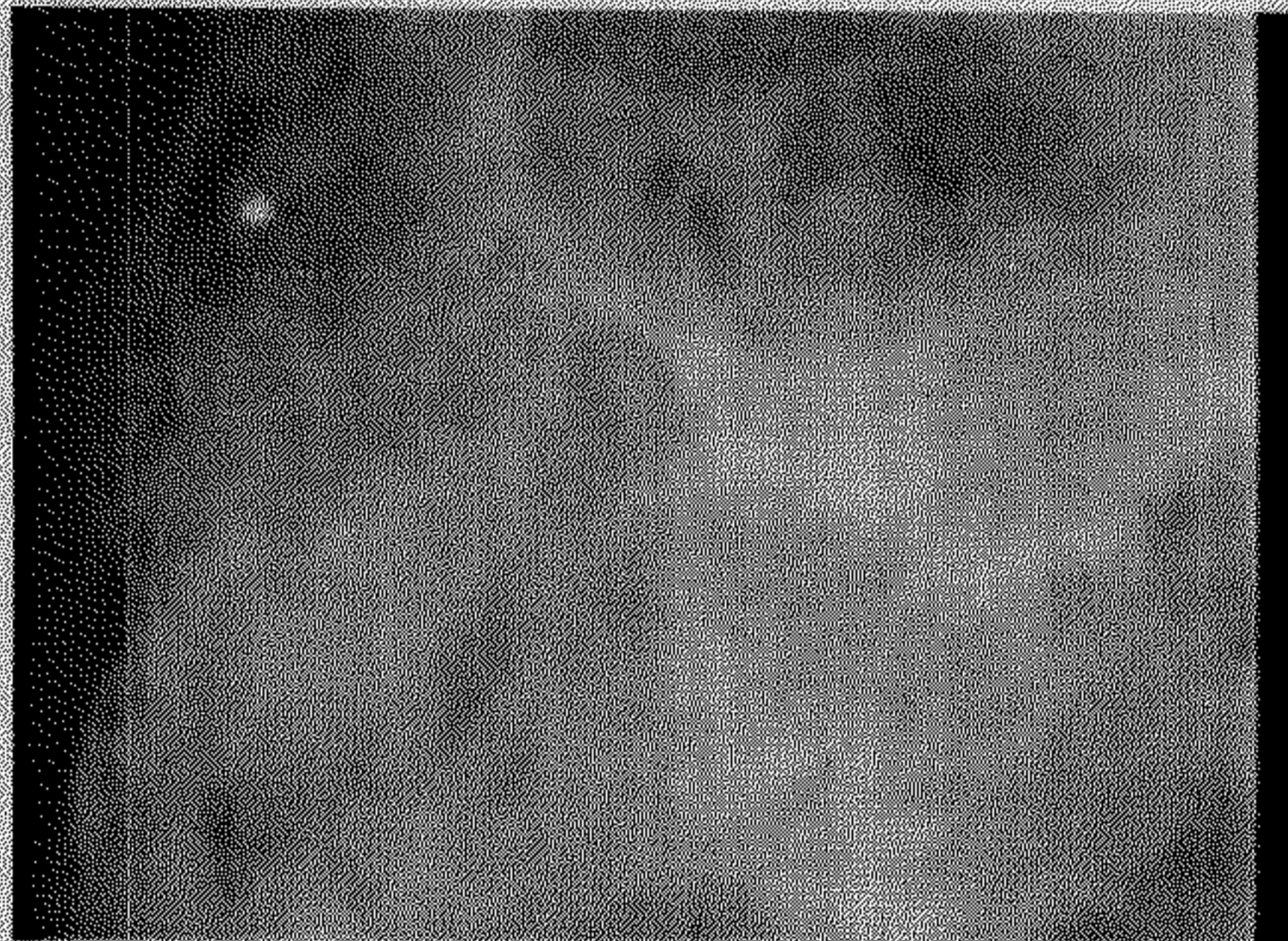


Figure 1.2: *Clustered Microcalcification.*



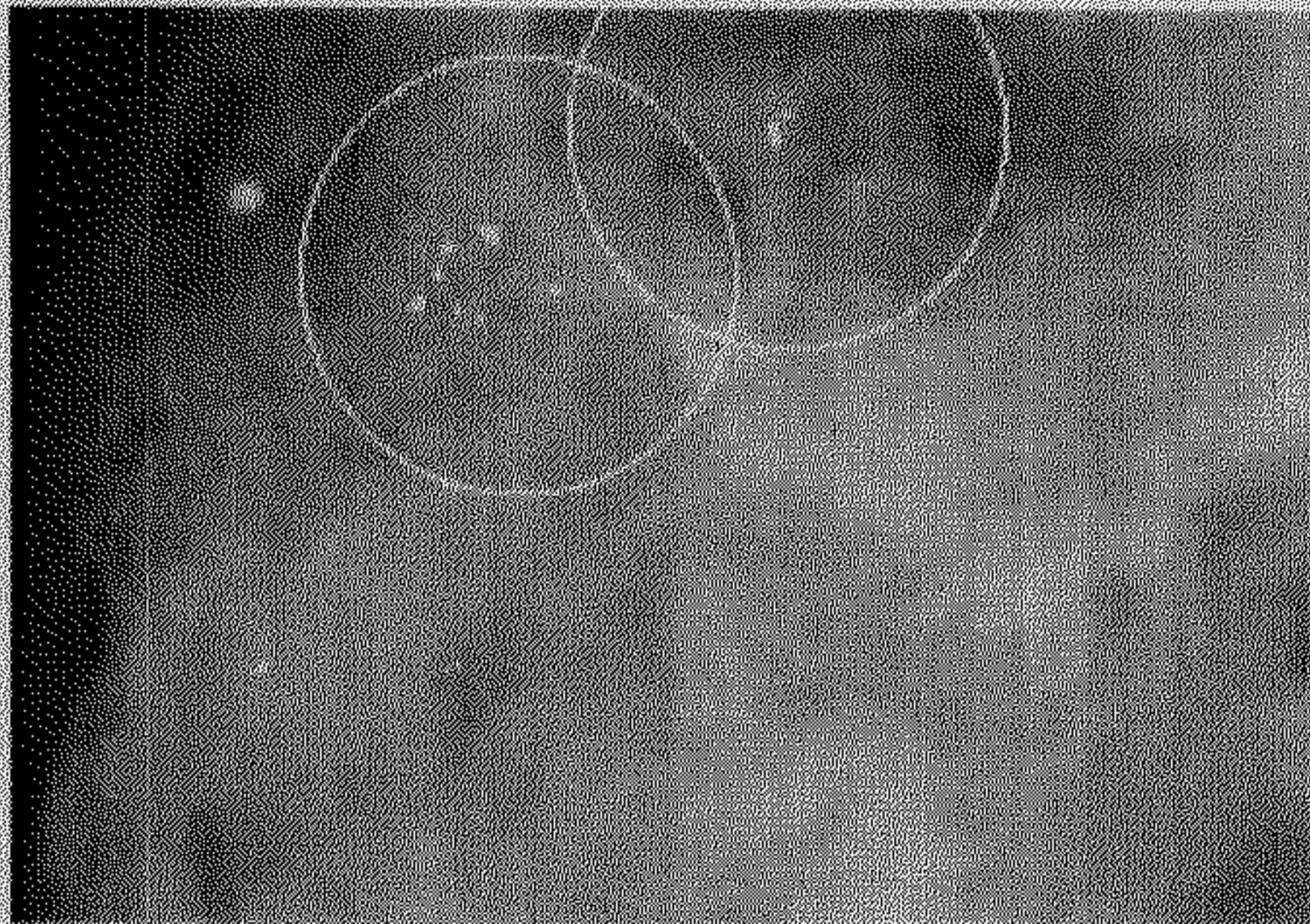


Figure 1.3: *Calcified Result of Fig 1.2.*

## 1.2 Mammography CAD system

A typical computer-aided mammography screening system can be described by the block diagram shown in Fig 1.4. The preprocessing block includes digitization of the mammograms with different sampling and quantization rates. Then, the regions of interests selected from the digitized mammogram are denoised and enhanced. The segmentation block is designed to find suspicious areas containing MCCs, and to separate the MCCs from the background that will be used for extracting features of MCCs. In the feature extraction and selection block, the features of MCCs will be extracted and selected, and MCCs will be classified into benign, malignant and normal.

### 1.2.1 Enhancement of Microcalcifications

Mainly, image enhancement includes intensity and contrast manipulation, noise reduction, background removal, edge sharpening, filtering, etc. The task of mammogram enhancement is to sharpen the edges or boundaries of ROIs, or to increase the contrast between ROIs and background. Although microcalcifications usually are brighter than their surroundings, microcalcifications in a dense breast are with quite low contrast that human eyes can hardly distinguish. The aim of contrast enhancement is to increase the contrast of microcalcifications against their background.

Enhancement of Mammogram can be done in various ways, conventional enhancement techniques includes contrast stretching techniques, histogram



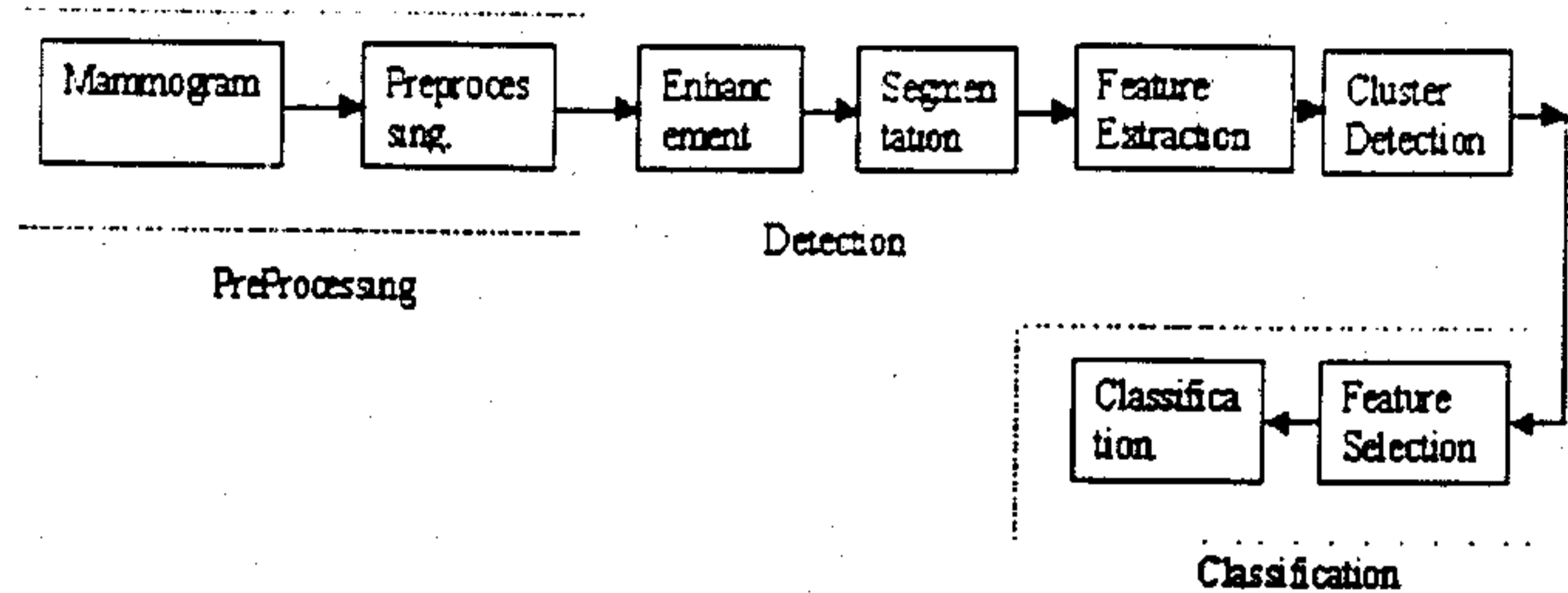


Figure 1.4: Block Diagram of mammography CAD system.

equalization, unsharp masking, spatial filtering. Region-based approach enhances the contrast of the mammographic features of ROIs with various sizes and shapes according to the change of their surroundings. Region-based method can enhance more anatomical detail without significantly introducing artifacts, and has demonstrated that it can identify calcifications more effectively in the image of dense breasts where the contrast between calcifications and breast tissue is quite low [5]. Feature based enhancement techniques can be performed for mammogram analysis through microcalcification features. Since mammograms have some degree of fuzziness such as indistinct borders, ill-defined shapes, and different densities, sometimes the original images are transformed into a fuzzified image on the basis of microcalcification features [2, 6] and fuzzified image is processed.

