

New Feature Extraction Method for Mammogram Computer Aided Diagnosis

Belal K. Elfarra¹ and Ibrahim S. I. Abuhaiba²

¹Department of Information Systems, Islamic University of Gaza, Palestine

²Department of Computer Engineering, Islamic University of Gaza, Palestine
fbelal@iugaza.edu.ps, ihaiba@iugaza.edu.ps

Abstract

Computer-aided diagnosis (CADx) is usually used as a second opinion by the radiologists. Improving CADx increases the treatment options and a cure is more likely. The main objective of this research is to introduce a new method for feature extraction in order to build a CADx model to discriminate between cancers, benign, and healthy parenchyma. For feature extraction, we use both human features, which are obtained by Digital Database for Screening Mammography (DDSM), and computational features, and we propose a new feature extraction method called Square Centroid Lines Gray Level Distribution Method (SCLGM). The experimental results are obtained from a data set of 410 images taken from DDSM for different types. Our method select 31 features from 145 extracted features; 18 of the selected features are from our proposed feature extraction method (SCLGM). We used both Receiver Operating Characteristics (ROC) and confusing matrix to measure the performance. In training stage, our proposed method achieved an overall classification accuracy of 96.3%, with 92.9% sensitivity and 94.3% specificity. In testing stage, our proposed method achieved an overall classification accuracy of 89%, with 88.6% sensitivity and 83.3% specificity.

Keywords: Breast Cancer, Mammogram, Feature Extraction, Feature Selection, Computer Aided Diagnosis, Genetic Algorithm, Forward Sequential

1. Introduction

Breast cancer is the most common form of cancer among women and is the second leading of death after lung cancer. The American Cancer Society [1] estimates that in 2011 approximately 230,480 women in the US will be diagnosed with tumor breast cancer, and about 39,520 women will die from breast cancer. Early detection and diagnosis of breast cancer increase the treatment options and a cure is more likely. One of the most effective tools for early detection of breast cancer is the mammography.

Visual interpretation of a mammogram is a tedious and fatiguing process that generally requires a magnifying glass.

The abnormality may be overlooked in a way that for each thousand cases we have only three to four cancerous. So, the probability of false negatives is high. Here the radiologists fail to detect 10% to 30% of cancers. Two thirds of these false negative results are due to missed lesions that are evident retrospectively [2].

Also, a significant level of false positives were reported, that the positive predictive value is less than 35% which means a high proportion of biopsies are performed on benign lesions. Avoiding benign biopsies would spare women anxiety, discomfort, and expense.

Two systems have been developed to help the radiologists in reading mammogram. The first system is computer-aided detection (CADe) which has improved radiologists' accuracy of detection of breast cancer [3, 4]. The second system is computer-aided diagnosis (CADx) which classifies the detected regions into malignant or benign categories to help the radiologists in recognizing the next step, biopsy or short-term follow-up mammography. Most diagnosis algorithms of CADx consist of one stage with five steps: preprocessing, segmentation, feature extraction, feature selection, and classification. It begins with a lesion region or a region of interest (ROI) that contains the abnormality and outputs the likelihood of malignancy or a management recommendation.

The performance of CADx depends more on the optimization of the feature selection than the classification methods. However, the feature space is very large and complex due to the wide diversity of the normal tissues and the variety of the abnormalities. Using excessive features may degrade the performance of the algorithm and increase the complexity of the classifier. For this, the main goal of this research is to evaluate methods for feature extraction, introduce other feature extraction techniques and enhance feature selection method to have best feature that guarantee the enhancement of classification with less dimension.

2. Related Work

The density tissues, which surround and mask the suspicious region, makes segmentation step very difficult. To overcome this problem we have to enhance feature extraction technique in order to obtain a large space of good features from which we could select discriminative features to improve the quality of lesion classification. One of the feature extraction techniques that are mostly used in literatures is Gray-Level Co-occurrence Matrix (GLCM). In [17, 18] authors used five co-occurrence matrices statistics extracted from four spatial orientations, with $\theta \in (0, \pi/4, \pi/2, 3\pi/4)$ and pixel distance ($d=1$) to detect masses in mammograms. The extracted features won't be able to be discriminative with cases of cancers due to their non-uniform shape and margins. So our contribution is to increase the number of spatial orientations and increase the range of pixel distances.

Blot [9] compares the GLCM of the local ROI and compare it with the mean GLCM obtained from a number of equal size areas surrounding the local ROI. The purpose is to compare the difference between these two matrices obtaining a probability estimate of the abnormal image structures in the ROI. This method will have bad results if ROI is surrounded by normal parenchyma with high gray level value such as nipple and lactation ducts.

A new approach of texture classification of 3-D Ultrasound (US) breast diagnosis using run difference matrix (RDM) with neural networks is developed in [10]. The accuracy reaches 91.9%. The results of this method are good, but not highly guaranteed because the used dataset is small; it includes just 54 malignant and 161 benign tumors.

Generally, the greater the number of features is, the higher the recognition rate will be. However, when the number of features is large but the number of training samples is small, features that have little or no discriminative information weaken the performance of classifiers. This situation is typically called the curse of dimensionality [11], in this situation we have to choose a feature subset yielding the highest performance.

It is very difficult to predict which feature or feature combinations will achieve better in classification rate. We will have different performances as a result of different feature combinations. Relatively few features used in a classifier can keep the classification performance robust [12]. Therefore, we have to select an optimized subset of features from a large number of available features. Two major methods for feature selection have been employed for CADx in mammography [13]: Stepwise feature selection and Genetic algorithm (GA). Stepwise feature selection are used in [14, 15, 16].

In [17], a feature selection method using stepwise Analysis Of Variance (ANOVA) Discriminant Analysis (DA) is used for classifying mammogram masses. This approach combines the 17 shape and margin properties of the mass regions and classifies the masses as benign or malignant. In ANOVA DA the discriminating power of each feature is estimated based on grouping class variable. Principal component analysis (PCA) does feature extraction but it doesn't consider the grouping class variable. The experiment is performed on 300 DDSM [6] database mammograms.

The main idea is to start with empty selected feature pool, and then at each step followed, one available feature is added to or removed from the selected feature pool with respect to the result of analyzing its effect on a selection criterion.

Wei [18], provide five selection criteria: (1) the minimization of Wilks' lambda, (2) the minimization of unexplained variance, (3) the maximization of the between-class F statistic value, (4) the maximization of Mahalanobis distance, and (5) the maximization of Lawley-Hotelling trace. Most studies in mass detection [15, 16, 19] employed the minimization of Wilks' lambda as the selection criterion, which is defined as the ratio of within-group sum of squares to the total sum of squares.

In [16], the authors test all available selection criteria. A set of 340 features is reduced to 41 features with the stepwise feature selection. Stepwise feature selection techniques based on sequential search method and the main disadvantages of these techniques is it's availability of falling in local minima. Veldkamp [20] used cluster shape features, cluster position features, and distribution features for the classification of calcifications. They used a sequential forward selection procedure for feature selection.

