International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161

Available at http://inpressco.com/category/ijcet

Research Article

Classification of Clustered Microcalcifications in Mammograms using Topological and Shape Features

Minu Baby†* and Vipin. V†

†Department of Electronics and Communication Engineering, M.G University, Kottayam, India

Accepted 30 Sept 2015, Available online 02 Sept 2015, **Vol.5, No.5 (Oct 2015)**

Abstract

Classification of benign or malignant microcalcification clusters in mammograms is a major diagnostic challenge for radiologists. Clinical studies have explained that malignant microcalcifications tend to be small and densely distributed while benign microcalcifications are generally larger, and more diffusely distributed. Most of the existing works are focused either on the shape, texture or the distribution of the clusters. Topological feature extraction method fail to discriminate malignant from benign when a single microcalcification is detected using segmentation approach. In this case shape feature of the individual microcalcification will discriminate malignant from benign. The size of the microcalcification is determined by the shape features and the distribution of the microcalcification is determined by the topological features of the cluster. Hence, topological features and shape features are extracted, which constitute the feature space for classifying microcalcification clusters. SVM is employed for classification. High classification accuracy and high true positive rate are obtained.

Keywords: *Mammography, Microcalcifications, Graphs, Topology, Shape, Perimeter, Classification*

1. Introduction

Breast cancer is presently the most probable cancer affecting women worldwide. Primitive prevention seems impractical since the origin of this disease still remain unknown. Detection in the beginning is the key to improve breast cancer prediction. A proper treatment can be provided if the cancer is detected earlier. Early detection methods of breast cancer include screening by mammography and analytic breast examination. Digital mammography is an x-ray analysis of the breasts. Microcalcifications are tiny deposits of calcium salts inside breast tissue that arise as small shining or bright spots in mammograms. These small bright spots forms a cluster in the breast tissue. It is very important to perform a suitable evaluation of microcalcification clusters to determine whether they are benign or malignant. However, not all microcalcification clusters vitally displays the presence of cancer, only certain types of microcalcifications are correlated with a high possibility of malignancy. Fig. 1 shows two mammographic image patches taken from the Mammographic Image Analysis Society (MIAS) database, containing a benign microcalcification cluster and a malignant microcalcification cluster, respectively.

It is crucial for radiologists to keep both precise and consistent evaluation for the huge number of mammograms generated in screening. Due to a high rate of avoidable biopsies, differentiating between malignant and benign clusters is a difficult and tedious task for radiologists. Only 20-30 percentage of breast biopsy cases confirmed by radiologists turn out to be of malignant nature. Radiologists have a favorable chance to improve their earlier prediction with the aid of digital image processing, pattern recognition and artificial intelligence. i. e, the diagnosis accuracy can be improved with the help of computer systems. In order to improve the prediction accuracy and to reduce the time duration of radiologists interpreting microcalcifications in mammograms, computer-aided diagnosis (CAD) systems have been developed.

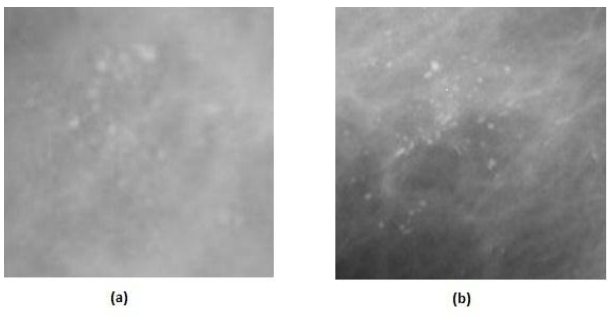


Fig.1 Example mammographic image patches containing microcalcification clusters:(a) benign (b) malignant

*Corresponding author: **Minu Baby**

Rest of the paper is organized as follows: Section 2 summarizes the system model. Section 3 describes the methodology and design in detail. Results and discussion in section 4. Experimental Results are discussed in section 5 and paper is concluded in section 6.

2. System model

The block diagram of the proposed method is shown in fig: 2. The proposed system consists of mainly four stages: preprocessing, segmentation, feature extraction and classification. The mammogram image is initially enhanced by top-hat transformation. The clusters in the image are detected by k-means clustering method. The segmented image so generated by k means clustering will have M number of microcalcifications. One of the most important steps for the classification task is extracting proper features efficient for distinguishing between classes. A method for classifying and modelling microcalcification clusters in mammograms is proposed based on their topological and shape features. Classification is performed using SVM. Also, when a single microcalcification is detected due to underdetection, the feature set is not able to discriminate as malignant or benign. Topological features can be extracted only for more than one microcalcification. This case is avoided by considering the perimeter of the single node.

