Feature extraction and classifications of mammograms

4.1 Introduction

The next stage after segmentation is the feature extraction. This stage is separate for masses as well as for micro-calcifications. Feature can be categorized into three groups such as texture, shape and scalar area features. Micro-calcifications are characterized by their shape, number, contrast and distribution.

4.2 Feature extraction of masses

In this study shape based and texture based features are extracted.

4.2.1 Shape Based Features

Several techniques have been developed for charactering the shape of masses in an image. Many features are in vogue in the literature to classify the masses into benign and malignant. The important shape based features includes the geometric parameters such as speculation, relative contrast, circulatory, compactness and radial angle are used for classifications [74, 80].

A. Spiculation

Firstly, the centroid is calculated. Euclidean distances are found from the centroid to every 20th pixel on the boundary. Now these radial distances are scanned sequentially and if the difference between the consecutive distances is greater than 5 then the spiculation count is incremented by 1. In this way spiculation count is calculated.
B. Relative contrast

Normally, the intensities of malignant masses are higher than those of benign masses. Therefore, the contrast of masses with respect to the background in malignancy is higher than in benignancy. In our approach, the contrast is computed as the difference between the gray level average of masses and that of the tissue within a defined ring-like operator. As shown in Fig. 4.1(b), the ring-like operator indicates the circle band around the region of masses and can be represented as

\[ \Omega = \left\{(x, y) \mid d \leq \sqrt{x^2 + y^2} \leq d + \Delta d \right\} \]  

(4.1)

Here \( d \) is the radius of the smallest circle containing the mass, \( \Delta d \) is the width of the ring-like operator. In the experiment, \( d \) is set as 5. The contrast is computed as:

\[ \text{contrast} = \frac{\text{mean}(R) - \text{mean}(\Omega)}{\text{mean}(R)} \]  

(4.2)

where \( R \) is the region of masses.

C. Circularity

The main purpose of circularity is to show the circular degree of masses. The higher the circularity, the more circular the object tends to be. Owing to the fact that the roundness is one of the criteria of benign masses, the probability of masses as being benign is higher when circularity is higher. Let \( R \) be defined as the region of masses. In the computation of mass circularity, a circle \( C_{eq} \) which has the same area as the mass is first located at the center of the mass, as shown in Fig. 4.1(c). The equations for calculating the circularity are:

\[ S = \sum_{(x,y) \in R}^{} 1 \]  

(4.3)
Where \( S \) is the area of the mass and the center of the mass, \((x,y)\) is given by

\[
\bar{x} = \frac{1}{S} \sum_{(x,y) \in R} x \tag{4.4}
\]
\[
\bar{y} = \frac{1}{S} \sum_{(x,y) \in R} y \tag{4.5}
\]

The radius of circle \( R_{eq} \) is represented as

\[
r_{eq} = \frac{\text{area} (C_{s})}{\pi} = \frac{\text{area} (R)}{\pi} = \sqrt{\frac{S}{\pi}} \tag{4.6}
\]

Then circularity can be calculated as:

\[
circularity = \frac{\text{area} (R \cap C_{s})}{\text{area} (R)} \tag{4.7}
\]

The value of circularity will lie between 0 and 1. If the ratio is 1, this means that the mass matches exactly with the \( C_{eq} \), implying the mass is circular. On the other hand, if the circularity is much smaller than 1, this implies the mass is far from being circular.

**D. Radial angle**

The radial angle is used to differentiate the shape of edges of the masses as spiculated or as round and smooth. The radial angle is the smaller included angle between the direction of the gradient and the radial direction of the edge, as shown in Fig. 4.1(d), where the radial angle of \( P1 \) is \( \theta \). As we know, when the mass tends to be more rounds, its radial angles tend to be near 180 and the average of the radial angles tends to be larger. Conversely, a mass with spiculated edge will have a smaller averaged radial angle. The distribution of the circular shape is aggregated around the largest value, 180, while that of a spiculated shape is evenly distributed into much smaller values. Therefore, the value of the averaged radial angles provides one indicator for differentiating between benign and malignant masses.