### 1.2.2 Segmentation of Microcalcifications

The goal of mammogram segmentation is to obtain the suspicious regions. In general, segmentation is to divide the image  $I$  into non-overlapping regions  $S_i$  such that,

$$\bigcup S_i = I \text{ and } S_i \cap S_j = \Phi \text{ where } i \neq j.$$

The extraction of objects from the background can be done by image segmentation. The segmentation can be done by choosing a threshold if the background is uniform. However, due to the variations in shapes, sizes and intensities of microcalcifications, it is difficult to choose a threshold to segment the calcification.

Some possible ways of segmentation include are region growing method

and edge detection method.

### 1.2.3 Micro-calcification detection based on feature extraction

Many researchers use features extracted from mammogram to directly describe individual micro-calcification. Some of the possible features are: the area of the object, average grey level of the object, gradient strength of the perimeter pixels of the object, contrast, perimeter, compactness, elongation, eccentricity, thickness, orientation, the mean intensity level of the background, fourier descriptor, low order moment based on shape descriptor and many more. Experimental results showed that the back propagation neural network could reduce the false detection rate by 42%, and a false detection rate of 3.15 per image was obtained [13].

### 1.2.4 Classification of Calcification

It is usually very difficult to distinguish benign from malignant MCCs because of the variability associated with their appearances. The human breast varies considerably in composition, and mammographic appearances vary from relative uniformity to complex patterns of bright streaks or blobs. The major features used to distinguish normal, benign and malignant MCCs are :

- Gray scale *descriptor*.
- Shape *descriptor*.
- Cluster *descriptor*.

The classifier can be of various type: Neural Network classifiers, K-nearest neighbor classifiers, Bayesian classifiers, Decision tree...etc.

## 1.3 Evaluation of Calcification Detection Algorithms

A decision for a detection result can be either correct (true) or incorrect (false). A decision for a detection result, therefore, will be one of four possible categories as shown in Table 1.1: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). FN and FP are two kinds of errors. A false negative error implies that a true abnormality was not detected, and a false positive error occurs when a normal region was falsely identified as abnormality. A TP decision is a correct judgment of an actual abnormality, and a TN decision means a normal region was correctly labeled.



Table 1.1: A decision for the detection of calcification

CAD	A cancer	Not a Cancer
Claimed a Cancer	True Positive	False Positive
Not Claimed a Cancer	False Negative	True Negative

The performance of diagnostic systems has been measured with two terms: *sensitivity* and *specificity*. where  $\text{sensitivity} = \text{TPs} / (\text{TPs} + \text{FNs})$  and  $\text{specificity} = \text{TNs} / (\text{TNs} + \text{FPs})$ . There is always a tradeoff between sensitivity and specificity which can be shown in a ROC(receiver operating characteristic) curve. ROC curve is a plot of operating points which can be considered a plot of true positive as a function of false positive. For evaluating true-positive detection, sometimes it is required not only to detect the existence but also the localization of the calcification. A better method for this case is free-response receiver operating characteristic (FROC) analysis which is a plot of operating points showing the tradeoff between the TP rate versus the average number of false positives per image.

## 1.4 Review of Some Existing Methods

In the literature, various techniques are described to detect and classify the presence of microcalcifications in digital mammograms as benign or malignant: classical image processing techniques [10], wavelet-based techniques [12], statistical techniques, neural networks based techniques [11].

Mascio [10] developed a microcalcification detection algorithm, which operates on digital mammograms by combining morphological image processing with arithmetic processing. The first analysis emphasizes any detail in the image that changes sharply in intensity and is larger than some specific size. The second analysis emphasizes any detail that is small and textured. Areas that are common to both analysis are segmented and kept for thresholding. This resulted in the detection of microcalcifications and suspicious areas.

Wavelet-based techniques have also been used in the detection of microcalcifications in digital mammograms. Yoshida [12] developed a system based on the wavelet transform. In the wavelet transform, all of the wavelets are derived from scaling and the translation of a single function. Yoshida used the least asymmetric Daubechies wavelets in combination with a difference image technique. These methods are useful in separating microcalcifications from

normal background tissues and achieve a detection rate of approximately 90%.

Woods modified the KNN algorithm, stating that unknown test pattern is assigned to a particular class if at least  $k$  of the KNNs are in that class. The KNN rule will be more sensitive to microcalcification detection and less sensitive to non-microcalcifications.

Verma and Zakos [3] have made an attempt to design a CAD (computer aided diagnosis) system for digital mammograms based on fuzzy-neural and feature extraction techniques. They tried various combinations of features to find important sets of features for classification of micro-calcification.

Automated breast cancer detection has been studied for more than 20 years, the CAD mammography systems for micro-calcification detection have gone from crude tools in the research laboratory to commercial systems. Yet there is well accepted system.

Although by now some progress has been achieved, there are still remaining challenges for future research, such as: developing better enhancement and segmentation algorithms, designing better feature detection and selection algorithms, integration of classifiers to reduce both FPs and FNs.

## Chapter 2

# On-line Feature Selection with Regularization using Neural Networks

Feature selection for classification aims to select a subset of features from the available feature set to obtain improved performance in classification of the given pattern set.

Several researchers have emphasized the importance of feature analysis and proposed methods. But these methods typically use feature ranking based on some criterion and then use a few top ranked features to design a classifier. There are few basic problems associated with such a method: (i) two correlated features may get high ranks, when only one of them may be enough for the task, (ii) it may not be able to detect the effect of a group of features together, (iii) it ignores the fact that importance of a feature depends on the problem being solved and the tool used to solve the problem and thus fail to capture the subtle interaction between features and tools. The third point emphasizes the fact that the best set of features, for example, for a k-NN classifier may not be the best for a neural network classifier. So the best strategy would be to use a scheme that can select the necessary features while learning (constructing) the classifier. We call such methods *Online Feature Selection (OFS)* methods [4].

The OFS is a neural network based technique to choose a subset of features while learning to classify the data set. In this scheme the inputs to the network are attenuated by an attenuation function before they pass into the network. After the training is over, the features may be selected based on their importance as reflected by the values of the attenuators associated with different input nodes[4]. It has been observed that if there are some features with very little discriminating power, then if the network is trained



long enough, the gates corresponding to those features get opened. Under such circumstances it may be difficult to select the top ranked features. That is why when the number of misclassification or the sum of square error(SSE) reduces to some acceptable level in OFS, the training is stopped to select the features. And the network is then retrained with the selected features. Although OFS has been successfully applied in several areas based on the above principle, it would be better, if we can eliminate the above problem. This may be achieved by using a regularizing term in the learning process, which compromises between the reduction in SSE(sum of square error)and the gate opening. We achieve this by adding a penalty term in objective function of the OFS method.

## 2.1 Feature selection Method Using OFS

In a standard multilayer perceptron network, the effect of some features (inputs) can be eliminated by not allowing them into the network, i.e., by equipping each input node (hence each feature) with a *gate* and closing the gate. For good features the associated gates can be completely opened. On the other hand, if a feature is bad, then the corresponding gate should be completely closed. Pal and Chintalapudi [4] suggested a mechanism for realizing such a gate so that *useful* features can be identified and attenuated according to their relative usefulness. In order to model the gates we associate a gate function to each node in the input layer. A gate function should produce a value of 1 or nearly 1 for a good feature; while for a bad feature, it should be nearly 0. To use the gate we multiply the input feature value by its gate function value and the modulated feature value is passed into the network. The gate functions attenuate the features before they propagate through the net, so we may call these gate functions *attenuating* functions. A useful gate function  $F_i : R \rightarrow [0, 1]$  should have a tunable parameter and should be differentiable with respect to the tunable parameter. It should be monotonic with respect to its tunable parameter. The sigmoidal function satisfies the above criteria and in this paper we have used it. Other choices are also possible.

The basic philosophy of learning would be to keep all gates almost closed at the beginning of the learning (i.e. no feature is important) and then open the gates as required during the training. To complete the description in connection with MLP, let  $F_i$  be the gate or attenuation function associated with the  $i^{th}$  input feature.  $F_i$  has an argument  $m_i$ ,  $F_i'(m_i)$  be the value of derivative of the attenuation function at  $m_i$ ,  $\mu$  be the learning rate of the attenuation parameter;  $\eta$  be the learning rate of the connection weights.

$X_i$  be the  $i^{th}$  input of an input vector  $\mathbf{X}$ :  $X'_i$  be the attenuated value of  $X_i$ .

$$X'_i = X_i F(m). \quad (2.1)$$

Let  $W_{ij}^0$  be the weight connecting the  $j^{th}$  node of the first hidden layer to the  $i^{th}$  node of the input layer; and  $\delta_j^1$  be the error term for the  $j^{th}$  node of the first hidden layer.

The output of the first hidden layer is  $O_i^1 = f(\sum_j X'_j W_{ij}^0)$  and for other layers  $O_i^k = f(\sum_j X_j^{k-1} W_{ij}^{k-1})$  for  $k=2,3..n+1$ . The weights and the gate parameters are adjusted to minimize  $\sum \epsilon$  using gradient descent on  $\epsilon$ . where

$$\epsilon = 1/2 \sum (E_i)^2, \text{ where } E_i = (T_i - O_i^{n+1}). \quad (2.2)$$

In (2.2),  $T_i$  is the target output from  $i^{th}$  output node and  $O_i^{n+1}$  is the computed output at the  $i^{th}$  output node.

The weights of the network are adjusted exactly in the same manner as per the backpropagation neural network rule except  $w_{ij}^0$ , which is modified by:

$$\Delta W_{ij}^0 = \eta \delta_i^1 X'_j = \eta \delta_i^1 X_j F(m_j), O_i^{n+1} = f(\sum_j W_{ij}^n O_j^n). \quad (2.3)$$

It can be easily proved that the adjustment of the attenuator parameter is done using:

$$\Delta m_i = \mu X_i F'_i \sum_j (W_{ji}^0 \delta_j^1). \quad (2.4)$$

As mentioned earlier, for the gate function, several choices are possible but we use here the sigmoidal function  $F(m) = 1/(1 + e^{-m})$ . The input gate parameters are so initialized that when the training starts  $F(m)$  is practically zero for all gates, i.e., no feature is allowed to enter the network. As the gradient descent learning proceeds, gates for the features that can reduce the error faster are opened faster. The learning of the gate function continues along with other weights of the network. At the end of the training we can pick up important features based on the values of the attenuation function.

Typically, the training is stopped when the training error or misclassification is reduced to an acceptable level. Note that, for different initializations different feature set may be selected. If this happens, then this indicates that there are different sets of features that can do the classification job equally well. One may rank the features based on the extent the gates are opened and use a set of top ranked features. But note that feature ranking is not the objective OFS. The algorithm is expected to do a good job because OFS looks at all the features at a time during the training process.

## 2.2 Regularization of OFS system

Our goal is not to allow the gates to open corresponding to the features with low discriminating power. Also we should prevent use of correlated features as far as possible. To achieve this, we add a penalty term to the instantaneous error  $\epsilon_j = \sum_i (T_i - O_i^{n+1})^2$ , thus making it:

$$\epsilon_j = \sum_i^c (T_i - O_i^{n+1})^2 + K_r \times g(\mathbf{M}). \quad (2.5)$$

In (2.5)  $K_r > 0$  is a constant and  $g(\mathbf{M})$  is a function of the  $p$  gate opening parameters.