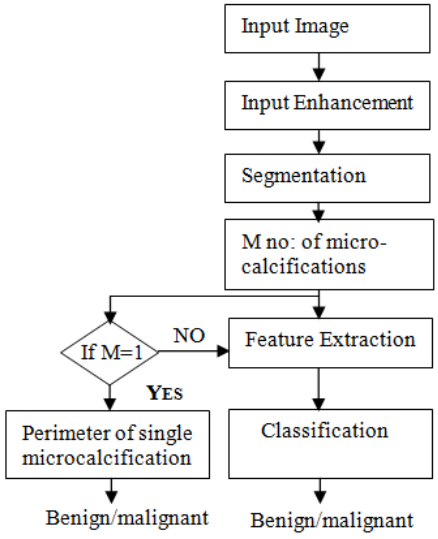


Fig.2 Block diagram of system model

3. Methodology and design

3.1 Image acquisition

The data used in the experiments consist of two datasets. These datasets are composed of image patches of different cases. The first dataset was extracted from the MIAS database, containing image patches with the same size of 256*256 pixels. The second dataset was taken from the digital database for screening mammography (DDSM) database containing image patches. The diagnostic standard (benign or malignant) of all microcalcification clusters in this experiment has been given by biopsy.

3.2 Image enhancement

The function of mammogram enhancement is to sharpen the edges or boundaries of ROIs, or to boost the contrast between ROIs and background. The morphological top-hat transformation is used for the mammogram image enhancement. In digital image processing and mathematical morphology, top-hat transform is an operation that takes small elements and details from given images. The top-hat transform is defined as the difference between the input image and its opening by some structuring element.(D. Betal *et.al, 1997).*

3.3 Image segmentation

After performing the top-hat transformation, the aim is to extract the microcalcification cluster part. For that segmentation algorithm is applied on the image. There are two different goals for the segmentation of microcalcifications. One is to obtain the locations of cautious areas to aid radiologists for diagnose. The other is to classify the abnormalities of the breast into benign or malignant. K-means clustering is used for segmentation. In the k-means algorithm at first define the number of clusters k. Then k-cluster center are selected randomly. The distance between the each pixel to each cluster centers are estimated. The distance is of simple Euclidean function. The grouping is done by diminishing the Euclidean distance between data and the corresponding cluster centroid.

3.4. Feature extraction

Clinical studies have revealed that malignant microcalcifications tend to be small and densely distributed while benign microcalcifications are generally larger, and more diffusely distributed. Hence, a method is proposed a for classifying microcalcification clusters in mammograms based on their topological and shape features.

3.4.1 Topological feature extraction

The topological feature extraction method is to find out the distribution of the clusters. This consists of three main steps: connectivity between microcalcifications are estimated using morphological dilation at multiple scales, a microcalcification graph is generated at each scale using the connectivity estimated, extracting topological features from these microcalcification graphs at multiple scales.(Zhili Chen *et.al*, 2015)

3.4.1.1 Connectivity evaluation

Here the connectivity between the individual microcalcification is estimated by performing morphological dilation at multiple scales. The scale

corresponds to the radius of the structuring element measured in pixels. As the scale increases dilation absorbs nearby pixels into individual microcalcifications. Therefore the connectivity between microcalcifications within the cluster is varied by the multiscale dilation. As the scale increases the connectivity will be higher for dense distribution and it will be lower for the wide distribution.

3.4.1.2 Microcalcification graph generation

The topology of microcalcification clusters is represented in graphical form. Based on the connectivity relationship between microcalcifications within a cluster a microcalcification graph is generated. In a microcalcification graph, each node serve as an individual microcalcification, and an edge is connected between the nodes if the microcalcifications are connected. The connectivity of the microcalcification cluster increases from small to large scales and the corresponding microcalcification graph becomes denser and denser and more edges are created in the graph.