**E. Compactness**

It is the measure of closeness its shape is to that of a circle. It is calculated by the perimeter and area of the segmented portion. The measure of compactness is given by:

\[
\text{Compactness} = \frac{\left( \frac{\text{perimeter}}{\pi \text{ (area)}} \right)^2}{4} \tag{4.8}
\]

Compactness measure of the circle is 1.
4.2.2 Texture Based Features

Texture features have been proven to be useful in differentiating masses and normal breast tissues. These features can be sub-divided into three categories such as statistical, structural and gray level dependency. Gray level dependency is related to the spatial gray level dependence matrix. The co-occurrence based features are extracted for classification because these consider the spatial arrangements of pixels in to consideration.

A. Gray Level Co-occurrence Matrices (GLCM)

The GLCM is a well-known established robust statistical tool for extracting second order texture information from images. The GLCM matrix element, \( P_{\theta, d(i, j)} \), is the joint probability density of the occurrence of grey levels \( i \) and \( j \) for two pixels with a defined spatial relationship on an image. The spatial relationship of the pixel pair is described by a selection rule that specifies the relative direction in degrees and the distance \( d \) between the two pixels at direction \( \theta \). Because of the discrete nature of the digital image, the choice of \( \theta \) is actually limited to \( 0^\circ, 45^\circ, 90^\circ, \) and \( 135^\circ \) and the distance \( d \) is limited to integral multiples of the pixel size [81-84].

B. Haralick's Features

The Haralick's texture features are used for image classification. These features capture information about the patterns that emerge in patterns of texture. A number of texture features can be derived from the GLCM. This study, evaluated the discriminant ability of eight features: correlation, entropy, energy, inertia, inverse difference moment, sum average, sum entropy and difference entropy [84]. These features describe the shape of the GLCM and generally contain information about the image characteristics such as homogeneity, contrast and the presence of organized structures as well as the complexity and grey level transitions within the image [81-84].

The texture descriptors derived from GLCM are contrast, energy, homogeneity and correlation of gray level values. Table 4.1 provides the equations for the four features. The contrast measures the amount of local variations present in an image, while energy is the sum of squared elements in GLCM. Energy may also be referred as uniformity or the angular second moment. The homogeneity descriptor refers to the closeness of the distribution of elements in...
GLCM to the GLCM diagonal. Lastly, correlation will show how correlated a pixel is to its neighbor over the whole image [81-84].

Where $P_{ij} = \text{Element } i, j$ of the normalized symmetrical GLCM and $N = \text{Number of grey levels in the image as specified by number of levels in under quantization on the GLCM.}$

$\mu = \text{The GLCM mean, given as: } \mu = \sum_{i,j=1}^{N} i P_{ij}$

$\sigma^2 = \text{The variance of the intensities of all reference pixels in the relationship that contributed to the GLCM, given as: } \sigma^2 = \sum_{i,j=1}^{N} P_{ij} (i - \mu)^2$

<table>
<thead>
<tr>
<th>Feature</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contrast</td>
<td>$\sum_{i,j=1}^{N} P_{ij} (i - j)^2$</td>
</tr>
<tr>
<td>Energy</td>
<td>$\sum_{i,j=1}^{N} P_{ij} i^2$</td>
</tr>
<tr>
<td>Homogeneity</td>
<td>$\sum_{i,j=1}^{N} P_{ij} 1 - (i - j)^2$</td>
</tr>
<tr>
<td>Correlation</td>
<td>$\sum_{i,j=1}^{N} P_{ij} (i - \mu)(j - \mu) / \sigma^2$</td>
</tr>
</tbody>
</table>

Table 4.1: Features of GLCM.

The GLCM depends on the spatial relationship of the pixel pairs. Because it is expected that the texture of a mass may be isotropic whereas the texture of normal breast structures may have a slightly stronger directional dependence as they diverge from the nipple, we first investigate the dependence of the classification accuracy on directional information. For a fixed pixel distance $d$, each of the eight texture features are calculated from the GLCM at $\theta = 0^\circ, 45^\circ, 90^\circ$ and $135^\circ$, resulting in a 32-dimensional feature space. One of the eight Haralick’s features is entropy. This approach replaces the Shannon’s entropy with new generalized entropy function. The proposed entropy function, has an exponential gain function given by

$$H = \sum_{k \in \mathcal{P}} p_k e^{-(a p_k + b p_k^2 + c p_k^3) + d}$$  \hspace{1cm} (4.9)$$

where $\sum_{k \in \mathcal{P}} p_k = 1$

Since this entropy function is parameterized, we used few parameter $(a, b, c, d)$ sets to evaluate the entropy value. Since this entropy function is parameterized requiring a few parameter sets to evaluate the entropy value. Now the feature space increases to 48-dimensions.
The features are computed by evaluating the co-occurrence matrix for various distances ranging from 2 pixels to 20 pixels with an interval of 2 pixels. It is found that the pixel distance of 12 pixels gave high value of classification accuracy as shown in Fig. 4.2. Use of new entropy function improved the classification accuracy by about 5%.