$\mathbf{M} = (m_1, m_2, \dots, m_p)^T$  gate opening parameter vector.

$p$  is the dimension of the input.

The function  $g(\mathbf{M})$  should increase with the opening of the gates. Since we are using sigmoidal function for the gate opening,  $g$  should increase with  $m_i$ . So the total error is  $E = \sum_j^N \epsilon_j$ . Our objective would be minimize  $E$ . To minimize  $E$ , preferably one should use *batch update* rules. However, like the back propagation learning algorithm, here also we use the instantaneous error to derive the learning rule.

As already discussed, regularization during the training of OFS system compromises between the SSE and gate opening. After adding this penalty-term in the error it can be shown that the learning rule of the attenuator becomes:

$$\Delta m_l = \mu X_l F_l' \sum_j (W_{ji}^0 \delta_j^1) - \mu K_r \times g'(\mathbf{M}). \quad (2.6)$$

For not allowing all the gates to open, one of the possible choice of  $g$  can be of the form  $g(\mathbf{M}) = \sum_l \exp^{-(k-m_l)}$ ,  $k > 0$  is a constant. So the equation for  $\epsilon_j$  becomes:

$$\epsilon_j = \sum_i (T_i - O_i^{n+1})^2 + K_r \times \sum_l \exp^{-(k-m_l)} \quad (2.7)$$

Here we considered  $k$  to be 6 because the gate opening becomes almost 1 when  $m_i$  equals to 6. And the learning rule for the attenuator becomes:

$$\Delta m_l = \mu X_l F_l' \sum_j (W_{ji}^0 \delta_j^1) - \mu K_r \times \exp^{-(k-m_l)}. \quad (2.8)$$



## 2.3 Data Set Description

We prepared three sets of synthetic data. Each set has 200 patterns, with four features. There are 100 patterns for each class. Our third data set is the Iris data set.

### Data set 1

The data set 1 has been created with four features. The first two features are shown in Fig 2.1, where the two classes are displayed with dots and stars. Then we have added two random components in the range 0-20. So feature 3 and 4 are not useful features.

### Data set 2

The data set 2 has been created with five features. The first two features are shown in Fig 2.1, where the two classes are displayed with dots and stars. Then we have added two random components in the range 0-20. So feature 3 and 4 are not useful features. The fifth component is twice the value of the first component. So the first and fifth features are strongly correlated features.

### Data set 3

The data set 3 has been created with four features. The first two features are shown in Fig 2.1. And the third and four features are shown in Fig 2.2, where the two classes are displayed with dots and stars. Note that here either  $(F_1, F_2)$  or  $(F_3, F_4)$  can do the classification job.

### Data set 4

Here we considered Iris data set. This is the well-known Anderson's Iris data set [8]. It contains a set of 150 measurements in four dimensions taken on Iris flowers of 3 different species or classes. The four features are sepal length, sepal width, petal length and petal width. The data set contains 50 instances of each of the three classes.

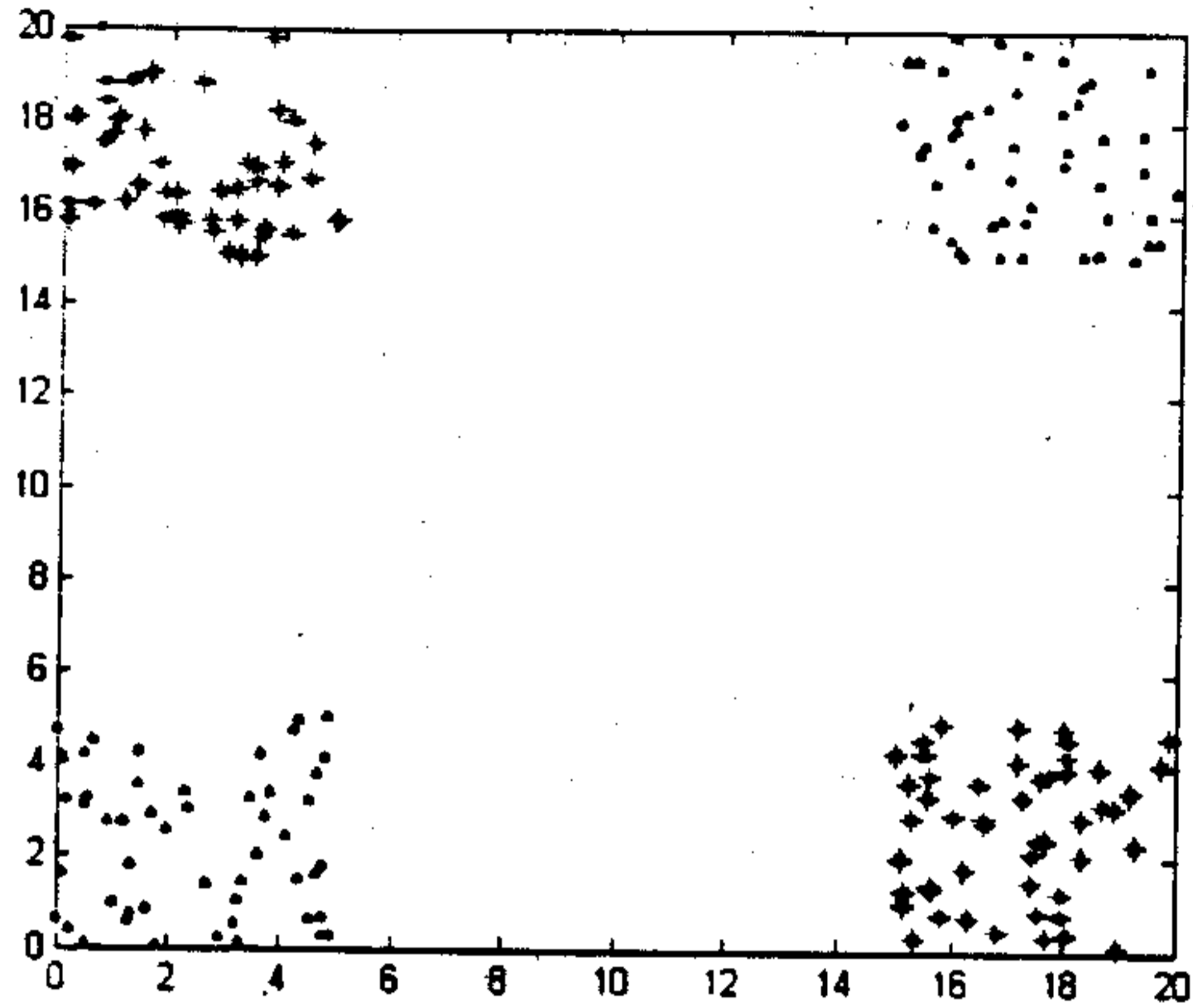


Figure 2.1: *Data set for feature 1 and 2. · for class 1 and \* for class 2.*

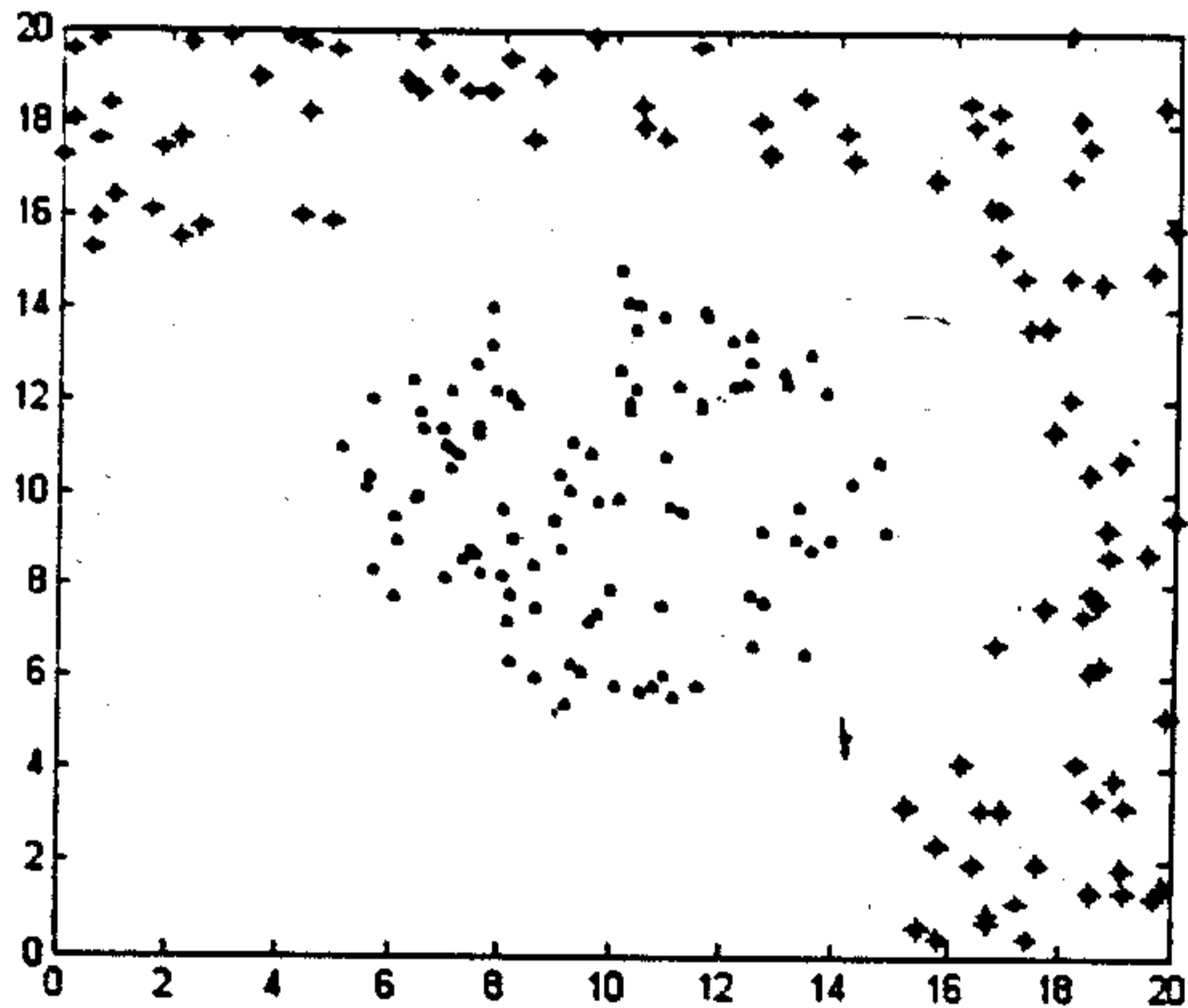


Figure 2.2: *Data set for feature 3 and 4. · for class 1 and \* for class 2.*

Table 2.1: Results on DataSet1 with and without Regularization :  $\eta = 0.9$  and  $\mu = 0.3$ .

$K_r$	Random initialization of network					
	Initialization 1		Initialization 2		Initialization 3	
	Gate opening	SSE, mis-classification	Gate opening	SSE, mis-classification	Gate opening	SSE, mis-classification
0.0	0.9800	0.0009, 0	0.9846	0.0012, 0	0.9795	0.0009, 0
	0.9812		0.9821		0.9800	
	0.0082		0.0089		0.0108	
	0.2596		0.2525		0.2443	
0.5	0.1090	0.0102, 0	0.1095	0.0101, 0	0.1063	0.0097, 0
	0.1072		0.1077		0.1045	
	0.0008		0.0004		0.0004	
	0.0001		0.0008		0.0008	

## 2.4 Results and Discussion

We have used a network architecture with four hidden nodes in a single hidden layer. We trained the feature selection network with 30000 epochs. The features are then selected based on the gate opening values.

### Data set 1

The data set 1 has been created in such a way that the two classes can be satisfactorily separated with only the first two features. This is because the third and fourth elements of each pattern have been chosen randomly. Table 2.1 shows the gate opening for three different initializations of the network. Note that,  $K_r = 0$  results in the original OFS network. From Table 2.1 it is clear that the original OFS method opens the first two gates more, as they correspond to the two dominating features. But the fourth gate is also open moderately, it indicates that the fourth feature has little discriminating power. But by applying the penalty with regularization factor 0.5, we find that, only the first two features have much high discriminating power than the other two. The other two gates are almost closed.