3. Methodology

When radiologists need to diagnose a mass in mammogram, they look for some significant features that discriminate malignant from benign masses. These visual features -which are based on shape, size and margin - could have different interpretation based on radiologist's opinion and experience. To solve the problem of these different interpretations, more discriminative features should be extracted. Computer provides multiple methods for obtaining these discriminative features, which are done in two steps:

Feature extraction: this step is responsible for extracting all possible features that are expected to be effective in diagnosing a ROI in mammogram, without concerning the disadvantages of excessive dimensionality.

Feature selection: this step is responsible for reducing the dimensionality by removing redundant features and searching for the best significant features to avoid the curse of dimensionality.

3.1 Feature Extraction

In [21] we use and modify two of known methods for computational feature extraction, which are Spatial Gray Level Dependence Method (SGLDM) and Run Difference Method (RDM) to enhance their power in describing textural characteristics of the mass patterns.

Table 1 shows the extracted features by using the two mentioned methods. More details about SGLDM and RDM methods are given in our previous work [21].

Table 1. The Extracted Features Obtained in our Previous Work [21]

Spatial Gray Level Dependence Method (SGLDM) features		
Correlation:	$\frac{\sum \sum [() ()]}{\dots}$	3.2
Homogeneity:	$\frac{\sum \sum [()]}{\dots}$	3.3
Variance:	$\sum \sum ()$	3.4
Contrast:	$\sum \sum ()$	3.5
Energy:	$\sum \sum ()$	3.6
Maximum probability:		3.7
Sum average:	$\sum \sum ()$	3.8
Cluster prominence:	$\sum \sum ()$	3.9
Mean:	$\frac{\sum}{\dots} \quad 3.10$	
Mean absolute deviation:	$\frac{\sum \dots }{\dots}$	3.11
Minimum:		3.12
Maximum:		3.13
Variance:	$\frac{\sum (\dots)}{\dots}$ Where \dots is the mean of feature \dots over all $d \in \{1, 2, \dots, L/2\}$.	3.14
Skewness:	$\frac{\sum (\dots)}{\dots}$ $\frac{\dots}{\dots}$	3.15

Run Difference Matrix (RDM)		
Large difference emphasis (LDE):	$\sum (\quad) (\quad)$	3.24
Sharpness:	$\sum (\quad) (\quad)$	3.25
Second Moment of DGD (SMG):	$\sum (\quad)$	3.26
Second Moment of DOD (SMO)	$\sum (\quad)$	3.27
Long distance emphasis for large difference (LDEL)	$\sum (\quad) (\quad)$	3.28
Human Features		
Age		3.29
Density		3.30

In this paper we propose a new feature extraction method; we call it “Square Centroid Lines Gray Level Distribution Method (SCLGM)”. We use the smallest square that includes the segmented mass with zero background. As shown in Figure 1, the relations between the pixels of each centroid line provide discriminative information about the mass type.

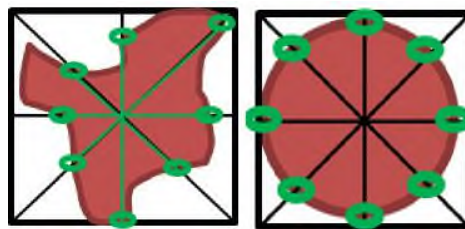


Figure 1. Centroid Lines Pixels Provide Discriminative Information About the Mass Type

In our method, we choose four centroid lines (C_i): C_1 at $\theta = 0$, C_2 at $\theta = \pi/4$, C_3 at $\theta = \pi/2$ and C_4 at $\theta = 3\pi/4$, which are the lines that pass through the square’s center point. We represent the gray level points at each centroid line by C_{gi} :

$$\left\{ \begin{matrix} (\quad) \\ (\quad) \\ (\quad) \\ (\quad) \end{matrix} \right\} \quad 3.29$$

After defining C_{gi} , and as we do in SGLDM, a set of statistics need to be computed before extracting textural features. In the following, we compute these statistics and extracted textural features:

1- **Mean:** which measure the average of the gray level of each line (C_{gi}):

$$\bar{C}_i = \frac{\sum_{j=1}^n C_{gij}}{n} \quad (3.30)$$

2- **Variance and its difference vector:** variance is used to measure the variability of the gray levels for each central line around their mean value:

$$V_i = \frac{\sum_{j=1}^n (C_{gij} - \bar{C}_i)^2}{n} \quad (3.31)$$

As a result, we have four variance values one for each line, then we compute the differences between these variances as a vector

$$D_i = (V_{i1} - V_{i2}, V_{i2} - V_{i3}, V_{i3} - V_{i4}) \quad (3.32)$$

Table 2 shows all extracted features using equations 3.31 and 3.32 based on our proposed feature extraction method SCLGM.

Table 2. The Extracted Features for Variance and its Difference Vector using SCLGM

Average variance:	$\bar{V}_i = \frac{\sum_{j=1}^4 V_{ij}}{4}$	3.33
Average difference of variance:	$\bar{D}_i = \frac{\sum_{j=1}^3 D_{ij}}{3}$	3.34
Minimum variance:	$V_{i_{min}} = \min(V_{i1}, V_{i2}, V_{i3}, V_{i4})$	3.35
Minimum difference of variance:	$D_{i_{min}} = \min(D_{i1}, D_{i2}, D_{i3})$	3.36
Maximum variance:	$V_{i_{max}} = \max(V_{i1}, V_{i2}, V_{i3}, V_{i4})$	3.37
Maximum difference of variance:	$D_{i_{max}} = \max(D_{i1}, D_{i2}, D_{i3})$	3.38
Range of variance	$R_i = V_{i_{max}} - V_{i_{min}}$	3.39
Range of difference of variance:	$R_{D_i} = D_{i_{max}} - D_{i_{min}}$	3.40
Median variance:	$V_{i_{med}} = \text{median}(V_{i1}, V_{i2}, V_{i3}, V_{i4})$	3.41
Median difference of variance:	$D_{i_{med}} = \text{median}(D_{i1}, D_{i2}, D_{i3})$	3.42
Standard deviation of variance:	$\sigma_{V_i} = \sqrt{\frac{\sum_{j=1}^4 (V_{ij} - \bar{V}_i)^2}{4}}$	3.43
Standard deviation of difference of variance:	$\sigma_{D_i} = \sqrt{\frac{\sum_{j=1}^3 (D_{ij} - \bar{D}_i)^2}{3}}$	3.44

Mean absolute deviation of variance:	$\frac{\sum x_i - \bar{x} }{n}$	3.45
MAD of difference of variance:	$\frac{\sum x_i - \bar{x} }{n}$	3.46
Skewness of variance:	$\frac{\sum (x_i - \bar{x})^3}{n \cdot (\sum (x_i - \bar{x})^2)^{3/2}}$	3.47
Skewness of difference of variance:	$\frac{\sum (x_i - \bar{x})^3}{n \cdot (\sum (x_i - \bar{x})^2)^{3/2}}$	3.48

3- Standard deviation (s) and its difference vector (As): 3.49

$$\sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

3.50

Table 3 shows all extracted features using equations 3.49 and 3.50 based on our proposed feature extraction method SCLGM.