![Dependence of classification accuracy on distance](image)

C. Laws’ texture energy measures

Laws’ texture energy measures determine texture properties by assessing average gray level, edge, spots, ripples and waves in texture. The textures are extracted based on the method proposed by Laws’ [85]. According to this approach, texture features are extracted from images that had previously been filtered by each of the 25 Laws’ masks. The filtered image is characterized as texture energy images. The first order statistics such as mean, standard deviation, range, skewness and kurtosis are computed for each ROI by applying masks and finding the average value of the filtered image in the central region [85].

4.3. Feature extraction of micro-calcification clusters

The micro-calcification clusters are classified using the important features such as the number of micro-calcifications, total area of the cluster, average area of each calcification, standard deviation of the area, mean of the compactness, mean of eccentricity and mean of the product of area and eccentricity [74, 80].
The number of Micro-calcifications

In the segmented image, the number of white regions is counted and this count is taken as one feature. More the count more likely that cluster is malignant.

The total area of cluster

The sum of the areas of all the regions is calculated and taken as one feature.

The average area of each calcification

The mean of the areas of the micro-calcifications is obtained and considered as the third feature. This is calculated by dividing the first feature by the second one.

The standard deviation of the area

The standard deviation of the areas of all the regions is taken as another feature.

Mean of compactness

Compactness is defined as

\[
Compactness \ s = \left( \frac{\text{perimeter}}{\text{area}} \right)^2
\]

This measure is calculated for every region and the mean of this measure is considered as the feature.

Mean of Eccentricity

Eccentricity is the measure describing the shape of the conic section. It can be thought of as a measure of how much the conic section deviates from being circular. This measure is calculated for every region and mean is taken as the feature.

Mean of the product of area and eccentricity

This measure is the mean of the eccentricity weighted by its area. The value of eccentricity of larger objects dominates the overall eccentricity measure. This is calculated by the formula

\[
EA = \frac{1}{N} \sum_{i=1}^{K} E(i) \times A(i)
\]

4.4 Feature Selection

A rough set [86] based feature selection scheme is used prior to classification. All the features may not be required for classification. Use of lesser number of features may improve the
classification accuracy. Rough set based reduct generation algorithm chooses the consistency of the data as the criteria for finding the reduct. The smallest set of attributes which gives highly consistent and having lesser number of conflicting rows is the reduct set. Application of reduct requires discretization. We chose entropy based discretization because it finds the cuts at the points where uncertainty is large. The selected features are then used for classification.

Rough set theory is an extension of conventional set theory that supports approximations in decision making. The rough set itself is the approximation of a vague concept by a pair of precise concepts, called lower and upper approximations, which are a classification of the domain of interest into disjoint categories. The lower approximation is a description of the domain objects which are known with certainty to belong to the subset of interest, whereas the upper approximation is a description of the objects which possibly belong to the subset [86-87].

Let \( I = (U, A) \) be an information system, where \( U \) is a non-empty set of finite objects and \( A \) is a non-empty finite set of attributes such that \( a: U \to V_a \) for every \( a \in A \).

With any \( P \subseteq A \) there is an associated equivalence relation \( IND(P) \):

\[
IND(P) = \{(x, y) \in U^2 \mid \forall a \in P \ a(x) = a(y)\}
\]

(4.12)

The partition of \( U \), generated by \( IND(P) \) is denoted \( U/P \) and can be calculated as follows:

\[
U/P = \bigotimes\{a \in P : U/IND([a])\}
\]

(4.13)

Where \( \trianglelefteq B = \{X \cap Y \mid X \in A, Y \in B, X \cap Y \neq \emptyset\} \)

If \((x, y) \in IND(P)\), then \( x \) and \( y \) are indiscernible by attributes from \( P \). The equivalence classes of the \( P \)-indiscernibility relation are denoted \([x]_P\).