3.4.1.3 Topological Feature Extraction at multiple scales

The topological features of the clusters are extracted from the generated microcalcification graph over a range of scales. Let G(V,E) represent a graph, where V is the vertex set and E is the edge set, and use |V| (cardinality of V) and |E| (the cardinality of E) to denote the number of vertices and the number of edges in G, respectively. The topological features extracted are:

1) Average vertex degree: The degree matrix *D(i, j)* of a graph *G(V,E)* is a $|V|\times|V|$ diagonal matrix containing the degree of vertex *i* at entry *(i, i),* defined as:

$$D(i,j) = \begin{cases} d(i), if\ i = j \\ 0, Otherwise \end{cases} \quad (1)$$

where $d(i) = \sum_{j\varepsilon V} a_{ij}$ is the number of edges incident to vertex *i*. This diagonal matrix contains the information about the degree of each vertex. i.e, it gives the number of edges attached to each vertices. The average vertex degree *F* is given by:

$$F = \frac{\sum_{i\varepsilon V} d(i)}{|V|} \quad (2)$$

Since the benign microcalcification cluster is less connected than malignant cluster the average vertex degree values of the benign cluster are smaller than those of the malignant cluster over the entire range of scales.

2) Maximum vertex degree: It is the maximum number of vertex degree in the graph. The maximum vertex degree *mvd* is given by:

$$mvd = max_{i\varepsilon V} d(i) \quad (3)$$

Since at higher scales less number of edges are connected for the benign microcalcification cluster the maximum vertex degree values for the benign cluster are also smaller than those of the malignant cluster.

3) Average vertex eccentricity: The eccentricity of a vertex *i* in a graph *G(V,E)*, denoted by *e(i)*, is defined as the maximum distance from itself to any of the reachable vertices in *G*, given by:

$$e(i) = max_{j\varepsilon V} dist(i,j) \quad (4)$$

The average vertex eccentricity, *C* is given by:

$$C = \frac{\sum_{i\varepsilon V} e(i)}{|V|} \quad (5)$$

where $|V|$ denotes the number of vertices. At the first few scales, most microcalcifications are seperated from others in the cluster, which results in small average eccentricity values. When the scale increases to a particular value, the maximum average eccentricity is achieved, in which case the formely isolated microcalcifications are engaged into a connected component with a relatively large diameter. After that, as the scale further increases, microcalcifications get connected to a greater extent and the average vertex eccentricity starts to decrease. The maximum average eccentricity of the malignant cluster is larger than that of the benign cluster.

4)Diameter: The diameter of the graph is defined as the maximum eccentricity. When all microclacifications are connected the diameter value evenly goes down toward the minimum value . The maximum diameter of the benign cluster is smaller than that of the malignant cluster

5)Percentage of isolated points: A vertex is considered isolated if it has degree equal to 0. The set *X* denotes the set of vertices within a graph *G(V,E)* that are isolated. The percentage of isolated points, *P* is given by:

$$P = \frac{|X|}{|V|} \quad (6)$$

Where |V| denotes the number of vertices and |X| denotes the number of isolated vertices. The percentage of isolated points for the malignant cluster are smaller than those of the benign cluster, and also, the malignant cluster tends to 0% at a much smaller scale than the benign cluster. This is due to the dense distribution of the malignant cluster.

Maximum vertex degree, average vertex degree , average vertex eccentricity, diameter, percentage of isolated points are the set of graph features that will be extracted from the generated microcalcification graphs. A set of microcalcification graphs is generated based on the connectivity relationship between microcalcifications after performing morphological dilation at multiple scales, denoted by $G = G_3, \ldots, G_S$ where S is the largest scale, and G_s ($s = 3, 6 \ldots, S$)

indicatess the microcalcification graph generated at the s^{th} scale. The largest scale used in this experiment is 30. The five graph metrics from each graph in *G*, which produces five graph feature sets covering *S* scales.

3.4.2 Shape feature extraction

One of the typical feature of the microcalcifications, which could be a sign of malignancy, is the size of calcifications. The features extracted are: mean of microcalcifications in a cluster, standard deviation of size of microcalcifications in a cluster, standard deviation of size of boundary pixels, maximum compactness and the perimeter. A good measure for classifying microcalcifications by their shape is perimeter. It is found by calculating the distance between successive boundary pixels. Compactness is the ratio of squared perimeter to the area. The topological and shape features are concatenated into the feature set for the subsequent classification process.

3.5 Classification

SVM(support vector machine) is used for classifying microcalcification clusters into malignant and benign. Support Vector Machine (SVM) is a useful tool to distinguish between malignant and benign micro-calcifications by set of all training vectors (support vectors). SVM aims to obtain a maximum margin between the two classes.