Table 3. The Extracted Features for Standard Deviation and its Difference Vector using SCLGM

Average of Standard deviation:	$\frac{\sum x_i}{n}$	3.51
Average of difference of Standard deviation:	$\frac{\sum x_i - \bar{x} }{n}$	3.52
Minimum Standard deviation:	$\min(x_i)$	3.53
Minimum difference of Standard deviation:	$\min(x_i - \bar{x})$	3.54
Maximum Standard deviation:	$\max(x_i)$	3.55
Maximum difference of Standard deviation:	$\max(x_i - \bar{x})$	3.56
Range of Standard deviation:	$\max(x_i) - \min(x_i)$	3.57
Range of difference of Standard deviation:	$\max(x_i - \bar{x}) - \min(x_i - \bar{x})$	3.58
Median Standard deviation:	$\text{median}(x_i)$	3.59
Median difference of Standard deviation:	$\text{median}(x_i - \bar{x})$	3.60

Standard deviation of Standard deviation:	$\sqrt{\frac{\sum (\dots)}{n}}$	3.61
Standard deviation of difference of Standard deviation:	$\sqrt{\frac{\sum (\dots)}{n}}$	3.62
MAD of Standard deviation:	$\frac{\sum \dots }{n}$	3.63
MAD of difference of Standard deviation:	$\frac{\sum \dots }{n}$	3.64
Skewness of Standard deviation:	$\frac{\sum (\dots)}{(\sum (\dots))^3}$	3.65
Skewness of difference of Standard deviation:	$\frac{\sum (\dots)}{(\sum (\dots))^3}$	3.66

4- **Mean absolute deviation (MAD) and its differences** MAD is a robust measure of the variability of gray level distribution around the mean, for each centroid line i we compute the MAD as follows:

$$\frac{\sum | \dots |}{n} \quad 3.67$$

The differences of MAD could be computed as follow:

$$\left(\frac{\sum | \dots |}{n} - \frac{\sum | \dots |}{n} \right) \quad 3.68$$

Table 4 shows all extracted features using equations 3.67 and 3.68 based on our proposed feature extraction method SCLGM.

Table 4. The Extracted Features for Mean Absolute Deviation and its Difference Vector using SCLGM

Average of MAD:	$\frac{\sum \dots }{n}$	3.69
Average of difference of MAD:	$\frac{\sum (\dots)}{n}$	3.70
Minimum of MAD:	$\min (\dots)$	3.71
Minimum of difference of MAD:	$\min (\dots)$	3.72
Maximum of MAD:	$\max (\dots)$	3.73

Maximum of difference of MAD:	$\max(f(x) - g(x))$	3.74
Range of MAD:	$\max(f(x) - g(x)) - \min(f(x) - g(x))$	3.75
Range of difference of MAD:	$\max(f(x) - g(x)) - \min(f(x) - g(x))$	3.76
Median of MAD:	$\text{median}(f(x) - g(x))$	3.77
Median difference of MAD:	$\text{median}(f(x) - g(x))$	3.78
Standard deviation of MAD:	$\sqrt{\frac{\sum (f(x) - g(x))^2}{n}}$	3.79
Standard deviation of difference of MAD:	$\sqrt{\frac{\sum (f(x) - g(x) - \text{median}(f(x) - g(x)))^2}{n}}$	3.80
Mean absolute deviation of MAD:	$\frac{\sum f(x) - g(x) }{n}$	3.81
Mean absolute deviation of difference of MAD:	$\frac{\sum f(x) - g(x) - \text{median}(f(x) - g(x)) }{n}$	3.82
Skewness of MAD:	$\frac{\sum (f(x) - g(x) - \text{median}(f(x) - g(x)))^3}{(\sum (f(x) - g(x) - \text{median}(f(x) - g(x)))^2)^{1.5}}$	3.83
Skewness of difference of MAD:	$\frac{\sum (f(x) - g(x) - \text{median}(f(x) - g(x)) - \text{median}(f(x) - g(x) - \text{median}(f(x) - g(x))))^3}{(\sum (f(x) - g(x) - \text{median}(f(x) - g(x)) - \text{median}(f(x) - g(x) - \text{median}(f(x) - g(x))))^2)^{1.5}}$	3.84

5- **Skewness (SK)**: skewness is a measure of the asymmetry of the gray level distribution, it is computed by the following formula:

$$\frac{\sum (x - \bar{x})^3}{(\sum (x - \bar{x})^2)^{1.5}} \quad 3.85$$

$$\frac{\sum (|f(x) - g(x)| - \text{median}(|f(x) - g(x)|))^3}{(\sum (|f(x) - g(x)| - \text{median}(|f(x) - g(x)|))^2)^{1.5}}$$

Skewness differences could be computed as follow:

$$\frac{\sum (| |f(x) - g(x)| - \text{median}(|f(x) - g(x)|) | - \text{median}(| |f(x) - g(x)| - \text{median}(|f(x) - g(x)|) |))^3}{(\sum (| |f(x) - g(x)| - \text{median}(|f(x) - g(x)|) | - \text{median}(| |f(x) - g(x)| - \text{median}(|f(x) - g(x)|) |))^2)^{1.5}} \quad 3.86$$

Table 5 shows all extracted features using equations 3.85 and 3.86 based on our proposed feature extraction method SCLGM.

Table 5. The Extracted Features for Mean Absolute Deviation and its Difference Vector using SCLGM

Average of SK:	Σ	3.87
Average of difference of SK:	$\frac{\Sigma}{\text{O}}$	3.88
Minimum of SK:		3.89
Minimum of difference of SK:	()	3.90
Maximum of SK:		3.91
Maximum of difference of SK:	()	3.92
Range of SK:		3.93
Range of difference of SK:		3.94
Median of SK:	()	3.95
Median of difference of SK:	(())	3.96
Standard deviation of SK:	$\Sigma ()$	3.97
Standard deviation of difference of SK:	$\frac{\Sigma}{\text{O}} () ()$	3.98
Mean absolute deviation of SK:	$\Sigma $	3.99
Mean absolute deviation of difference of SK:	$\frac{\Sigma}{\text{O}} () $	3.100
Skewness of SK:	$\frac{\Sigma ()}{(\Sigma)}$	3.101
Skewness of difference of SK:	$\frac{\Sigma (())}{(\Sigma (()))}$	3.102

6- **Kurtosis (K):** Kurtosis is a **measure** of whether the gray level distribution is peaked or flat relative to a normal distribution.

$$\frac{\Sigma (())^{\equiv}}{(\Sigma | (()))^{\equiv}} \quad 3.103$$

Kurtosis differences vector could be computed as follows:

$$K = \frac{1}{N} \sum_{i=1}^N (k_i - \bar{k})^2 \quad (3.104)$$

Table 6 shows all extracted features using equations 3.103 and 3.104 based on our proposed feature extraction method SCLGM.