This happens consistently for all the three runs. Note that, with regularization, the extent of gate opening for the good features is also reduced significantly, but even with that the network brings the misclassification to



Table 2.2: Results on DataSet 2 (5 features) with and without Regularization :  $\eta = 0.9$  and  $\mu = 0.3$ .

$K_r$	Seed for the random initialization of network					
	Initialization 1		Initialization 2		Initialization 3	
	Gate opening	SSE. mis-classification	Gate opening	SSE. mis-classification	Gate opening	SSE. mis-classification
0.0	0.0098	0.0010, 0	0.0073	0.0010, 0	0.0082	0.0010, 0
	0.9857		0.9842		0.9843	
	0.0082		0.0075		0.0078	
	0.4779		0.4351		0.4312	
	0.9535		0.9459		0.9458	
0.2	0.0010	0.0060, 0	0.0010	0.0060, 0	0.0010	0.0065, 0
	0.1472		0.1470		0.1471	
	0.0011		0.0010		0.0010	
	0.0032		0.0032		0.0033	
	0.1071		0.1084		0.1085	

Table 2.3: Results on DataSet3 with and without Regularization :  $\eta = 0.9$  and  $\mu = 0.3$ .

$K_r$	Random initialization of network					
	Initialization 1		Initialization 2		Initialization 3	
	Gate opening	SSE. mis-classification	Gate opening	SSE. mis-classification	Gate opening	SSE. mis-classification
0.0	0.9631	0.0071, 0	0.9644	0.0058, 0	0.9631	0.0071, 0
	0.9515		0.9513		0.9516	
	0.9587		0.9595		0.9586	
	0.9801		0.9807		0.9800	
2.0	0.1464	0.0978, 0	0.2842	0.3348, 0	0.1460	0.0972, 0
	0.1160		0.2742		0.1157	
	0.0070		0.0051		0.0071	
	0.0049		0.0043		0.0050	

Table 2.4: Results on Iris data set with and without Regularization :  $\eta = 0.9$  and  $\mu = 0.9$ .

$K_r$	Random initialization of network					
	Initialization 1		Initialization 2		Initialization 3	
	Gate opening	SSE, mis-classification	Gate opening	SSE, mis-classification	Gate opening	SSE, mis-classification
0.0	0.9402	4.1850, 2	0.4578	5.093, 2	0.8424	5.2710, 3
	0.9514		0.7967		0.9694	
	0.9513		0.6969		0.9144	
	0.9555		0.8713		0.9447	
2.0	0.0005	7.4498, 6	0.0002	7.9156, 6	0.0002	7.5775, 6
	0.1095		0.1099		0.1007	
	0.0419		0.0190		0.0368	
	0.6203		0.6536		0.6309	

zero.

### Data set 2

The data set 2 has been created in such a way that the first and fifth features are correlated. So our feature selection method expected to select one of them, if required; but not both. Table 2.2 shows the result of the OFS method with and without regularization. And we find that only the second and fifth features are emphasized by the network with regularization. But the original OFS, opens three gates.

### Data set 3

The data set 3 has been created in such a way that the two classes can be satisfactorily separated with the first two features or the last two features. Table 2.3 summarizes the results for this data set. From Table 2.3 it is clear that the original OFS method opens all four gates to get very low sum of squared error. But using the modified OFS with regularization factor 2.0, we find that the network picks only the first two features.

#### Data set 4

This is the Iris data set. Table 2.4 depicts the results for Iris data set. Table 2.4 reveals that the original OFS method opens each of the four gates to a great extent. But the modified OFS with regularization factor 2.0, opens only the second and fourth features. Row 1 of Table 2.4 suggests all four features as important while row 2 suggests only features 2 and 4. It shows that these two features have adequate discriminating power. With usual MLP all four features result in 2 misclassification and only features 2 and 4 can get 2 misclassification. For all three cases, the network emphasizes more on feature four, which is consistent with the known facts.



## Chapter 3

# Extracting Fuzzy Rules With higher Specificity for Classification

The fuzzy logic approach is a different conceptual model to classify objects. It is based on approximate reasoning. The fuzzy set theoretic framework provides a degree of support to each of the potential classes. A set of fuzzy rules is used to define (describe) the class label of each data point. The rules are defined on some attributes which are computed for each data point. After the rule base is obtained, for every data point, the attributes are computed and the degree of match of these attributes with each fuzzy rule is computed. The class label associated with the rule having the strongest match defines the class of the data point.

The methodology that we are going to discuss consists of three steps: generation of the prototypes, conversion of the prototypes into the fuzzy rules, tuning of the fuzzy rules for the given training set. It has been observed that if there are some atypical patterns (which are far away from the center of its class) in the training set then if the rules are tuned for long enough then the spread of the membership function of the rules may go on increasing. This is because during the tuning of the rules we want to reduce the error function  $E$  (see section 3.1.3), which depends on the firing strength and hence on spread of the membership functions. When the spreads of the membership function increases, the specificity of rules becomes low. This is usually not a desirable property as it can lead to poor generalization. That is why when the misclassification or the sum of error reduces to some acceptable level, the training is stopped to get the parameters of the fuzzy rules. Therefore, there is a need to regularize the tuning process, which can compromise between the increase in spreads and the reduction of error and the misclassification.

We achieve this by adding a penalty term with the error, which is a function of the spreads of the rules.

### 3.1 Fuzzy rule based classification method

A fuzzy rule based classifier consists of a set of fuzzy rules of the form:  
 $R_i$ : if  $x_1$  is  $A_{i1}$  AND  $x_2$  is  $A_{i2}$  AND...AND  $x_p$  is  $A_{ip}$  then class is  $j$ . Where  $A_{ik}$  is a fuzzy set used in the  $i^{th}$  rule and is defined on the domain of  $x_k$ , i.e defined on the universe of  $k^{th}$  feature.

When a sample data point  $x \in R^p$  is presented to the system for classification, the fuzzy rules fire to produce outputs. The magnitude of the outputs (also known as firing strengths) are used for deciding the class membership of the sample data  $x$ .

#### 3.1.1 Generation of prototypes

We use the k-means clustering algorithm to find few clusters in the data corresponding to each class *separately*. Consider a three class training set  $X_1, X_2$  and  $X_3$  such that  $X = X_1 \cup X_2 \cup X_3$ ,  $X_q \cap X_r = \Phi$ ,  $r \neq q = 1, 2, 3$  be the training data,  $X_q$  be the training data corresponding to class  $q$ .  $X \subset R^p$ . The number of prototypes for each class may be different. Here we considered two prototypes ( $v_i \in R^p$ ) for each of the classes.

#### 3.1.2 Fuzzy rule generation from the prototypes

A prototype  $v_i$  for class  $k$  can be translated into a fuzzy rule of the form:

$R_i$ : if  $x$  is CLOSE TO  $v_i$  then the class is  $k$ .

where the fuzzy set 'CLOSE TO' is represented by a multidimensional membership function:

$$\mu_{CLOSETO}(x) = \exp \frac{-\|x - v_i\|^2}{\sigma_i^2} \quad (3.1)$$

where  $\sigma_i > 0$  is a constant. This above membership function represents a hyperspherical zone of with center at  $v_i$ . This above function may not perform well when different features have different variances. So this  $x$  'CLOSE TO'  $v_i$  is written as :  $x_1$  is CLOSE TO  $v_{i1}$  and  $x_2$  is CLOSE TO  $v_{i2}$  and ... and  $x_p$  is CLOSE TO  $v_{ip}$ . Here  $v_i = (v_{i1}, v_{i2}, \dots, v_{ip})^T$  and  $x = (x_1, x_2, \dots, x_p)^T$ . In this way we get a set of initial rules. In general, the  $i^{th}$  rule representing one of the  $c$  classes takes the form :

$R_i : x_1 \text{ CLOSE TO } v_{i1} \text{ and...and } x_p \text{ CLOSE TO } v_{ip} \text{ then the class is } k.$

Here  $p$  is the number of features and hence the number of atomic clauses. The fuzzy set CLOSE TO  $v_{ij}$  is modeled by a Gaussian membership function:

$$\mu_{ij}(x_j : v_{ij}, \sigma_{ij}) = \exp\left(-\frac{(x_j - v_{ij})^2}{\sigma_{ij}^2}\right).$$

In this above case each fuzzy rule is characterized by two parameters  $v_{ij}$  and  $\sigma_{ij}$ . The  $v_{ij}$ s of the rules are initialized by the components of the prototypes obtained by the k-means algorithm. The initial estimation of  $\sigma_{ij}$  is computed by the standard deviation of  $j^{\text{th}}$  feature of the  $i^{\text{th}}$  cluster from the training set.

For a given data point  $\mathbf{x}$ , we first find the firing strength of each rule using the product

$$\alpha_i(\mathbf{x}) = \prod_{j=1}^{j=p} \mu_{ij}(x_j : v_{ij}, \sigma_{ij}).$$

Here  $\alpha_i(\mathbf{x})$  is the firing strength of the  $i^{\text{th}}$  rule on a data point  $\mathbf{x}$ . This gives the degree of match between the data point  $\mathbf{x}$  and the antecedent of the  $i^{\text{th}}$  rule. Now class label of the rule having the maximum firing strength determines the class of the data point  $\mathbf{x}$ . Let  $l = \underbrace{\text{argmax}}\{\alpha_i(\mathbf{x})\}$ , and suppose

the  $l^{\text{th}}$  rule represents class  $c$ , then  $\mathbf{x}$  is assigned to class  $c$ .

### 3.1.3 Tuning of the fuzzy rules

To improve the performance of the rule base, we now refine the rules to minimize the training error using gradient descent technique. In this regard, appropriate learning rules are also derived.

Let  $\mathbf{x} \in X$  be from the class  $c$  and  $R_c$  be the rule from class  $c$  giving the maximum firing strength  $\alpha_c$  for  $\mathbf{x}$ .

Also let  $R_{\neg c}$  be the rule from the incorrect classes having the maximum firing strength  $\alpha_{\neg c}$  for  $\mathbf{x}$ . We use the error function  $E$  to refine the rules:

$$E = \sum_{\mathbf{x} \in X} (1 - \alpha_c(\mathbf{x}) + \alpha_{\neg c}(\mathbf{x}))^2.$$

The rules are refined by minimizing  $E$  with respect to the centroids ( $v_{cj}$ ,  $v_{\neg cj}$ ) and spreads ( $\sigma_{cj}$ ,  $\sigma_{\neg cj}$ ) associated with  $R_c$  and  $R_{\neg c}$ . This may be viewed as refining the rules with respect to their *contexts* in the feature space. The rule refinement algorithm is given below.



The rule refinement algorithm

Begin

Choose learning parameters  $\eta_m$  and  $\eta_s$

Choose a parameter reduction factor  $0 < \varepsilon < 1$

Choose the maximum number of iterations. *maxiter*.

Compute the error  $E_0$  for the initial rule base  $R^0$ .

Compute the misclassification  $M_0$  corresponding to initial rule base  $R^0$ .

$t \leftarrow 1$

While ( $t \leq \text{maxiter}$ ) do

For each vector  $\mathbf{x} \in X$

Find the rules  $R_c$  and  $R_{\neg c}$ .

Modify the parameters of rules and as follows.

