Mammogram Computer Aided Diagnosis

Belal K. Elfarra[1] and Ibrahim S. I. Abuhaiba[2]

[1]Dept. of Information Systems, [2]Dept. Computer Engineering,
Islamic University of Gaza, Palestine

fbelal@iugaza.edu.ps, ihaiba@iugaza.edu.ps

Abstract

Computer-aided diagnosis (CADx) is used to help radiologists in interpretation mammograms and 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 enhance and introduce a new method for feature extraction and selection 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. For computational feature extraction, we enhance and use two pre-existed feature extraction methods, which are the Run Difference Method (RDM) and the Spatial Gray Level Dependence Method (SGLDM). Then, we evaluate and introduce a new method for feature selection by running both of forward sequential and genetic algorithm search methods individually. Later we evaluate the results. Experimental results are obtained from a data set of 410 images taken from DDSM for different types. Our method select 14 features from 65 extracted features. 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 94.6%, with 95.2% sensitivity and 84.8% specificity. In testing stage, our proposed method achieved an overall classification accuracy of 87%, with 88.6% sensitivity and 78.6% 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 [4].

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 [5, 6]. 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.

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

In segmentation step, the suspected objects are separated from background, this step is extremely important as the success of the classification algorithm depends on this stage [8]. The two major categories of segmentation methods are region growing and discrete contour models.

In [9], histogram based thresholding is used in segmentation step, where pixels with gray level values greater than 40 are retained while all others are set to zero, then statistical features are extracted from this region which signify the important texture features of breast tissue. These features are fed to the support vector machine (SVM) classifier to classify its tissue. By using threshold for segmentation in this method, some important information may be lost and other normal region will be added to the ROI such as nipple and lactation ducts. This addition will have bad effect on the results. A supervised method for the segmentation of masses in mammographic images using texture features was proposed in [10, 11], which presented a homogeneous behavior inside the selected region. The results of this method are very good in segmentation but it requires a very large data to represent all types of cancers with all its possible shapes and margins.

Edge-detection, which is a traditional method for image segmentation, is used in [12, 13]. One of the main disadvantages of this method is its dependence on the size of objects and sensitivity to noise, Further, since conventional boundary finding relies on changes in the grey level, rather than their actual values, it is less sensitive to changes in the grey scale distributions over images as against region based segmentation.

In [16], the lesion was segmented from the surrounding background using an adaptive region growing technique. Ninety-seven percent of the lesions were segmented using this approach.

In our experiments we use Region Growing technique; it is so suitable for mass segmentation [18].

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.

A new approach of texture classification of 3-D Ultrasound (US) breast diagnosis using run difference matrix (RDM) with neural networks is developed in [20]. 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 [21], 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 [22]. 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 [23]: Stepwise feature selection and Genetic algorithm (GA). Stepwise feature selection are used in [24, 25, 26].

In [27], 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 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 [28] 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 [25, 26, 29] 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 [26], 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 [30] 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.

Genetic algorithm (GA) has been studied in [24, 26, 7, 29]. By GA we create a population of solutions based on the chromosomes and evolve the solutions by applying genetic operators such as mutation and crossover to find best solution(s) based on the predefined fitness function. The GA method with different fitness functions can reduce a set of 340 features to 39–62 features [15].

In [33], the authors developed a classification method that used only morphologic features. These features were designed after an in-depth study of the clinical characteristics of calcifications and produced very impressive results.

Aoki [21] assume that each feature subset has different discriminative information for different classes; they proposed a class decision tree classifier using class dependent feature subsets.

In [34], authors classify mammogram images into normal image, benign image and malignant image. They used GLCM techniques to extract 26 features, and by a hybrid approach of feature selection they selected just 8 features, which are used by the decisions tree for mammogram classification. They applied their algorithms on just 113 mammogram image. Results are satisfactory but not highly guaranteed because the used dataset is so small and not enough. Another disadvantage in their work is about the extracted features which are so little, and this weakens the probability of selecting a discriminative features.