Table 6. The Extracted Features for Mean Absolute Deviation and its Difference Vector using SCLGM

Average of Kurtosis:	$\frac{1}{N} \sum_{i=1}^N k_i$	3.105
Average of difference of Kurtosis:	$\frac{1}{N} \sum_{i=1}^N (k_i - \bar{k})$	3.106
Minimum of Kurtosis:	$\min(k_i)$	3.107
Minimum of difference of Kurtosis:	$\min(k_i - \bar{k})$	3.108
Maximum of Kurtosis:	$\max(k_i)$	3.109
Maximum of difference of Kurtosis:	$\max(k_i - \bar{k})$	3.110
Range of Kurtosis:	$\max(k_i) - \min(k_i)$	3.111
Range of difference of Kurtosis:	$\max(k_i - \bar{k}) - \min(k_i - \bar{k})$	3.112
Median of Kurtosis:	$\text{Median}(k_i)$	3.113
Median of difference of Kurtosis:	$\text{Median}(k_i - \bar{k})$	3.114
Standard deviation of Kurtosis:	$\sqrt{\frac{1}{N} \sum_{i=1}^N (k_i - \bar{k})^2}$	3.115
Standard deviation of difference of Kurtosis:	$\sqrt{\frac{1}{N} \sum_{i=1}^N (k_i - \bar{k})^2}$	3.116
Mean absolute deviation of Kurtosis:	$\frac{1}{N} \sum_{i=1}^N k_i - \bar{k} $	3.117
Mean absolute deviation of difference of Kurtosis	$\frac{1}{N} \sum_{i=1}^N k_i - \bar{k} $	3.118
Skewness of Kurtosis:	$\frac{\sum_{i=1}^N (k_i - \bar{k})^3}{(\sum_{i=1}^N (k_i - \bar{k})^2)^{3/2}}$	3.119
Skewness of difference of Kurtosis:	$\frac{\sum_{i=1}^N (k_i - \bar{k})^3}{(\sum_{i=1}^N (k_i - \bar{k})^2)^{3/2}}$	3.120
	$\frac{\sum_{i=1}^N (k_i - \bar{k})^3}{(\sum_{i=1}^N (k_i - \bar{k})^2)^{3/2}}$	


```

19.           ( )
20.           ( )
21. (         )
22.           ( )
23. (         )
24.           ( )
25. (         )
26.           ( )
27. end loop;
28. for each measure vector  $M_i$  do
29.   (         ) (         ) (         ) (         ) (         )
30.   (         ) (         ) (         ) (         ) (         )
31. end loop;
32.
    
```

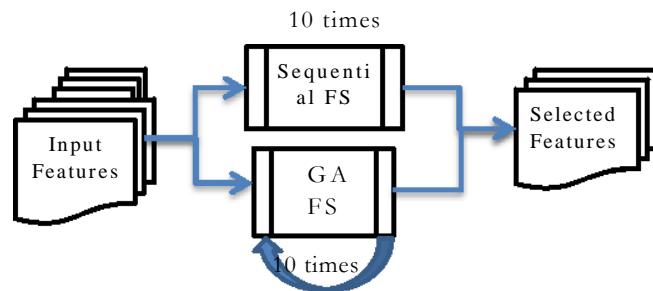


Figure 2. Feature Selection Technique

4. Experiments

To enable direct comparison with our previous work [21], we used the same datasets, methodology and implementation environment. Table 7 describes the used datasets.

Algorithm 2: The pseudo code of the proposed feature selection technique.

Purpose:

1. Selecting the best and discriminative features from the extracted set.

Input:

2. 145 extracted features.

Output:

3. The selected features are represented by the following sets: *SFF*, *GAF*, *USF*, and *ISF*.

Procedure:

```

4. begin
5. for  $i = 1$  to 10 do
6.           (         )
7.           (         )
8. end loop;
    
```

9 . ()
 1 0 . ()
 1 1 .
 1 2 .

Table 7. The Experiments are Done with Three Datasets

Dataset	# Cancers	# Benign	# Normal	Totals
Training dataset	126	105	179	410
Testing dataset-1	35	42	23	100
Testing dataset-2	60	14	26	100

Figure 3 shows three cases representing three types of mammogram: normal, benign and cancer, downloaded from DDSM.

As mentioned before, we extract 145 features, some of these features are obtained from DDSM database –human features- and the others are extracted mathematically.

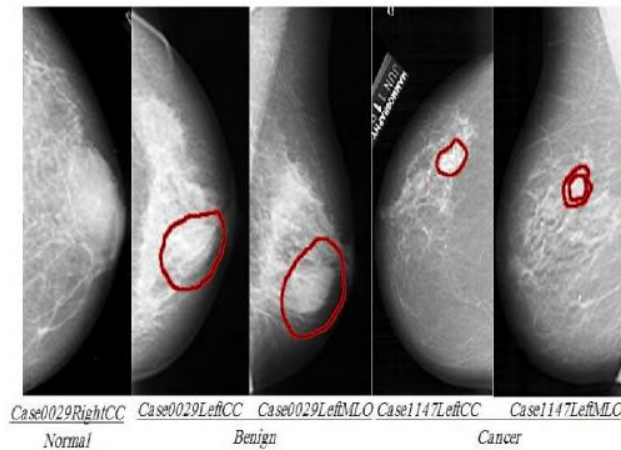


Figure 3. An Example for Each Normal, Benign and Cancer

For feature selection, we run forward sequential feature selection for ten times and we choose SVM classification algorithm -for its simplicity and high possibility of finding a global minimum- as a classifier in the criteria function of the forward sequential method. We use GA as a randomized feature selection method with the following parameters:

- The population size (P) which is the number of chromosomes in each generation; we use a random value for population size computed by the following Matlab formula:

$$P = \text{round}((L-1) * \text{rand}(DF, 200 * DF)) + 1; \quad 5.1$$

Where L is the number of input features, DF is the desired number of selected features; we use its value in range of 10 to 15.

- Max_Generations = 200, Max_Generations stands for the maximum number of generations and is used as a termination criterion. The GA will be stopped before the generated number of chromosome populations exceeds the Max_Generations, and the top scoring chromosome will be returned as the search's answer;
- The generations with no change in highest-scoring (elite), this parameter is used as a second termination criterion. It represents the number of generations that may pass with no changes in the elite chromosome before that elite chromosome will be returned to as the search's answer. In our method we set elite equal to one.

We apply GA methods 10 times for feature selection; the obtained selected features are ranked with respect to the number of occurrences of each feature in the 10 rounds and its criteria-or fitness- function. Table 8 shows the number of occurrences of each feature in the 10 rounds.

After feature selection we start classification phase using the Feed-Forward Artificial Neural Network (FFANN) with back-propagation, which is one of the most popular techniques, as a classifier. We use three levels to represent the forward neural network, the input layer with a number of neurons equal to the number of selected features, the output layer with a number of three nodes to represent the target classes –normal, benign and cancer-, and a hidden layer.

We use “Log-sigmoid” for transfer function, which is used to calculate the layer's output from its net input as follow:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (5.2)$$

To avoid falling in local minima and to made the result independent on the initial values, we use 10-fold cross validation.