For  $k=1$  to  $p$

$$\begin{aligned} v_{ck}^{\text{new}} &= v_{ck}^{\text{old}} - \eta_m \frac{\partial E}{\partial v_{ck}^{\text{old}}} \\ &= v_{ck}^{\text{old}} + \eta_m (1 - \alpha_c + \alpha_{\neg c}) \frac{\alpha_c}{\sigma_{ck}^{\text{old}^2}} (x_k - v_{ck}^{\text{old}}) \end{aligned}$$

$$\begin{aligned} v_{\neg ck}^{\text{new}} &= v_{\neg ck}^{\text{old}} - \eta_m \frac{\partial E}{\partial v_{\neg ck}^{\text{old}}} \\ &= v_{\neg ck}^{\text{old}} - \eta_m (1 - \alpha_c + \alpha_{\neg c}) \frac{\alpha_{\neg c}}{\sigma_{\neg ck}^{\text{old}^2}} (x_k - v_{\neg ck}^{\text{old}}) \end{aligned}$$

$$\begin{aligned} \sigma_{ck}^{\text{new}} &= \sigma_{ck}^{\text{old}} - \eta_s \frac{\partial E}{\partial \sigma_{ck}^{\text{old}}} \\ &= \sigma_{ck}^{\text{old}} + \eta_s (1 - \alpha_c + \alpha_{\neg c}) \frac{\alpha_c}{\sigma_{ck}^{\text{old}^3}} (x_k - v_{ck}^{\text{old}})^2 \end{aligned}$$

$$\begin{aligned} \sigma_{\neg ck}^{\text{new}} &= \sigma_{\neg ck}^{\text{old}} - \eta_s \frac{\partial E}{\partial \sigma_{\neg ck}^{\text{old}}} \\ &= \sigma_{\neg ck}^{\text{old}} - \eta_s (1 - \alpha_c + \alpha_{\neg c}) \frac{\alpha_{\neg c}}{\sigma_{\neg ck}^{\text{old}^3}} (x_k - v_{\neg ck}^{\text{old}})^2 \end{aligned}$$

End For

End For

Compute the error  $E_t$  for the new rule base  $R^t$ .

Compute the misclassification  $M_t$  for  $R^t$ .

If  $M_t > M_{t-1}$  or  $E_t > E_{t-1}$   
 then  
 $\eta_m \leftarrow (1 - \varepsilon)\eta_m$   
 $\eta_s \leftarrow (1 - \varepsilon)\eta_s$   
 $R^t \leftarrow R^{t-1}$   
 /\* If the error is increased, then possibly the  
 learning coefficients are too large. So, decrease  
 the learning coefficients and restore the rule base to  $R^t$  . \*/  
 If (  $M_t \leq$  a threshold ( $T_1$ ) or  $E_t \leq$  a threshold ( $T_2$ ) )  
 then Stop  
 $t \leftarrow t + 1$   
 End while  
 End

At the end of the rule base tuning we get the final rule base  $R^{final}$  which is expected to give a very low error rate.

### 3.2 Regularization of Fuzzy Rules

The rules extracted by the above scheme or its variants work fine and have been used in many applications. However, the rules extracted by this method may have a very low specificity, particularly when the training data set has a few outlier points. This happens because, to reduce  $E$ , the learning process may increase the spreads significantly. To avoid this problem we add a penalty for larger spreads. Our goal is to make a compromise between the increase in spreads of the rules and the reduction in the error and misclassification. Hence we add a penalty term in the error equation  $E = \sum_{\mathbf{x} \in X} (1 - \alpha_c(\mathbf{x}) + \alpha_{-c}(\mathbf{x}))^2$ . So the error function changed to:

$$E = \sum_{\mathbf{x} \in X} (1 - \alpha_c(\mathbf{x}) + \alpha_{-c}(\mathbf{x}))^2 + K_r \times g(\sigma). \quad (3.2)$$

In (3.2)  $K_r > 0$  is a constant.

We have already discussed about regularization during the training of OFS system that makes a compromise between the error and gate opening. Here we use a similar concept. Here  $g(\sigma)$  is a function of all spreads.  $\sigma = (\sigma_{11}, \sigma_{12}, \dots, \sigma_{np})^T$  the vector of all spreads,  $n$  is the number of rules.  $g(\sigma)$  should be monotonically increasing function of  $\sigma$ . For every data point, since we adjust only the parameters of the best and worst rules for that data point, it is enough to find the update rules for ( $\sigma_{ck}$  and  $\sigma_{-ck}$ ) for  $k=1,2,\dots,p$ . After

adding this Regularization-term to the error function it can be easily shown that the adjustment of spreads ( $\sigma_{ck}$  and  $\sigma_{-ck}$ ) in the tuning algorithm(see section 3.1.3) are modified to:

$$\sigma_{ck}^{new} = \sigma_{ck}^{old} + \eta_s(1 - \alpha_c + \alpha_{-c}) \frac{\alpha_c}{\sigma_{ck}^{old^3}} (x_k - v_{ck}^{old})^2 + \eta_s K_r \times g'(\sigma_{ck}). \quad (3.3)$$

$$\sigma_{-ck}^{new} = \sigma_{-ck}^{old} - \eta_s(1 - \alpha_c + \alpha_{-c}) \frac{\alpha_{-c}}{\sigma_{-ck}^{old^3}} (x_k - v_{-ck}^{old})^2 - \eta_s K_r \times g'(\sigma_{-ck}) \quad (3.4)$$

For not allowing the uncontrolled increase of the spreads, one of the possible penalty terms can be of the form  $g(\sigma) = \sum_{ij} \exp^{-(D-\sigma_{ij})}$ ,  $D > 0$ . So the equation for new E becomes:

$$E = \sum_{i \in X} (1 - \alpha_c + \alpha_{-c})^2 + K_r \times \sum_{ij} \exp^{-(D-\sigma_{ij})} \quad (3.5)$$

Here  $D > 0$  is a constant. D is used just to keep the value of penalty moderate i.e, to scale the value of penalty.

### 3.3 Results and Observations

We use the same data sets except data set 2 as discussed in Section 2.3 to demonstrate the fuzzy rule based system. We do not use data set 2 as for OFS data set 2 was used to show the effect of correlated features. The parameters used for the tuning process are  $\eta_m = 0.8$ ,  $\eta_s = 0.3$  and the maximum iteration used is 500.

#### Data set 1

The result is shown the Table 3.1. During the tuning process the sum of squared error(SSE) is reduced with the increase in the spreads. Row 2 suggests that some of  $\sigma$  values go to as high as 22.58, making a rule with very low specificity. The proposed system checks the uncontrolled growth of  $\sigma$ s at the cost of higher SSE, but still it attains zero misclassification. It is observed that with the higher values of regularization factor  $K_r$ , the increase in spread is controlled more.

#### Data set 3

Table 3.2 shows the result for data set 3. The tuning process reduces the sum of squared error(SSE) with the increase in the spreads. But by applying the penalty the uncontrolled growth of  $\sigma$ s has been checked at the cost of higher SSE, but still it attains zero misclassification.



Table 3.1: Result on Data Set 1 with and without regularization  $\eta_s = 0.3$  and  $\eta_m = 0.2$

	Fuzzy sets $\sigma$ and $\mu$								SSE. mis-classification
	spread( $\sigma$ )				prototype ( $v$ )				
Initial Rules	1.511	1.522	5.510	6.583	17.694	17.315	10.481	9.821	165.575. 0
	1.511	1.350	5.359	5.878	2.754	2.682	10.237	10.363	
	1.416	1.435	5.541	4.632	2.629	17.634	10.692	10.500	
	1.499	1.421	5.596	5.373	17.755	2.049	9.509	11.355	
Tuned Rules	7.248	7.324	20.677	22.580	17.781	17.447	10.461	9.903	10.631. 0
	7.091	7.134	20.380	21.304	2.651	2.529	10.019	10.159	
	7.081	7.160	20.765	18.798	2.486	17.701	10.694	10.333	
	7.257	7.161	20.708	20.373	17.886	2.098	9.459	11.341	
$k_r = 0.5$ Rules	6.973	7.057	14.945	16.348	17.777	17.438	10.497	9.919	18.974. 0
	6.845	6.743	14.627	15.394	2.681	2.586	10.118	10.334	
	6.780	6.866	14.901	13.521	2.517	17.690	10.717	10.396	
	6.950	6.881	15.076	14.554	17.849	2.069	9.405	11.584	
$k_r = 1.5$ Rules	6.819	6.885	14.069	15.373	17.765	17.422	10.502	9.904	27.251. 0
	6.700	6.535	13.776	14.535	2.698	2.608	10.164	10.393	
	6.600	6.700	14.020	12.728	2.536	17.677	10.742	10.432	
	6.782	6.696	14.243	13.694	17.826	2.070	9.397	11.643	

#### Data set 4

The result on Iris data is shown the Table 3.3. During the tuning process the sum of squared error(SSE) is reduced with the increase in the spread. In this case also regularization leads to rules with higher specificity. Note that, for Iris, the number of rules can be easily reduced, but we did not try to optimize the number of rules. Here our objective is to demonstrate the effectiveness of regularization.

Table 3.2: Result on Data Set 3 with and without regularization  $\eta_s = 0.3$  and  $\eta_m = 0.2$

	Fuzzy sets $\sigma$ and $\mu$								SSE, mis- classification
	spread( $\sigma$ )				prototype ( $\nu$ )				
Initial Rules	1.369	1.357	2.477	2.439	17.292	17.470	10.380	10.493	164.384, 0
	1.638	1.455	2.402	2.489	2.257	2.950	10.212	10.089	
	1.434	1.287	6.551	6.828	17.506	3.070	14.161	12.387	
	1.332	1.523	6.018	5.884	2.156	17.305	13.568	14.176	
Tuned Rules	7.469	7.330	12.656	12.839	17.448	17.541	10.288	10.358	8.546, 0
	7.852	7.175	12.355	12.845	2.134	2.831	10.179	10.129	
	7.346	6.871	22.605	22.969	17.596	2.962	13.535	12.154	
	7.289	7.181	21.486	21.395	2.041	17.406	13.005	13.716	
$k_r = 0.5$ Rules	7.301	7.299	12.412	12.587	17.451	17.550	10.294	10.359	11.706, 0
	7.833	6.988	12.078	12.538	2.126	2.826	10.177	10.133	
	7.382	6.892	20.71	20.714	17.599	2.966	13.604	12.141	
	7.410	7.304	16.204	16.194	2.011	17.423	13.054	13.714	
$k_r = 5.0$	7.311	7.290	11.019	11.113	17.437	17.522	10.310	10.387	13.5187, 0
	7.896	7.033	10.708	11.056	2.148	2.832	10.172	10.142	
	7.278	6.784	22.032	22.032	17.595	2.972	13.969	12.120	
	7.270	7.251	14.279	14.274	2.021	17.422	13.217	13.880	

Table 3.3: Result on Iris Data set with and without regularization  $\eta_s = 0.3$  and  $\eta_m = 0.2$

	Fuzzy sets $\sigma$ and $\mu$								SSE and misclassification
	spread( $\sigma$ )				prototype ( $v$ )				
Initial Rules	0.236	0.280	0.153	0.119	5.256	3.670	1.504	0.289	119.6803. 3
	0.212	0.247	0.188	0.060	4.713	3.122	1.417	0.191	
	0.307	0.266	0.363	0.149	5.533	2.604	3.883	1.183	
	0.369	0.278	0.223	0.136	6.308	2.923	4.608	1.458	
	0.395	0.277	0.291	0.269	6.168	2.864	5.175	1.943	
	0.457	0.329	0.413	0.250	7.123	3.114	6.032	2.132	
Tuned Rules	1.426	1.470	0.991	0.817	5.004	3.449	1.448	0.251	30.8614. 3
	0.279	0.148	0.006	0.460	4.621	2.909	1.160	0.007	
	1.041	1.140	1.292	0.732	5.644	2.660	4.025	1.208	
	0.537	0.683	0.654	0.499	6.741	3.091	4.665	1.488	
	1.954	0.616	0.432	0.623	4.869	2.633	4.922	1.990	
	1.368	1.121	1.189	1.035	6.964	3.126	5.937	2.151	
$k_r = 0.1$	0.241	0.284	0.158	0.130	5.253	3.669	1.502	0.284	117.3722. 3
	0.218	0.248	0.193	0.072	4.713	3.122	1.419	0.192	
	0.310	0.271	0.366	0.158	5.535	2.604	3.885	1.185	
	0.372	0.280	0.226	0.141	6.308	2.924	4.607	1.456	
	0.398	0.280	0.297	0.270	6.168	2.864	5.174	1.942	
	0.459	0.330	0.415	0.254	7.122	3.114	6.031	2.132	



## Chapter 4

# Detection of Micro-calcifications

Now we get back to our original problem, detection of micro-calcification. As we already discussed that the identification of breast cancer needs two steps, detection of micro-calcification and classification of the calcification into one of the three class, normal, benign or malign.