3.6 Single node detection case

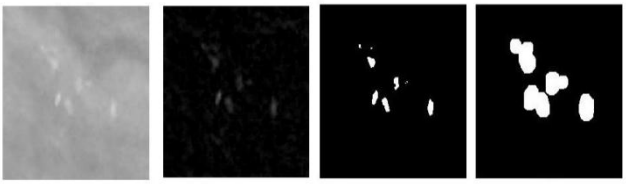
Another shape feature that has proven to be a good measure for classifying microcalcifications by their shape is average perimeter. When only a single microcalcification is detected from the cluster due to underdetection, the concatenated topological and shape feature set will fail to discriminate malignant from benign. The topological feature extraction works only for the case with more than one node. Therefore if one node is detected due to underdetection the node is classified by using the perimeter of the node. The perimeter is found by calculating the distance between successive boundary pixels.

4. Results and discussion

Simulation is done using MATLAB. The images are taken from MIAS dataset and DDSM dataset. The diagnostic standard of all microcalcification clusters in this experiment has been given by biopsy.

4.1 Benign case

The benign microcalcifications are widely distributed. The enhanced and segmented output of the benign case is shown in fig. 3 (b) and (c) respectively. In the experiment connectivity is estimated at multiple scales. i.e, for scales = 3, 6, 9....30. The corresponding graphs are also obtained at multiple scales. The connectivity estimation at scale = 9 is shown in fig. 3 (d). The microcalcification graph generation at scale = 9 is shown in fig. 4 .



(a) (b) (c) (d)

Fig.3 Example benign case(MIAS database: mdb252): (a) Input image, (b) Enhanced image, (c) Segmented output, (d) Connectivity estimation at scale = 9

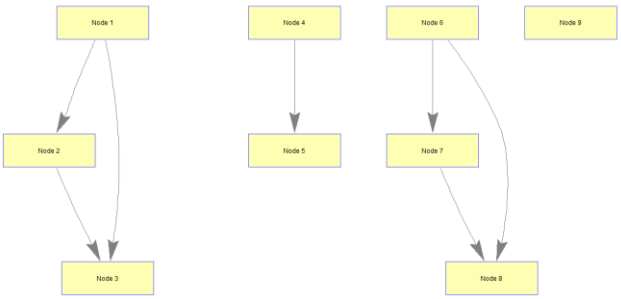
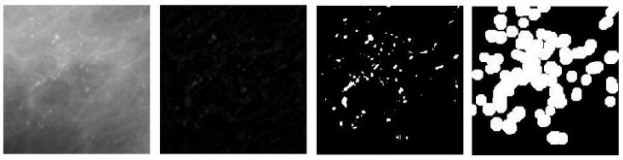


Fig.4 Microcalcification graph generation of the benign case in fig. 3 at scale = 9.

4.2 Malignant case

The malignant microcalcifications are closely distributed. The enhanced and segmented output of the malignant case is shown in fig. 6 (b) and (c) respectively. The connectivity estimation at scales = 9 is shown in fig. 6 (d). The microcalcification graph generation at scale = 9 is shown in fig. 6 . The graph is generated at different scales.the scale corresponds to scale=3, 6, 9,...., 30. At each scale set of features are extracted from the graph.



(a) (b) (c) (d)

Fig.5 Example malignant case(MIAS database: mdb233): (a) Input image, (b) Enhanced image, (c) Segmented output, (d) Connectivity estimation at scale = 9

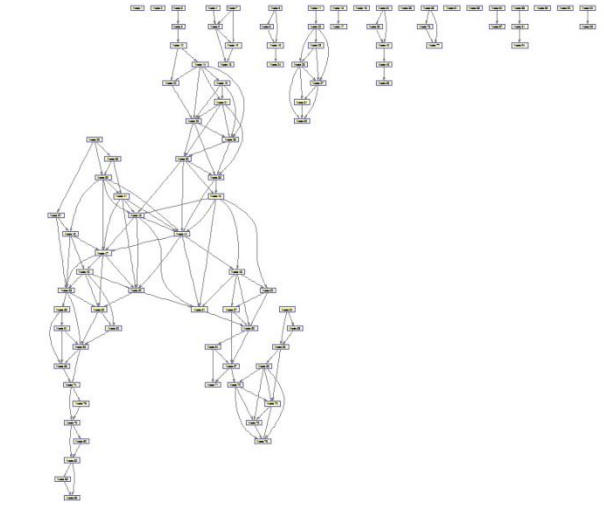


Fig.6 Microcalcification graph generation of the malignant case in fig. 5 at scale = 9.