Table 8. The Selected Features and the Number of Times it is Selected by Each Technique

	Forward Sequential	Genetic Algorithm
<i>x₁₄₅</i>	10	10
<i>x₁₄₃</i>	9	10
<i>x₁₃₅</i>	7	8
<i>x₇₆</i>	5	5
<i>x₁₄₂</i>	5	10
<i>x₅₉</i>	4	10
<i>x₆₄</i>	2	10
<i>x₁₀₃</i>	1	2
<i>x₁₂₆</i>	1	2
<i>x₄₃</i>	1	10
<i>x₁₂₈</i>	1	10
<i>x₃₃</i>	1	10
<i>x₉₆</i>	1	10

<i>x</i> ₁₂₀	2	2
<i>x</i> ₁₃₂	0	3
<i>x</i> ₁₄₄	0	9
<i>x</i> ₆₁	0	10
<i>x</i> ₃₁	0	10
<i>x</i> ₁₃₉	3	0
<i>x</i> ₁₃₈	2	0
<i>x</i> ₈₀	2	0
<i>x</i> ₁₁₄	1	0
<i>x</i> ₂₂	1	0
<i>x</i> ₆₉	1	0
<i>x</i> ₄₄	1	0
<i>x</i> ₆₀	1	0
<i>x</i> ₃₉	1	0
<i>x</i> ₁₆	1	0
<i>x</i> ₉₈	1	0
<i>x</i> ₈₄	1	0
<i>x</i> ₈₂	1	0

We implement the FFANN four times, each time we use one of the four sets that are obtained in feature selection stage, which are SFF, GAF, USF, and ISF. In the following we preview the classifications results using mentioned performance metrics:

1- Classification using SFF

The total number of the features selected by forward sequential technique is 27 features; we use these features as input to the FFANN with 29 neurons in the hidden layer and three neurons at the output layer.

Training stage: Figure 4 shows the ROC curve and the Confusion matrix of feed-forward ANN classifier in training stage. When we have a look on this figure we can judge that we have a very good classification with high performance as shown by the largest area under the ROC curve. The Confusion matrix shows that:

1-From 126 cancer cases, 105 are classified truly as cancer while the remained 21 are classified as benign,

2-From 105 benign cases, 89 are classified truly as benign, and 16 cases are classified in false as a cancer,

3- And 179 normal cases are classified truly as normal.

The overall classification accuracy is 91%, with 83.33% sensitivity and 84.7% specificity.

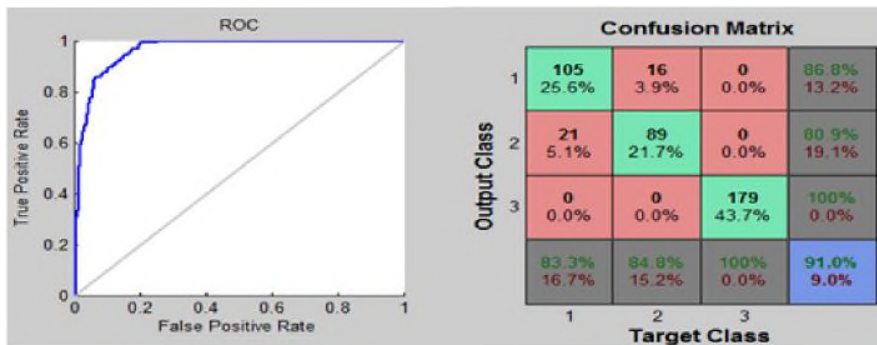


Figure 4. The ROC Curve and the Confusion Matrix Analysis of FFANN Classifier for Training using SFF Features

Test stage: As mentioned before, we use two different and difficult datasets for testing the model obtained in training stage. The difficulties of the used test datasets lead to decrease the accuracy of classification. With dataset-1 the accuracy is 82% as shown in the confusion matrix in Figure 5, and with dataset-2 the accuracy is 86% as shown in the confusion matrix in Figure 6.

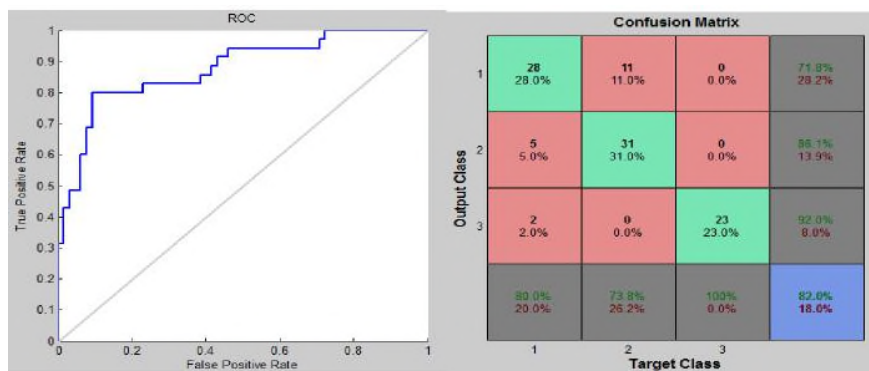


Figure 5. The ROC Curve Analysis and the Confusion Matrix of Testing Dataset-1 using SFF Features

2- Classification using GAF: GA method selects 18 features as the best for the classification; we use these features as input to the FFANN with 20 neurons in the hidden layer and three neurons at the output layer.

Training stage: The classification performance is demonstrated by the ROC curve and the Confusion matrix shown In Figure 7, as shown the area under ROC curve is large like that in Figure, which means that the performance here will be near that in case 1. The Confusion matrix, could be interpreted as follow:

1- From 126 cancer cases, 103 are classified truly as cancer while the remained 23 are classified as benign,

2- From 105 benign cases, 90 are classified truly as benign and 15 cases are classified in false as a cancer,

3- And 179 normal cases are classified truly as normal.

The overall classification accuracy is 90.7%, with 82% sensitivity and 85.7% specificity.

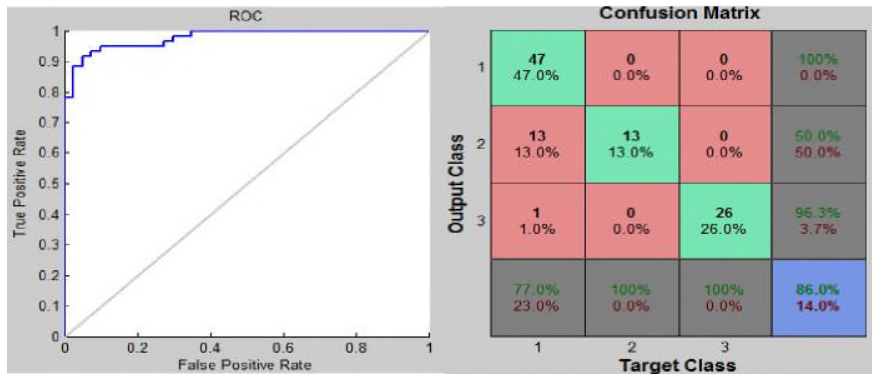


Figure 6. The ROC Curve Analysis and the Confusion Matrix of Testing Dataset-2 using SFF Features

Testing stage: After training we test the created model using two different and difficult datasets. As shown in Figure 8, the accuracy with dataset-1 is 84%. In cancer cases, 31 of 35 cases are classified truly as cancer, 2 as benign and 2 as normal. In dataset-2, the accuracy is 86%, which is better than that in dataset-1, see Figure 9.