Here we start with one of the approaches from the literature to detect micro-calcification. We first analyze the fuzzy logic based micro-calcification scheme by Heng-Da Cheng [2] and then we propose some modification of this method to get more accurate result with less computational complexity.

The rest of this chapter is organized as follows: The next section presents the steps used for micro-calcification detection by Heng-Da Cheng. In the following section we suggest to use S-function instead of  $\Pi$ -function for the fuzzification of image for enhancement of mammogram. We also propose an easy way of thresholding for the segmentation of mammogram to detect the micro-calcification.

### 4.1 Detection of Micro-Calcification [2]

In [2] authors used fuzzy logic techniques for micro-calcification detection. Micro-calcifications are first enhanced based on brightness and non-uniformity. Then, irrelevant breast structures are excluded by a curve detector. Finally microcalcifications are located using an iterative threshold selection method. The major advantage of this method, as claimed by author, is its ability to detect micro-calcifications even in very dense breast mammograms.

The proposed method follows five major steps: image fuzzification, image enhancement, irrelevant structure removal, segmentation, and image recon-

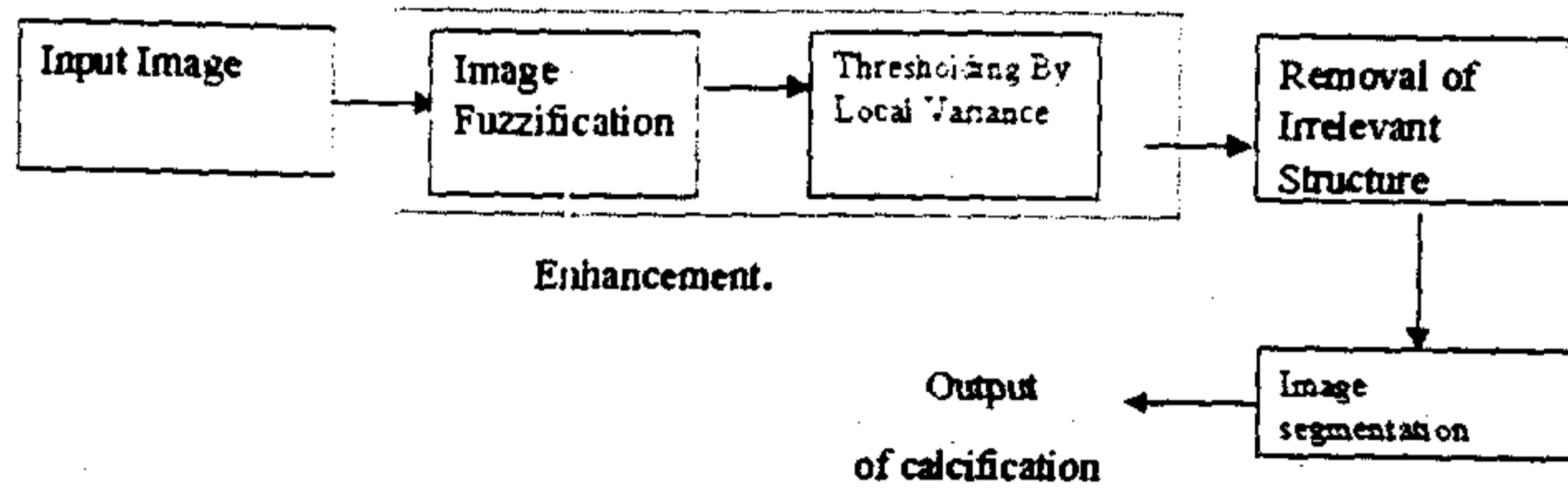


Figure 4.1: The system diagram of the proposed algorithm.

struction. These steps are shown in the Fig 4.1

#### Common assumptions usually made in mammogram analysis

Here are few assumptions that are considered valid for mammogram analysis.  
 (I) In digital mammography the suspicious areas are brighter than their surrounding tissues.

(II) The intensities of microcalcifications are higher than the average intensity of the breast tissues.

(III) The regions of microcalcifications are usually not homogeneous.

#### 4.1.1 Mammogram Fuzzification

This technique employs fuzzy sets theory to increase the contrast of microcalcifications. The mammogram is fuzzified by using a  $\pi$ -function shown in Fig 4.2. The value of the  $\pi$  represents the degree of the closeness of the gray value  $g$  to  $c$ . The  $\pi$ -function equation is defined as:

$$S(g; x, y, z) = \begin{cases} 0 & g \leq x \\ 2\left(\frac{g-x}{z-x}\right)^2 & x \leq g \leq y \\ 1 - 2\left(\frac{g-x}{z-x}\right)^2 & y \leq g \leq z \\ 1 & g \leq z \end{cases}$$

$$\pi(g; b, c) = \begin{cases} S(g; c - b, c - b/2, c) & \text{if } g \leq c \\ 1 - S(g; c, c + b/2, c + b) & \text{otherwise} \end{cases}$$

The selection of the points  $c$  and  $b$  could be viewed as an object-background classification problem. The point  $c$  is selected using an entropy thresholding method.

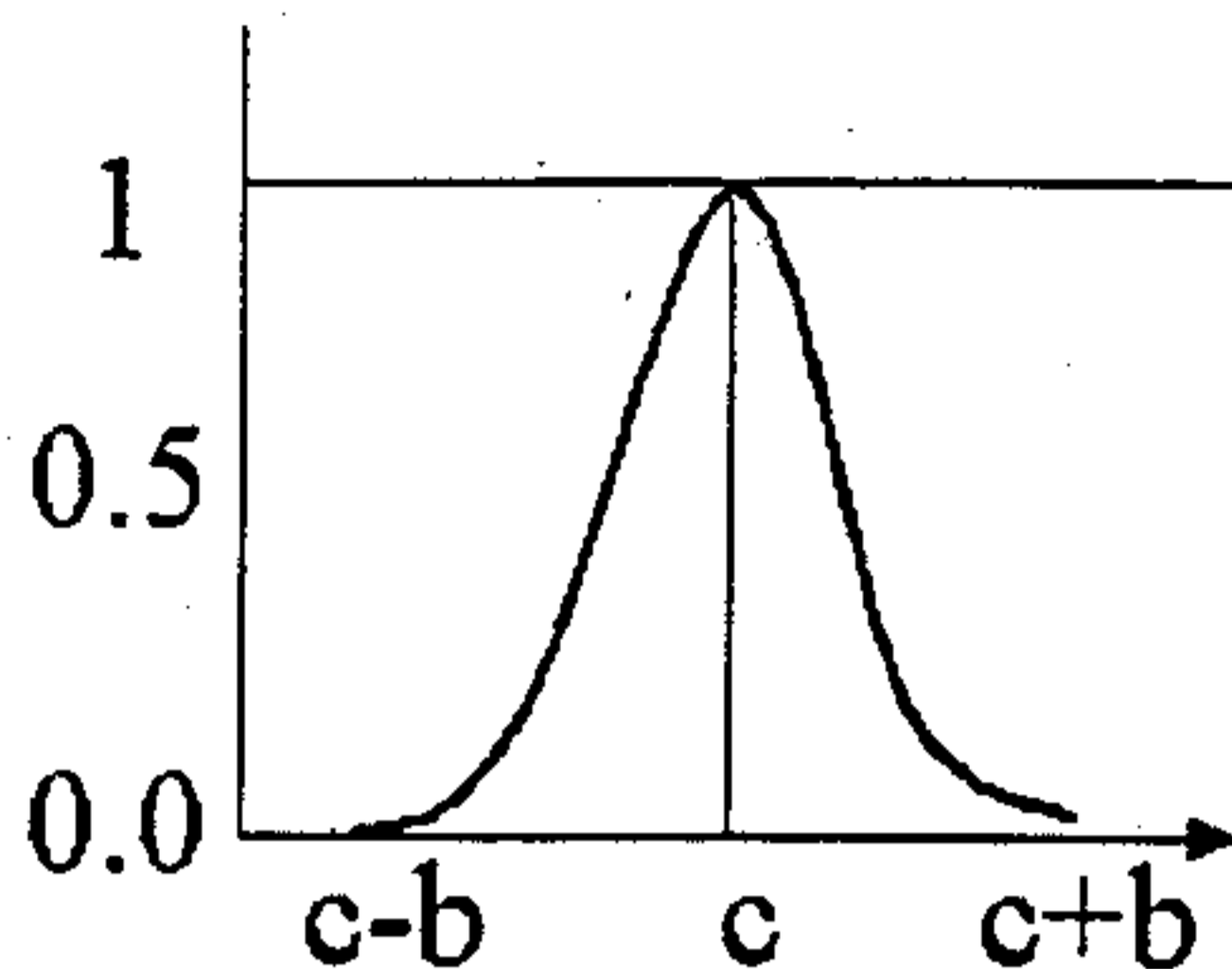


Figure 4.2: A  $\pi$  function for image fuzzification.

$$c = \underset{t}{\text{Argmax}}(H_b(t) + H_o(t)) \text{ for } k \leq t \leq N \quad (4.1)$$

Where

$H_b(t)$  and  $H_o(t)$  are entropies of object and Background regions.

$t$  is the value of the threshold.

$b = \max(c-k, N-c)$ .

$k$  is the mean gray value of breast tissue.

$N$  is the maximum gray value of the image.

The areas containing microcalcifications (i.e., ROI) are usually inhomogeneous and the variances of these areas would be larger than those of tissue background regions. Therefore, a threshold can be used to separate the microcalcifications from the breast tissues according to nonuniformity. This threshold is computed from the local variance occurrence function equations (4.2) and (4.3) and is determined by a minimum error thresholding technique. Here the objective is to find out the threshold of the local variance to classify the micro-calcification and background tissues. The optimum threshold  $T$  is found corresponding to the minimum value of the minimum-error thresholding criterion function.

$$\mu_{xy} = \frac{1}{M \times M} \sum_{j=1}^{M^2} g_{xy} \quad (4.2)$$

$$\sigma^2 = \frac{1}{M \times M} \sum_{j=1}^{M^2} (g_{xy} - \mu_{xy})^2 \quad (4.3)$$

$$\nu_{xy} = \begin{cases} \frac{\sigma_{xy}^2}{T} & \text{if } \sigma_{xy}^2 \leq T \\ 1 & \text{otherwise} \end{cases}$$

### 4.1.2 Image Enhancement.

The image after enhancement is obtained by:

$$g'_{xy} = \pi(g; b, c) \times \nu_{xy} \times N, \quad (4.4)$$

where N is the maximum gray value of the image.

### 4.1.3 Irrelevant Breast Structure Removal

Most of the irrelevant breast structures are found to exhibit line-like or curve-like patterns after the microcalcification enhancement. A curve detector is, therefore, employed to remove these irrelevant breast structures. Whether a pixel is considered a microcalcification or a curve-like pattern is determined by the ratio of its tracing length and width. The curve detector traces the length and width of the detected micro-calcification in the enhanced image to decide whether pixels are in a curve or not.

The traced object is said to be a curve if one of the following conditions holds.

$$\frac{\text{Length}}{\text{Width}} > T_1. \quad (4.5)$$

$$\text{Length} > T_2. \quad (4.6)$$

If the above conditions are not satisfied it will be considered a calcification.