Let \( X \subseteq U \), the \( P \)-lower approximation \( P_X \) and upper approximation \( \overline{P}_X \) of set \( X \) can now be defined as:

\[
P_X = \{x \mid [x]_P \subseteq X\}
\]

(4.14)

\[
\overline{P}_X = \{x \mid [x]_P \cap X \neq \emptyset\}
\]

(4.15)

It is such a tuple \(<P_X, \overline{P}_X>\) that is called a rough set. Consider the approximation of concept \( X \) in Fig.4.3.
Let $P$ and $Q$ be equivalence relations over $U$, then the positive, negative and boundary regions can be defined as:

$$\text{POS}_P (Q) = \bigcup_{x \in U/Q} PX$$  \hspace{1cm} \text{(4.16)}

$$\text{NEG}_P (Q) = U - \bigcup_{x \in U/Q} \overline{PX}$$ \hspace{1cm} \text{(4.17)}

$$\text{BND}_P (Q) = \bigcup_{x \in U/Q} \overline{PX} - \bigcup_{x \in U/Q} PX$$ \hspace{1cm} \text{(4.18)}

The positive region contains all objects of $U$ that can be classified to classes of $U/Q$ using the knowledge in attributes $P$.

An important issue in rough set based feature selection is discovering dependencies between attributes. Intuitively, a set of attributes $Q$ depends totally on a set of attributes $P$, denoted $P \Rightarrow Q$ if all attribute values from $Q$ are uniquely determined by values of attributes from $P$. If there exists a functional dependency between values of $Q$ and $P$, then $Q$ depends totally on $P$. Dependency can be defined in the following way:

For $P, Q \subseteq A$, it is said that $Q$ depends on $P$ in a degree $k$ ($0 \leq k \leq 1$), denoted $P \Rightarrow_k Q$, if

$$k = \gamma_P (Q) = \frac{|\text{POS}_P (Q)|}{|E|}$$ \hspace{1cm} \text{(4.19)}

If $k = 1$, $Q$ depends totally on $P$, if $0 < k < 1$ $Q$ depends partially on $P$, and if $k = 0$ then $Q$ does not depend on $P$.

The reduction of attributes is achieved by comparing equivalence relations generated by sets of attributes. A reduct is defined as a subset $R$ of the conditional attribute set $C$ such that $R \cap (D) = C \cap (D)$. A given dataset may have many attribute reduct sets, so the set $R$ of all reducts is defined as:

$$R = \{X: X \subseteq C; \gamma_X (D) = \gamma_C (D)\}$$ \hspace{1cm} \text{(4.20)}
The intersection of all the sets in \( R \) is called the core, the elements of which are that attributes which cannot be eliminated without introducing more contradictions to the dataset. In RSAR, a reduct with minimum cardinality is searched for or an attempt is made to locate a single element of the minimal reduct set \( R_{\text{min}} \subseteq R \):

\[
R_{\text{core}} = \bigcap_{X \in R} R
\]

(4.21)

The quick reduct algorithm, attempts to calculate a minimal reduct without exhaustively generating all possible sub-sets. This quick reduct algorithm stops when maximum dependency of one is reached, it ignores other potential features. Thus it is necessary to modify the algorithm in order to consider the effect of all potential features. The modified quick reduct algorithm works by removing features from full feature set as long as the value of dependency does not change. To find a minimal reduct, a new modified quick reduct algorithm has been introduced in Fig. 4.4 to accomplish the task.

```plaintext
Input: C, the set of all features
      D, the set of class attributes
Output: R, the attribute reduct, \( R \subseteq C \)
(1) \( R \leftarrow D \)
(2) do
(3) \( T \leftarrow R \)
(4) for each \( x \in R \)
(5) if \( \gamma(x, \ldots, D) = \gamma(D) \)
(6) \( T \leftarrow R - \{x\} \)
(7) \( R' \leftarrow R \)
(8) \( R \leftarrow T \)
(9) until \( \gamma(x, D) < \gamma(D) \lor R = R' \)
(10) return \( R' \)
```

Fig. 4.4: The modified quick reduct algorithm.

### 4.5 Classifiers

#### 4.5.1 Neural Network

The classification is performed by means of artificial neural network with input neuron, a number of hidden neurons which is used to obtain the best classification performance and one output neuron. Fig. 4.5 shows a three layer back propagation neural network and is used as a classifier. The optimized features are given as input to the neural network [88-89].
Optimal Feature Set

Input Neurons

Hidden Neurons

Output Neurons

Fig. 4.5: Three-layer back propagation neural network.