3- Classification using ISF

Shared features are the features that are selected by both sequential and GA techniques; the number of these features is 13 features. We create a FFANN with 13 neurons for input, 15 neurons for the hidden layer and 3 neurons for output layer.

Training stage: Figure 10 shows the ROC curve analysis the confusion matrix for FFANN classifier with ISF features:

- 1- 112 of 126 cancer cases are classified truly as cancer while the remained 14 cases are classified as benign.
- 2- With 105 benign cases, 88 are classified truly as benign and 17 cases are classified in false as a cancer,
- 3- And 179 normal cases are classified truly as normal.

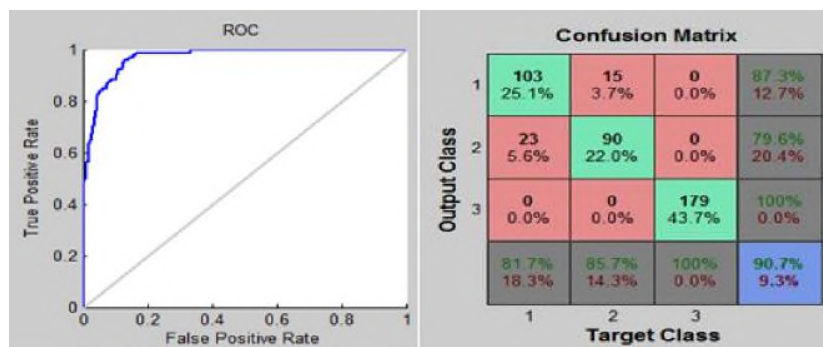


Figure 7. The ROC Curve Analysis and the Confusion Matrix of FFANN Classifier for Training using GAF Features

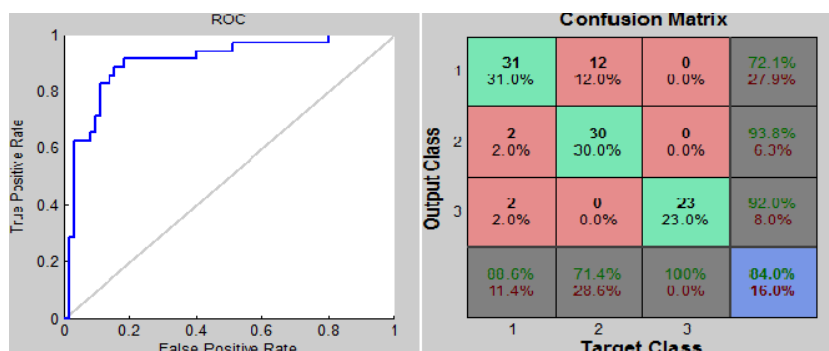


Figure 8. The ROC Curve Analysis and the Confusion Matrix of Testing Dataset-1 using GAF Features

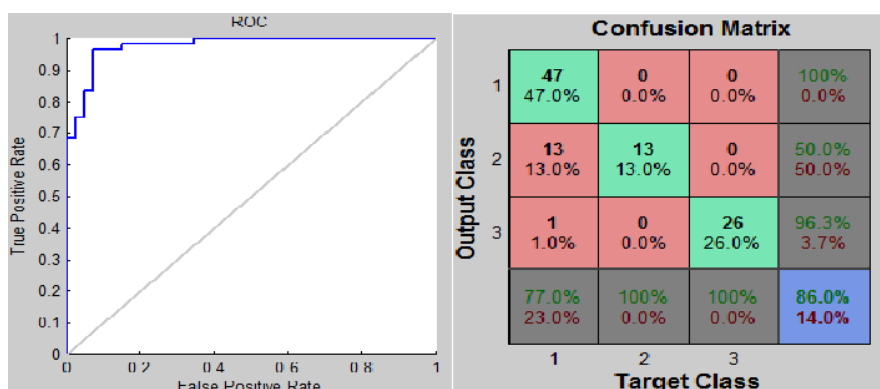


Figure 9. The ROC Curve Analysis and the Confusion Matrix of Testing Dataset-2 using GAF Features

The overall classification accuracy is 92.4%, with 88.9% sensitivity and 83.81% specificity. As shown the results is better than that obtained with either GAF or SFF.

Test stage: As shown in Figure 11, the accuracy with dataset-1 is 86%. In cancer cases, 31 of 35 cases are classified truly as cancer, 2 as benign and 2 as normal. In dataset-2, the accuracy is 89% as demonstrated in Figure 12. In general the accuracy of classification with testing datasets is better with ISF than with either SFF or GAF. But it is less than the accuracy that we obtained with training dataset.

4- Classification using USF

We have 31 features as a total number of the selected features by each of sequential and GA techniques. In the following we show the experiment results with training and testing data set using USF.

Training stage: Figure 13 shows the ROC curve of the classification using USF, the area under the curve is so small which means that we have the best accuracy. The total accuracy reaches 96.3% as shown in the “all” confusion matrix in Figure 13; in which we have:

- 1- 117 of 126 cancer cases are classified truly as cancer while the remained 9 cases are classified as benign,
- 2- 99 of 105 benign cases are classified truly as benign while 6 cases are classified in false as a cancer,
- 3- And 179 normal cases are classified truly as normal.

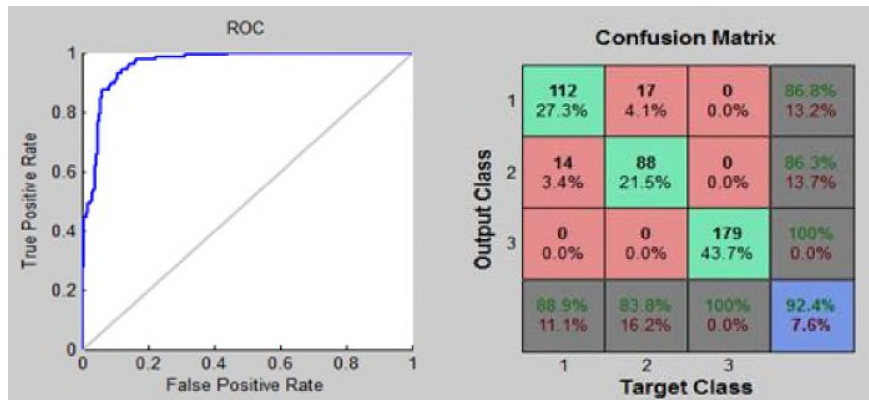


Figure 10. The ROC Curve Analysis and the Confusion Matrix of FFANN Classifier for Training using ISF Features