4.3 Topological features at multiple scales

The topological features extracted from the graph such as average vertex degree, maximum vertex degree, eccentricity ,diameter will be higher for malignant case and lower for benign case at different scales. Also, the percentage of isolated points will be smaller for malignant and higher for benign. Fig. 7,8,9,10,11 shows the feature set extracted from the example benign (green line) and malignant (red dotted line) microcalcification clusters in fig 3 and 5.

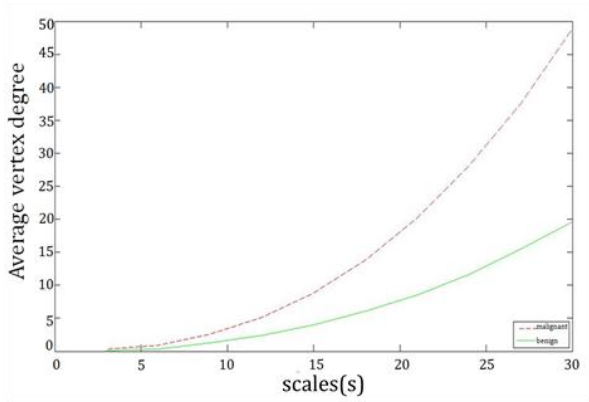


Fig.7 Average vertex degree

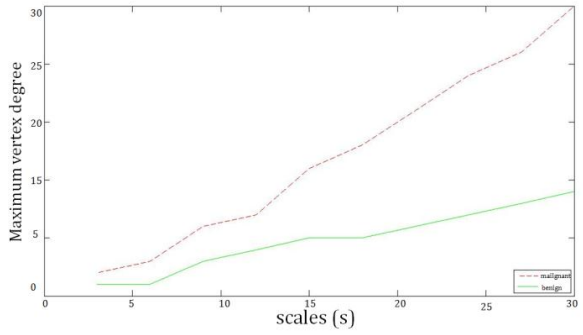


Fig.8 Maximum vertex degree

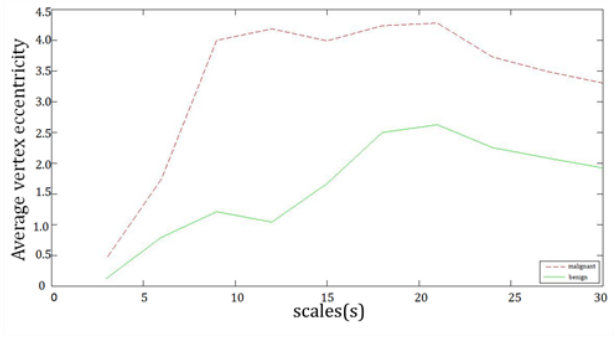


Fig.9 Average vertex eccentricity

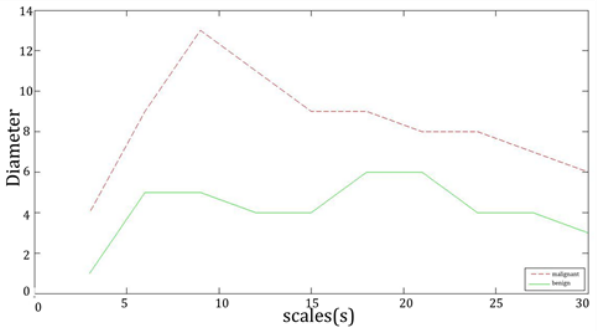


Fig.10 Diameter

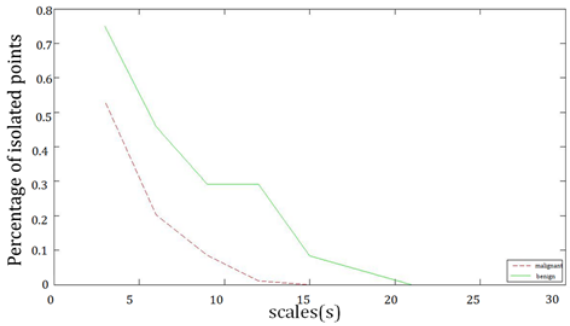


Fig.11 Percentage of isolated points

4.4 Shape features at zero scale

The shape features extracted from the segmented output of both cases and the corresponding values are listed in Table 1.