Where \( \lambda = 1 \), \( X = \sum W_i M_i \) and \( W_{ij} \) is the weight assigned between input and hidden layer and \( M_i \) is the input value. The output from the output layer is calculated using the sigmoid

\[
O = \frac{1}{1 + e^{-\lambda}}
\]

Where \( \lambda = 1 \), and \( X = \sum V_j H_j \) where \( W_{ij} \) is the weight assigned between hidden and output layer, and \( H_j \) is the output value from hidden neurons.

4.5.2 Support Vector Machine (SVM)

A SVM first maps the input points into a high-dimensional feature space and then finds a separating hyperplane that maximizes the margin between two classes in the space. A non-linear SVM is a supervisory binary classification algorithm. SVM constructs a maximal margin linear classifier in a high dimensional feature space, by mapping the original features via a kernel function \([90-95]\). The Gaussian radial basis function kernel is defined as

\[
K(x_i, x_j) = \exp\left(-\frac{|x_i - x_j|^2}{2\sigma^2}\right)
\]

Where \( x_i \) and \( x_j \) are two feature vectors and \( \sigma \) controls the size of the Gaussian kernel.
The decision function is taken as

\[ f(x) = \text{sgn}\left(\sum_{i=1}^{n} y_i \alpha_i K(x_i, x) + b\right) \]  \hspace{1cm} (4.25)

Function \( f(x) \) will have two classes; one corresponding to positive and other to negative sign. In this way the best possible kernel function are chosen to carry out the results. A proper combination of the kernel function and the training-test partition can maximize the performance of SVM-based classifier. A SVM-based classifier yields a much higher rate of classification accuracy than most other classifiers.

4.6 Results and discussions

This section presents the results of various modules. The first part of this section presents the results of classification of mass ROIs from normal ROIs. Next part presents the results of classification of masses and micro-calcifications respectively from mammograms.

Result of classification of mass ROIs from normal ones using rough set model are given in Table 4.2. Table 4.3 gives the classification with the new entropy function using Haralick’s features, Table 4.4 provides classification with energy, inertia and new entropy features with rough set feature selection and Table 4.5 gives the classification using the rough set model with Laws’ feature extraction and rough set based feature selection. Table 4.6 gives the classification using neural networks with back propagation learning.

Table 4.2: Classification with Harlick’s Features

<table>
<thead>
<tr>
<th>Actual</th>
<th>Predicted Masses</th>
<th>Predicted Normals</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Masses</td>
<td>27</td>
<td>20</td>
<td>57.45 %</td>
</tr>
<tr>
<td>Normals</td>
<td>30</td>
<td>63</td>
<td>67.74 %</td>
</tr>
<tr>
<td>Accuracy</td>
<td>47.37 %</td>
<td>75.90 %</td>
<td>64.29 %</td>
</tr>
</tbody>
</table>

True Positive Ratio: 57.45%
False Positive Ratio: 15.1%
Classification Rate: 64.29%

Table 4.3: Classification with Haralick’s features with new entropy

41
### Table 4.4: Classification with energy, inertia and new entropy features with rough set feature selection.

<table>
<thead>
<tr>
<th>Predicted Masses</th>
<th>Predicted Normals</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual Masses</td>
<td>33</td>
<td>14</td>
</tr>
<tr>
<td>Actual Normals</td>
<td>29</td>
<td>64</td>
</tr>
<tr>
<td>Accuracy</td>
<td>53.23%</td>
<td>82.05%</td>
</tr>
</tbody>
</table>

True Positive Ratio: 70.21%
False Positive Ratio: 18.95%
Classification Rate: 69.29%

### Table 4.5: Classification with Laws' feature extraction with rough set feature selection.

<table>
<thead>
<tr>
<th>Predicted Mass</th>
<th>Predicted Normal</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual Mass</td>
<td>35</td>
<td>12</td>
</tr>
<tr>
<td>Actual Normal</td>
<td>15</td>
<td>78</td>
</tr>
<tr>
<td>Accuracy</td>
<td>70%</td>
<td>86.67%</td>
</tr>
</tbody>
</table>

True Positive Ratio: 74.46%
False Positive Ratio: 14.3%
Classification Rate: 80.71%

### Table 4.6: Classification using neural networks with back propagation learning.

<table>
<thead>
<tr>
<th>Predicted Masses</th>
<th>Predicted Normals</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual Masses</td>
<td>32</td>
<td>15</td>
</tr>
<tr>
<td>Actual Normals</td>
<td>19</td>
<td>74</td>
</tr>
<tr>
<td>Accuracy</td>
<td>62.74%</td>
<td>83.14%</td>
</tr>
</tbody>
</table>

42
4.6.1 Classification results for classifying masses based on shapes

The results of classification are shown in Table 4.7 using rough set based feature selection and SVM based classification. Table 4.8 shows the results of classification with neural networks.