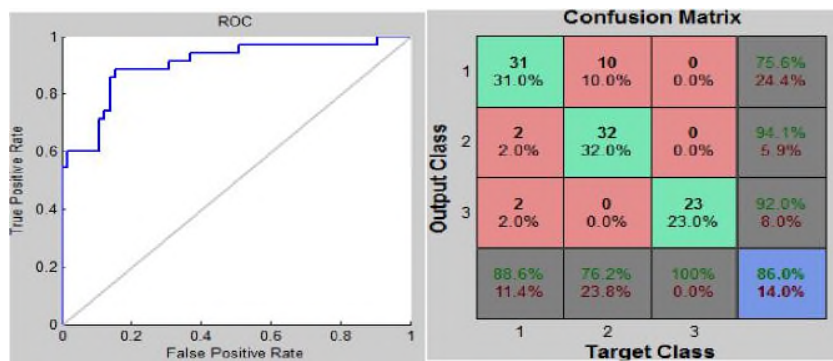


Figure 11. The ROC Curve Analysis and the Confusion Matrix of Testing Dataset-1 using ISF Features

So with union features we have the best performance, but it takes more time. The overall classification accuracy is 96.3%, with 92.9% sensitivity and 94.3% specificity.

Test stage: As shown in Figure 14, the accuracy of the classification with dataset-1 is 88%. In cancer cases, 31 of 35 cases are classified truly as cancer, 2 as benign and 2 as normal. Figure 15 shows the confusion matrix and ROC curve for the classification with USF, The accuracy of the classification is 89%, it is better than that with SFF, GAF, or ISF.

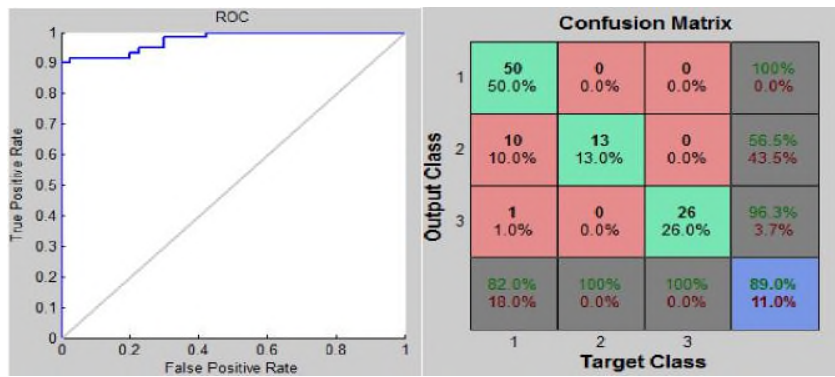


Figure 12. The ROC Curve Analysis and the Confusion Matrix of Testing Dataset-2 using ISF Features

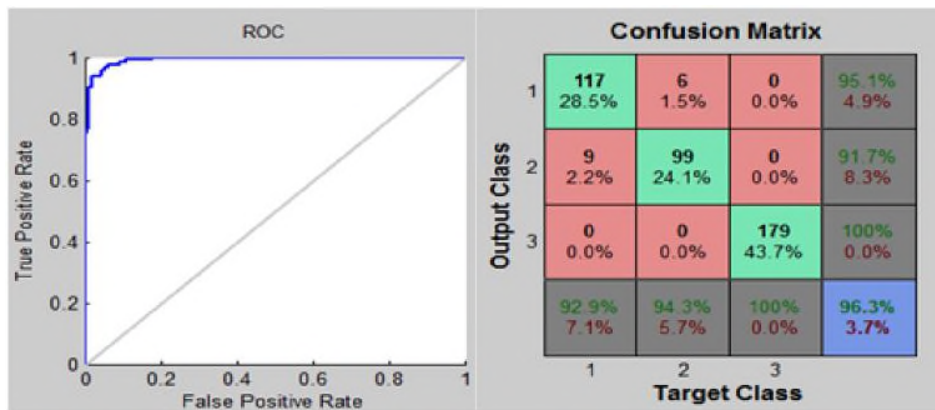


Figure 13. The ROC Curve Analysis and the Confusion Matrix for FF-ANN Classifier using USF Features

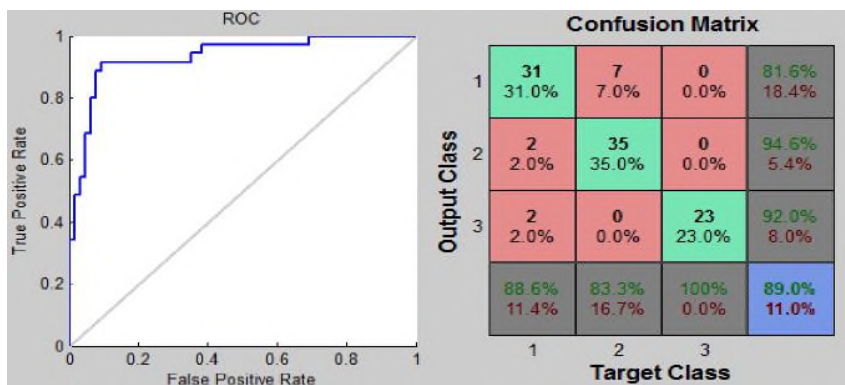


Figure 14. The ROC Curve Analysis and the Confusion Matrix of Testing Dataset-1 using USF Features

In, Table 9, we show a simple comparison in accuracy between our method and others. Actually, direct comparison of these systems is so difficult because most of these studies done on different databases and different cases.

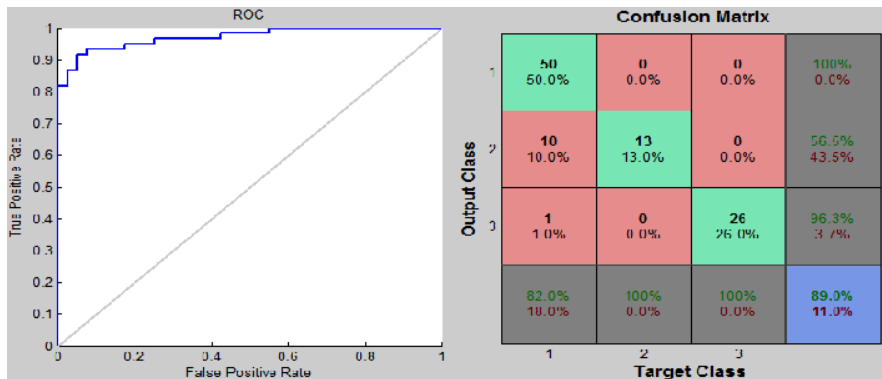


Figure 15. The ROC Curve Analysis and the Confusion Matrix of Testing Dataset-2 using USF Features

Table 9. A Simple Comparison in Accuracy between our Method and Others

Author	Training dataset	
	size	accuracy
Pohlman[8]	51	76%-93%
Hadjiiski [22]	348	81%
Jiang [7]	211	72.7%
Surendiran [27]	300	87%
Zheng [5]	3000	87%
Our previous work [21]	410	87% - 94.6%
Proposed method	410	89% - 96.3%

5. Conclusions and Future Work

We propose a new feature extraction method. We call it: Square Centroid Lines Gray level distribution Methods (SCLGM). By choosing four centroid lines we compute six mathematical measures, which we use to extract 79 features. To enable direct comparison with our previous work [21], we used the same datasets, methodology and implementation environment. The total number of the extracted features is 143, and by using feature selection techniques we select 31 features, 18 of the selected features are from our proposed feature extraction method (SCLGM). The Results were encouraging and if we compare the results with that in our previous work [21], we find that the accuracy enhanced by 2%.