Table 1 Shape features of benign and malignant cluster in fig. 3 and fig. 5

No:	Features	Benign	Malignant
1	Average perimeter	30.835	11.2328
2	Standard deviation of size of boundary pixels	20.72	9.7009
3	Maximum compactness	19.2013	35.0972
4	Mean of microcalcifications	0.247	0.0241
5	Standard deviation of size of microcalcification	0.9912	0.127

4.5 Single node case

Based on the clinical studies, malignant microcalcification are smaller than benign. So the perimeter will be smaller for malignant case. Here it is found that the perimeter of the single calcification is 40.557. Therefore it is assigned to the benign stage.

(a) (b) (c)

Fig.12 Single node detection case: benign case (MIAS,mdb240) (a) Input image, (b) Enhanced output, (c) Segmented output

5. Experimental results

The accuracy of the proposed method is 93.9 percentage with one malignant case and one benign case are misclassified. The existing method concentrates only on topological features for which two benign cases and two malignant cases are misclassified in this experiment. One benign case is the one with single node detection. It is correctly classified in the proposed method by considering the perimeter of the node. One of the misclassified malignant case is correctly classified by the concatenation of shape and topological features. The true positive rate of the proposed method is improved from 0.89 to 0.94 and the false positive rate is reduced from 0.14 to 0.07.

Table 2 Experimental results

Met-hods	Total images	Tested	Corre-ctly classified	True posit-ive rate	False posit-ive rate	Accuracy
Existing	74	33	29	0.89	0.14	87.8
Proposed	74	33	31	0.94	0.07	93.9

Conclusion and future scopes

This paper has successfully investigated a method for classifying microcalcification clusters in mammograms based on morphological topology analysis and shape features. This is an approach to analyze microcalcifications in terms of the connectivity and shape for classifying malignant from benign clusters. Since shape features are also considered the problem of single node detection in previous work of topological feature extraction method can be avoided.Also the one of the misclassified benign case is correctly classified. The topology or connectivity of microcalcification clusters was analyzed using multiscale morphology. A set of microcalcification graphs were created to explain the topological structure of microcalcification clusters at multiple scales. When examining the topology of microcalcification clusters, five graph metrics are extracted from microcalcification graphs created at different scales, which are average vertex degree, maximum vertex degree, average vertex eccentricity, diameter, and percentage of isolated points. The resulting five graph feature sets were aggregated with the shape features like mean, standard deviation of size of microcalcifications, average perimeter, maximum compactness and standard deviation of size of boundary pixels constituted the feature vector, which has been used to discriminate the clusters into benign and malignant. The perimeter is used for classifying the single node case. Also, the topological measures and shape features can be linked to clinical understanding. The proposed method can be continued by the evaluation using a greater number of digital mammograms and different classifiers like random forests, ANN, and kNN could also be considered in the future.

References

Zhili Chen ,Harry Strange, Arnau Oliver, Erika R. E. Denton, Caroline Boggis, and Reyer Zwiggelaar(2015), Topological Modeling and Classification of Mammographic Microcalcification Clusters, *IEEE transactions on biomedical engineering*,62, 1203-1214.

Liang Shen, Rangaraj M. Rangayyan, and J. E. Leo Desautels(1994),Application of shape analysis to mammographic calcifications, *IEEE Trans. Med. Imag*, 13, 263-274.

Heang-Ping Chany, Berkman Sahiner and Nicholas Petrick(1997), Computerized classification of malignant and benign microcalcifications on mammograms: Texture analysis using an artificial neural network*, Phys. Med. Biol*, 42, 549–567.

A. Papadopoulos, D.I. Fotiadis and A. Likas(2005), Characterization of clustered microcalcifications in digitized mammograms using neural networks and support vector machines, *Artif. Intell. Med*, 34 ,141-150.

D. Betal, N. Roberts and G.H Whitehouse(1997), Segmentation and numerical analysis of microcalcifications on mammograms using mathematical morphology, *Brit. J. Radiol*, 70, 903–917.

A. Wroblewska,P. Boninski, A. Przelaskowski, and M. Kazubek(2003), Segmentation and feature extraction for reliable classification of microcalcifications in digital mammograms, *Opto-electronics review*, 11, 227-235.

J. Suckling(1994),The mammographic image analysis society digital mammogram database, *Proc. Excerpta Med. Int. Congr. Ser*, 375–378.

M. Heath (2000), The digital database for screening mammography,'' *Proc. 5th Int. Workshop Dig. Mammography*, 212–218.

H.D. Cheng, Xiaopeng Cai, Xiaowei Chen and Liming Hu(2003), Computer-aided detection and classification microcalcifications in mammograms: A survey, *Pattern Recog*, 36, 2967-2991.