**Table 4.7: Classification of masses with rough set feature selection and SVM.**

<table>
<thead>
<tr>
<th>Actual Mass</th>
<th>Predicted Mass</th>
<th>Predicted Normal</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>2</td>
<td>90%</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>17</td>
<td>85%</td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>85.71%</td>
<td>89.47%</td>
<td>91.37%</td>
</tr>
</tbody>
</table>

The performance due to the proposed approach is the following:

True Positive Ratio: 90%
False Positive Ratio: 10.5%
Classification Rate: 91.37%

**Table 4.8: Classification using neural networks with back propagation learning.**

<table>
<thead>
<tr>
<th>Actual Masses</th>
<th>Predicted Masses</th>
<th>Predicted Normals</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>3</td>
<td>85%</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>17</td>
<td>85%</td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>85%</td>
<td>85%</td>
<td>85%</td>
</tr>
</tbody>
</table>

True Positive Ratio: 85%
False Positive Ratio: 15%
Classification Rate: 85%

4.6.2 Classification of micro-calcification clusters

The results of classification are shown in Table 4.9 using rough set based feature selection and SVM based classification. Table 4.10 gives the classification with neural networks.

**Table 4.9: Classification of masses with rough set feature selection and SVM based classifier.**
Predicted Mass | Predicted Normal | Accuracy
---|---|---
Actual Mass | 18 | 2 | 90%
Actual Normal | 1 | 11 | 91.66%
Accuracy | 94.73% | 84.61% | 93.54%

The improved performance is now indicated by the following ratios:
- True Positive Ratio: 90%
- False Positive Ratio: 15.39%
- Classification Rate: 93.54%.

Table 4.10: Classification using Neural Networks with back propagation learning.

| Predicted Masses | Predicted Normals | Accuracy |
---|---|---|
Actual Masses | 17 | 3 | 85%
Actual Normals | 3 | 9 | 75%
Accuracy | 85% | 75% | 85%

From the above tabulated results we infer that the classification accuracy is improved by selecting the features using rough set theory. Moreover the use of new entropy function in place of Shannon’s entropy improves the classification accuracy drastically. The new entropy function has four parameters that can be varied. Because of this freedom one can derive several functions by the choice of these parameters. Two texture feature extraction techniques, viz., co-occurrence matrix based scheme and Laws’ texture scheme, are compared. Out of the two the co-occurrence based method is found to be superior to the Laws’ method. The results of classification of both masses and micro-calcifications using both rough set based SVM and neural networks indicate that the former is superior to the latter.

4.7 Conclusions

The co-occurrence matrix based texture analysis suits well for mammographic images because it captures the spatial arrangement of pixels into consideration. The drawback of this technique is
that it gives rise to a number of features. Malignancy in masses is recognized mainly by its spiculated shape and high contrast or any of these. Certain important shape features like spiculation, relative contrast, circulatory, compactness and radial angle which can categorize the masses are also used for the classifications. In micro-calcification cluster shape is important feature. Others include number of calcifications and their contrast. The features used in this work are best suited for classifying the masses and micro-calcifications.

False positive (FP) and false negative (FN) cases are considered errors in these experiments because they will degrade the overall performance of the detection techniques. Otsu's method with classification based on GLCM features shows the best performance as it produces less error than the other two segmentation methods. In brief, Otsu's method with GLCM features for classification of masses obtained good results in detecting any types of masses in digital mammogram.

For classifying the masses from the normal ROIs co-occurrence matrix based feature selection scheme is developed by using new entropy in place of Shannon's entropy. It shows drastic improvement in the classification results. SVM when used in combination with the tuned kernel function and the training-test partition performs well for the classification of masses and micro-calcifications. This approach is found to be very effective. The classification system is based on rough set theory and it uses the SVM and neural networks for classification. The classification accuracy of 80% is achieved using SVM in classifying masses from normal regions. The classification accuracies of 91.37% and 93.54% are achieved using SVM in classifying the masses and micro-calcifications respectively into their specific classes, i.e. benign and malignant. With neural networks, the classification achieved is 70% only.