## 4.2 Proposed Modifications

### 4.2.1 Use of S-function instead of $\Pi$ -function for image fuzzification

When we are using the  $\pi$ -function for the image fuzzification, it can be easily observed that the gray values of the pixel greater than c is reduced from its actual value. Similarly, the gray values of the pixels that are less than c will also be reduced. So it will not intensify the contrast in a desirable manner. But, the pixels with gray value c will become prominent. So, if c is the intensity level for micro-calcification then this will work. This is shown



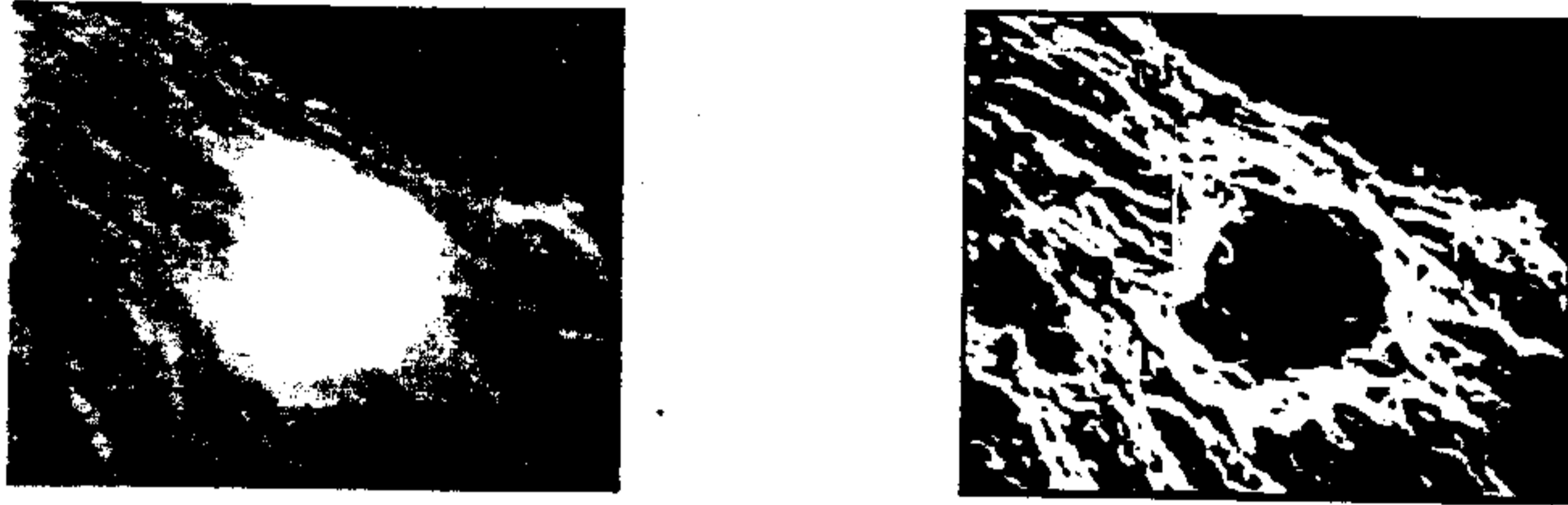


Figure 4.3: Result of  $\pi$  function image Enhancement.

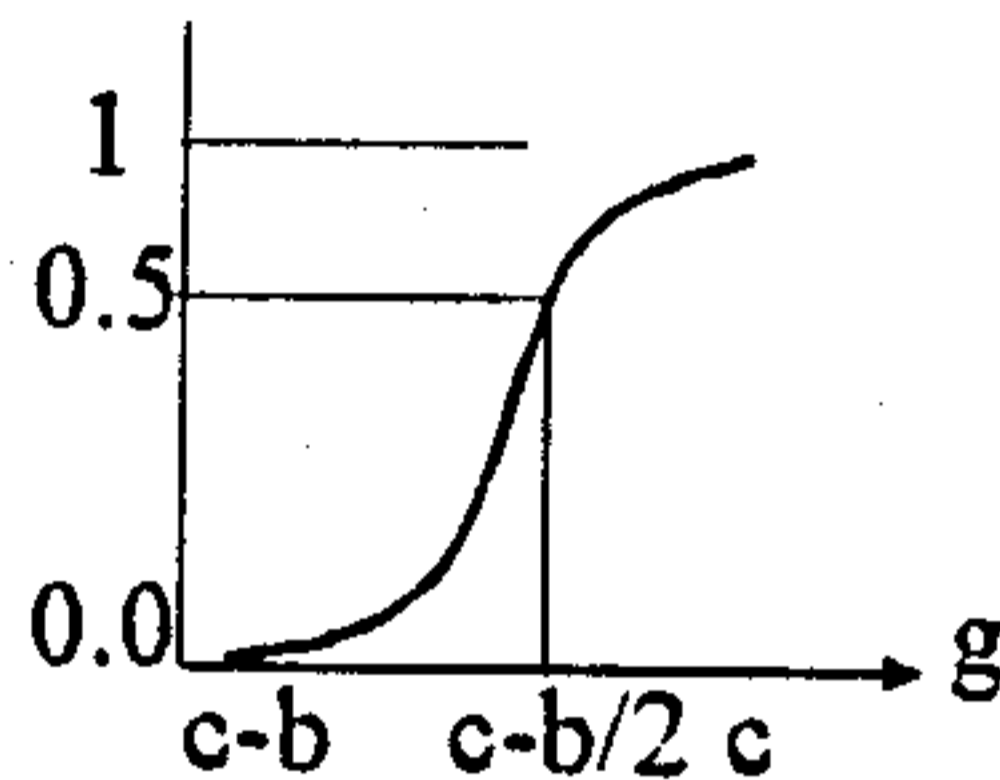


Figure 4.4: A S-function for image fuzzification.

in the Fig 4.2. Since usually the intensities of the microcalcification is higher than the intensity of the breast tissue and it is brighter towards the center of the calcification. the  $\pi$ - function often creates a black patch inside the calcification during the enhancement of the mammogram. Fig 4.3 shows an image before and after the enhancement of the calcification using the method in [2].

This problem can be solved by using a S-function instead of  $\pi$ - function. By using a S-function for image enhancement the pixel value above  $c$  ( see Fig 4.4 ) will remain as it is. This is because the S-function gives a value of 1 after the point  $c$ . The comparison of the result between  $\pi$  and S-function for image enhancement is shown in the Fig 4.5. Note that,  $\pi$ -function creates a black patch inside the classified area which is absent with S-function.

#### 4.2.2 Setting of Threshold on Local variance for Non-uniformity of calcification

In [2] we have to compute two threshold values. one is  $c$ (cross over point for fuzzification) and  $T$ (optimum threshold for non uniformity of calcification).



**Calcification with  $\pi$ -function and S-function**

**Figure 4.5: Comparison of  $\pi$  and S-function enhancement.**

As discussed earlier, Cheng proposed two methods for these. The value of  $c$  is found maximizing entropy of the segmented image. But, histogram entropy thresholding tries to find a threshold that often makes the object and background class of comparable sizes. So it may not produce the desired threshold.

It is observed that the enhancement algorithm gives higher values to the pixels with gray values nearer to  $c$  when their local variance is greater than  $T$ . And it assigns gray value almost zero to the pixel with gray value outside the range  $c-b$  to  $c+b$ , where  $b = \max\{c-k, N-c\}$ . Clearly  $N-c$  dominates the other one. With a considerable amount of experiments we found that a threshold of 150 is good enough for the background gray value for a pixel to be calcification. Since, for a given application mammograms are taken with the same machine, such a threshold can be used with a high confidence.

Again we found that the threshold for local variance  $T$ , is always between 10-15. This is found to be very low for local variance of the calcification pixel. Hence in the image enhancement equation (4.4), this  $T$  has very little contribution. So based on our extensive experiments we have fixed a threshold value of 25 for the local variance of a pixel to be calcification.

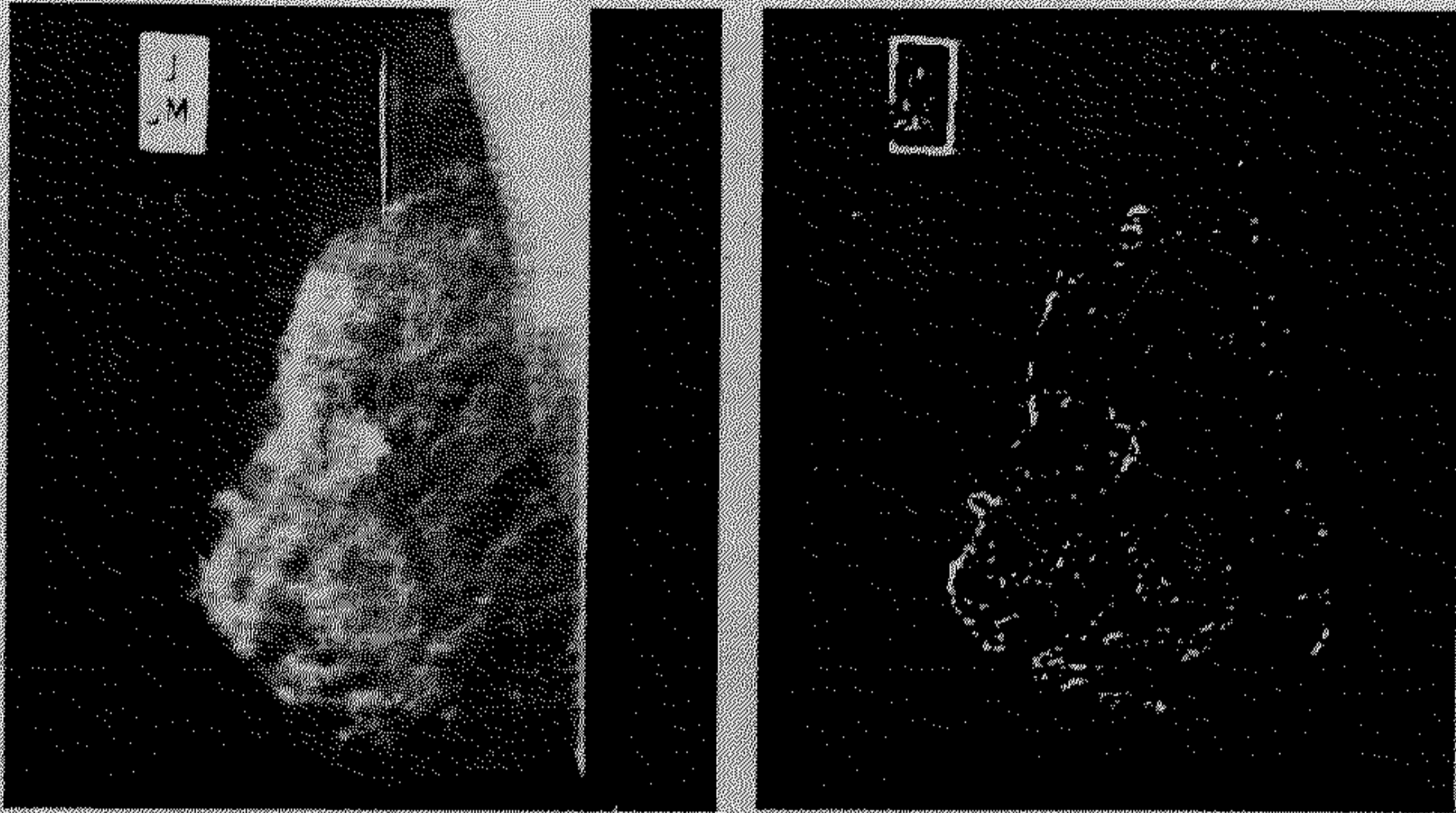
We have seen that these above two cutoff values do not degrade the performance of the calcification detection. But using these thresholds we achieved a good amount of decrease in computational complexity.

### 4.3 Calcification Detection method

We have followed the following method for the detection of calcification.

- Threshold the image with local variance greater than 25 and less than 900





Mdb219 (Benign)

Variance (25-900) and Background > 150

Figure 4.6: Thresholding with background and variance.



removal of irrelevant structure.

Removal of two pixel and Isolated Component

Figure 4.7: Removal of line structure and isolated component.



and background pixel value greater than 150.

- Remove the irrelevant curve like structure.
- Remove the unnecessary isolated components.
- Remove all the components from the outside of the breast tissue.
- Find out the calcification clusters and predict the circle around it.