The performance could be enhanced more and more by extracting and selecting the best and the most discriminative features so for future work we suggest the following:

- Here in our thesis we use gray level region grow methods for segmentation, and it does well, but not with ill-defined cases. So enhancing segmentation methods will be one of our future works.

- Enhancing our new proposed feature extraction technique SCLGM by finding more measures to extract more features will be one of our future works.
- We apply SCLGM on mammograms with mass types. In the future work we will apply it on the calcification cancer type, and it will extract a discriminative feature and will do best.
- After feature extraction we have to enhance the feature selection method, as shown in our experiments the selected features play the main role in enhancing the classifications performance.
- Finally, we should do comparison between different classifiers' methods.

References

- [1] Cancer facts and figures 2011, (2011), <http://www.cancer.org/Cancer/BreastCancer/DetailedGuide/breast-cancer-key-statistics>.
- [2] S. Theodoridis and K. Koutroumbas, "Pattern Recognition", 4th ed., Elsevier's Science & Technology, Ed. California, United States of America: Academic Press, (2009).
- [3] M. L. Giger, N. Karssemeijer and S. G. Armato, "Computer-aided diagnosis in medical imaging", IEEE, (2001), pp. 1205–1208.
- [4] C. J. Vyborny, M. L. Giger and R. M. Nishikawa, "Computer-aided detection and diagnosis of breast cancer", Radiol., (2000), pp. 725–740.
- [5] B. Zheng, *et al.*, "Interactive Computer-Aided Diagnosis of Breast Masses: Computerized Selection of Visually Similar Image Sets From a Reference Library", Acad Radio, vol. 14, (2007), pp. 917–927.
- [6] University of South Florida, "Digital Database for Screening Mammography (DDSM)", (2008), <http://marathon.csee.usf.edu/Mammography/Database.html>.
- [7] L. Jiang, E. Song, X. Xu, G. Ma and B. Zheng, "Automated Detection of Breast Mass Spiculation Levels and Evaluation of Scheme Performance", Acad Radiol, vol. 15, (2008), pp. 1534–1544.
- [8] S. Pohlman, K. A. Powell, N. A. Obuchowski, W. A. Chilcote and S. Grundfest-Bronlatowski, "Quantitative classification of breast tumors in digitized mammograms", Medical Physics, vol. 23, (1996), pp. 1337-1345.
- [9] L. Blot and R. Zwiggelaar, "Extracting background texture in mammographic images: a co-occurrence matrices based approach", in Proceedings of the 5th International Workshop on Digital Mammography, Toronto(Canada), (2000), pp. 142-148.
- [10] W. -M. Chen, *et al.*, "3-D Ultrasound Texture Classification Using Run Difference Matrix", World Federation for Ultrasound in Medicine & Biology, vol. 31, no. 6, (2005), pp. 763–770.
- [11] K. Aoki and M. Kudo, "Feature and Classifier Selection in Class Decision Trees", SSPR&SPR, (2008), pp. 562–571.
- [12] M. L. Giger, Z. Huo, M. A. Kupinski and C. J. Vyborny, "Computer-aided diagnosis in mammography", J.M. Fitzpatrick M. Sonka, (Eds.): SPIE Press, vol. 2, (2000).
- [13] H. D. Cheng, *et al.*, "Approaches for Automated Detection and Classification of Masses in Mammograms", Pattern Recognition, vol. 39, no. 4, (2006), pp. 646-668.
- [14] F. N. Che, M. C. Fairhurst, C. P. Wells and M. Hanson, "Evaluation of a two-stage model for detection of abnormalities in digital mammograms", in Proceedings of the 1996 IEE Colloquium on Digital Mammography, London, UK, (1996).
- [15] B. Sahiner, *et al.*, "Image feature selection by a genetic algorithm: application to classification of mass and normal breast tissue on mammograms", Med. Phys., no. 23, (1996), pp. 1671–1684.
- [16] B. Sahiner, H. P. Chan, N. Petrick, M. A. Helvie and M. M. Goodsitt, "Design of a high-sensitivity classifier based on a genetic algorithm: application to computer-aided diagnosis", Phys. Med. Biol., (1998), pp. 2853–2871.
- [17] B. Surendiran and A. Vadivel, "Feature Selection using Stepwise ANOVA Discriminant Analysis for Mammogram Mass Classification", International J. of Recent Trends in Engineering and Technology, vol. 3, no. 2, (2010), pp. 55-57.
- [18] D. Wei, *et al.*, "Classification of mass and normal breast tissue on digital mammograms: multiresolution texture analysis", Phys. Med., no. 22, (1995), pp. 1501–1513.
- [19] M. A. Kupinski and M. L. Giger, "Feature selection and classifiers for the computerized detection of mass lesions in digital mammography", in IEEE International Conference on Neural Networks, (1997), pp. 2460–2463.

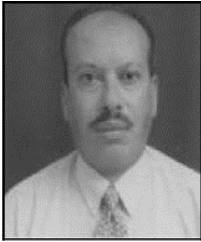
- [20] W. J. H. Veldkamp, N. Karssemeijer and J.D. M. Otten, "Automated classification of clustered microcalcifications into malignant and benign types", *Med. Physics*, no. 21, (2000), pp. 2600–2608.
- [21] B. K. Elfarrar and I. S. Abuhaiba, "Mammogram Computer Aided Diagnosis", *International Journal of Signal Processing, Image Processing and Pattern*, vol. 5, no. 4, (2012) December, pp. 1-30.
- [22] L. Hadjiiski, B. Sahiner, H. P. Chan, N. Petrick and M. Helvie, "Classification of malignant and benign masses based on hybrid ART2LDA approach", *IEEE Transactions on Medical Imaging*, vol. 18, (1999), pp. 1178-1187.

Authors



Belal K. Elfarrar

He is received his B.Sc. degree in Computer Engineering in 2000 from Islamic University of Gaza (IUG) in Palestine. In 1999 he had worked as a programmer developer at PdMAIN Company in Gaza. He is currently working in Islamic university as a system analyst. He works in this paper towards his MS at the department of Computer Engineering at the Islamic University of Gaza. He focuses on multi-agent system, artificial intelligence and data clustering analysis.



Ibrahim S. I. Abuhaiba

He is a professor at the Islamic University of Gaza, Computer Engineering Department. He obtained his Master of Philosophy and Doctorate of Philosophy from Britain in the field of document understanding and pattern recognition. His research interests include computer vision, image processing, document analysis and understanding, pattern recognition and artificial intelligence. Prof. Abuhaiba published tens of original contributions in these fields in well-reputed international journals and conferences.