## 4.4 Results and Discussion

For this experiment we considered 30 mammograms from the MIAS database: 10 normal cases, 8 Benign cases and 12 malign cases. These 30 images generate 178 suspected calcified areas. Out of these 10 normal images, for 3 images, the proposed system detected (false positive) calcified areas. Out of the 20 images with calcification in 6 images it could not detect the actual calcified areas. Note that, the false positive cases may be correctly classified by the classification stage.

Using the method in [2], the number of false positive cases increases a lot, but it reduces the number of false negative cases to four.

Fig 4.6 and 4.7 depict the results with one of the difficult images.



## Chapter 5

# Classification of Microcalcifications.

We have already discussed that, we need to classify the detected calcification into one of the three classes: normal, benign or malignant. According to the class to which the calcification belongs, we can predict the presence or absence of the carcinoma in a mammogram.

The major problem to classify the calcification is mainly due to its fuzzy nature, low contrast, and low distinguishability from their surroundings. In this chapter we discuss about the classification of the detected microcalcifications. In the first section, we deal with feature selection for calcifications using the neural network based feature selection method discussed in Chapter 2 and use the selected features to classify micro-calcification. And in the following section we discuss about fuzzy rule extraction for classification of micro-calcification.

### 5.1 Classification of Micro-Calcification using OFS

Here we considered ten features. These are size of the calcification, cluster size, contrast, mean gray value of the calcification, compactness[9], linearity, irregularity, homogeneity, moments[9], distance of the calcification from the center of the calcification cluster. We applied the neural network based feature selection technique with regularization to select the major features for classification of micro-calcifications. In OFS technique the features are selected based on their importance as reflected by the values of the attenuators associated with the input nodes. Table 5.1 shows the gate opening for three different initializations of the network. From Table 5.1 we find that

the original OFS method opens most of the gates to achieve a low sum of square error and low mis-classification. By applying a regularization factor 1.5 we found that the network emphasizes on five features in all the runs. Hence we considered these features to be useful features for classification of micro-calcification. These features are cluster size, contrast, mean gray value of the calcification, linearity, and distance of the calcification from the center of the calcification cluster.

The network is trained for 30000 iterations, with the parameters  $\eta = 0.9$  and  $\mu = 0.9$ . *Note that for the second and third initializations we get a better classification performance indicating that OFS with regularization does a good job of feature selection.*

### Classification of Calcifications using Neural Networks.

In the 30 images, there are 178 suspected calcification areas. Of these 178 cases, we use 110 for training, 30 for validation (of neural network) and 38 for testing. We use only the five selected features and train 3 networks each with 5 nodes in the hidden layer. Table 5.2 shows the classification rate on the training and test data set. Table 5.2 reveals that the selected features result in an excellent performance.

## 5.2 Fuzzy rules for Micro-Calcification Classification

The result of fuzzy rules on the training data is shown the Table 5.3. During the tuning process the sum of squared error(SSE) and misclassification are reduced with the increase in the spread. By this method we find that the mis-classifications reduces to 9. Hence the classification rate is 93.5%. With regularizing constant of 0.5, the reduction in the spread is not much here. It could be because of the data are sparsely distributed as reflected by the high initial spreads for 3 of the features. Table 5.4 displays the performance of the training and test data sets. Comparing Table 5.4 with Table 5.2 we find that neural networks marginally perform better than fuzzy rules. It is observed that with the greater value of regularization factor  $K$ , the increase in spread is controlled more. Hence the specificity of the fuzzy rules increases by applying the regularization to the tuning process. The misclassification can further be improved by optimizing the number of rules that we could not investigate due to lack of time.

The fuzzy rules have two distinct advantages: it is readable by human beings and it is not likely to give poor generalization.

The comparatively poor performance by the fuzzy rules may be attributed to the fact that the feature set that we used may not be the best for fuzzy rules. The feature set is selected by an online feature selection scheme using neural networks, so it may not be the best set for a fuzzy rule base classifier. Because the utility of a feature depends also on the tool used to solve the problem.

Table 5.1: Results of Mammogram data set (10 features) with and without Regularization :  $\eta = 0.9$  and  $\mu = 0.9$ .

$K_r$	Random initialization of network					
	Initialization 1		Initialization 2		Initialization 3	
	Gate opening	SSE, mis-classification	Gate opening	SSE, mis-classification	Gate opening	SSE, mis-classification
0.0	0.024	0.0084, 2	0.957	2.008, 2	0.945	2.011, 1
	0.999		1.000		0.999	
	0.994		0.996		0.998	
	0.919		0.983		0.966	
	0.996		0.991		0.990	
	0.939		0.697		0.703	
	0.433		0.022		0.686	
	0.994		0.905		0.013	
	0.891		0.869		0.799	
	0.999		0.999		0.999	
1.5	0.001	6.272, 5	0.064	0.596, 0	0.001	0.478942, 0
	0.514		0.357		0.435	
	0.275		0.222		0.252	
	0.262		0.155		0.196	
	0.090		0.073		0.001	
	0.141		0.117		0.099	
	0.001		0.000		0.026	
	0.001		0.076		0.000	
	0.066		0.091		0.074	
	0.253		0.197		0.300	

Table 5.2: Result of classification of calcifications by neural networks using 5 selected features :  $\eta = 0.8$

Network Initialization	Classification rate	
	Training	Testing
1	100%[110/110]	92.1%[35/38]
2	96.3%[106/110]	89.4%[34/38]
3	96.3%[106/110]	92.1%[35/38]



Table 5.3: Result on Mammogram Training set with and without regularization  $\eta_s = 0.3$  and  $\eta_m = 0.2$

	Fuzzy sets $\sigma$ and $\mu$										SSE, mis- classification
	spread( $\sigma$ )					prototype ( $\nu$ )					
Initial Rules	2.32	6.82	16.93	0.46	6.95	3.68	2.21	171.41	0.90	5.79	78.32, 27
	3.17	26.52	17.34	0.83	3.97	3.50	10.30	197.75	0.49	10.10	
	4.24	10.22	11.04	0.86	9.03	7.25	14.48	192.23	0.55	13.46	
	4.40	4.31	9.86	0.70	6.84	6.50	12.51	171.74	0.49	9.77	
	10.68	9.61	30.15	0.73	28.99	12.30	12.54	188.61	0.63	26.71	
	0.01	7.14	16.50	0.60	0.01	42.00	14.31	180.97	0.55	90.86	
Tuned Rules	5.17	6.75	18.66	1.19	8.86	4.05	1.28	171.51	0.96	5.22	51.94, 9
	2.98	26.80	17.24	2.24	2.90	3.09	9.89	197.83	0.78	11.70	
	2.84	12.22	14.15	3.63	10.81	8.12	14.67	191.94	0.60	14.79	
	4.14	7.35	12.14	2.77	7.48	7.35	12.58	171.96	0.42	10.59	
	9.44	11.07	32.04	2.60	30.31	15.90	12.01	188.43	0.70	28.37	
	0.01	11.35	19.77	2.87	0.01	42.00	14.51	181.02	0.56	90.86	
$k_r = 0.5$ Rules	5.04	6.61	18.60	1.14	8.74	4.01	1.25	171.50	0.96	5.20	51.44, 10
	2.92	26.67	17.16	2.22	2.81	3.07	9.91	197.83	0.78	11.70	
	2.81	12.20	14.02	3.59	10.76	8.12	14.65	191.95	0.60	14.77	
	3.99	7.16	11.98	2.69	7.36	7.32	12.60	172.02	0.42	10.54	
	10.05	11.06	25.26	2.56	24.90	15.81	11.99	188.32	0.69	28.49	
	0.01	11.27	19.69	2.84	0.01	42.00	14.51	181.01	0.56	90.86	

Table 5.4: Result of Classification of calcifications by fuzzy rules with 5 selected features :  $\eta = 0.8$

Rule base	Classification rate	
	Training	Testing
Initial Rules	80.7%[113/140]	76.3%[29/38]
Tuned Rules	93.5%[131/140]	86.8%[33/38]
Regularized Rules	92.8%[130/140]	86.8%[33/38]

# Chapter 6

## Conclusion

Automated breast cancer detection has been studied for more than 20 years. Although by now some progress has been achieved, there are still remaining challenges for future research.

In this thesis we first proposed two general purpose methods: feature selection using neural networks and extraction of fuzzy rules with higher specificity. The proposed modification of the neural network based feature selection method is validated with three sets of synthetic data and iris data. We found, by applying regularization to the OFS that can avoid the selection of features with low discriminating power and correlated features. The proposed method to extract fuzzy rules with higher specificity for classification is validated with two sets of synthetic data and iris data. It is observed that by applying penalty for higher spreads of membership functions the increase of spreads is controlled.

We then critically analyzed an existing method for detection of micro-calcification, found some of its problems and suggested some easy solutions. Finally, we classified the detected calcifications using neural networks and fuzzy rules. The neural network method selects only 5 features from a set of 10 features which are found to produce better results than that by all the features. Our fuzzy rule extraction scheme finds rules with higher specificity that are interpretable by human beings.

To demonstrate the effectiveness of the system we need to do more experiments. We did not try to optimize the number of rules, which should be done to improve the performance of the system further. We considered only 10 features for classification of micro-calcification and hence there is enough opportunities to improve the performance of the system by expanding the feature set.

# Bibliography

- [1] H.D. Cheng, X Cai, X Chen, L Hu, X Lou, Computer-aided detection and classification of microcalcifications in mammograms: a survey, *Pattern Recognition* 36(2003)2967-2991.
- [2] H.D. Cheng, Y.M. Lui, R.I. Freimanis, A novel approach to microcalcification detection using fuzzy logic technique, *IEEE Trans. Med. Imag.* 17 (3) (1998) 442-450.
- [3] B. Verma, J. Zakos. A computer-aided diagnosis system for digital mammograms based on fuzzy-neural and feature extraction techniques, *IEEE Trans. Inform. Technol. Biomed.* 5 (1) (2001) 465-4.
- [4] Pal NR, Chintalapudi KK. A connectionist system for feature selection. *Neural, Parallel and Scientific Computations* 5(1997) 359-382
- [5] K. Woods, L.P. Clarke, R. Velthuisen, Enhancement of digital mammograms using a local thresholding technique. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society* 13 (1) (1991) 114-115.
- [6] H.D. Cheng, H.J. Xu, Fuzzy approach to contrast enhancement, 14th International Conference on Pattern Recognition, Brisbane, Australia, August 1998, pp. 1722.
- [7] N. R. Pal, A. Laha, J. Das. Designing fuzzy rule based classifier using self-organizing feature map for analysis of multispectral satellite images. *International Journal of Remote Sensing*, 2005.
- [8] E. Anderson. "The Irises of the Gaspé peninsula", *Bulletin of the American IRIS Society*, vol 59, pp 2-5. 1935.
- [9] L. Shen, M Rangaraj, J. E. Leo Dasautels Application of shape analysis to mammographic Calcifications. *IEEE transaction on medical imaging*, vol 13, No. 2, June 1994.

- [10] L. Mascio, M. Hernandez, and L. Clinton. Automated analysis for microcalcifications in high resolution mammograms. Proc. SPIE Int. Soc. Opt. Eng., vol. 1898, pp. 472479, 1993.
- [11] Y. Chitre, A. Dhawan, and M. Moskowtz. Artificial neural network based classification of mammographic microcalcifications using image structure features. in State of the Art of Digital Mammographic Image Analysis. Singapore: World Scientific, 1994, vol. 7, pp. 167197.
- [12] H. Yoshida, R. Nishikawa, K. Muto, K. Doi, and M. Tsuda, Application of the wavelet transform to automated detection of clustered microcalcifications in digital mammograms, Tokyo Inst. Polytech., Tokyo, Japan, Academic Rep., vol. 16, 1994.
- [13] D.H. Davies, D.R. Dance, C.H. Jones. Automatic detection of clusters of calcifications in digital mammograms, SPIE Med. Imag. IV: Image Process. 1233 (1990) 185191.