In our method we extract 65 features, and we use three datasets: the first dataset contains 410 mammogram images and is used for selecting features and creating CADx model and the others two datasets each of which contain 100 mammogram images and are used for testing the created model.

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

With computational feature extraction, we use and modify two of known methods, 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.

3.1.1 Spatial Gray Level Dependence Method (SGLDM): also called Co-occurrence matrix, it is a statistical method proposed by Julesz in 1975, who induced that: "no texture pair can be

discriminated if they agree in their second-order statistics" [19]. Then Julesz used the definition of the joint probability distributions of pairs of pixels for texture feature extraction.

The SGLDM matrix is formed by computing the number of occurrences of each pixel with gray level i that are away by distance d from any pixel with gray level j in a direction defined by angle θ. The choice of distance and angle combination, as well as the quantization level, is somewhat arbitrary. Figure 1 shows the co-occurrence for one pixel (yellow pixel) with d=3 pixel and $\theta \in \{0, \pi/4, 2\pi/4, 3\pi/4\}$.

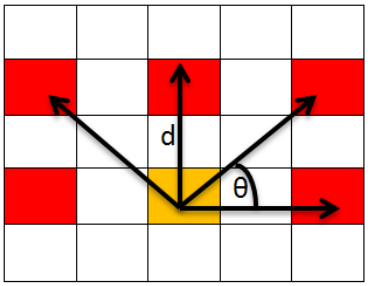

Figure 1. The Co-occurrence for One Pixel (yellow pixel) with d=3 pixel and θ ∈ {0, π/4, 2π/4, 3π/4}.

The choice of distance and angle combination, as well as the quantization level, is somewhat arbitrary [36]. Usually, the selection of parameter d is based on texture granularity, which means that with fine texture we use small values for d, and with coarse texture –as in malignant mass- large values of d will be used. In our method, we compute the co-occurrence matrix for a square ROI at eight different angles: $\theta \in \{0°, \pi/8, 2\pi/8, 3\pi/8, 4\pi/8, 5\pi/8, 6\pi/8,$ and $7\pi/8\}$, and at all distances $d \in \{1, 2, 3.., L/2\}$, where L is the length of ROI's side. We use these variant values of distances and angles to have discriminative features for the mass pattern to distinguish between benign or normal mass and malignant mass.

As a result, eight co-occurrence matrices are generated for each value of distance d; these matrices are summed to obtain a rotation invariant matrix Md:

$$M_{i,j}^{d} = \sum_{\theta} M_{i,j}^{\theta,d} \qquad 3.1$$

Then we extract the eight of 14 co-occurrence features proposed by Haralick [31]. These features are computed by equations that follow, where G denotes the number of gray levels, σ_i^d, σ_j^d and μ_i^d, μ_j^d denotes the standard deviations and the mean of the sum of co-occurrence matrix at distance value d.

1- **Correlation:**

$$x_1 = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} M_{i,j}^{d} \left[\frac{(i - \mu_i^d)(j - \mu_j^d)}{\sigma_i^d \sigma_j^d} \right] \qquad 3.2$$

2- **Homogeneity:**

$$x_2 = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \left[\frac{M_{i,j}^{d}}{1 + (i-j)^2} \right] \qquad 3.3$$

3- **Variance:**

$$x_3 = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} M_{i,j}^{d}(i-\mu^{d})^2 \quad 3.4$$

4- **Contrast:**

$$x_4 = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} M_{i,j}^{d}(i-j)^2 \quad 3.5$$

5- **Energy:**

$$x_5 = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} \left(M_{i,j}^{d}\right)^2 \quad 3.6$$

6- **Maximum probability:**

$$x_6 = \boldsymbol{max_{i,j}} M_{i,j}^{d} \quad 3.7$$

7- **Sum average:**

$$x_7 = \frac{1}{2}\sum_{i=0}^{G-1}\sum_{j=0}^{G-1} M_{i,j}^{d}(i+j) \quad 3.8$$

8- **Cluster prominence:**

$$x_8 = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} M_{i,j}^{d}\left(i-\mu_i^{d}+j-\mu_j^{d}\right)^4 \quad 3.9$$

If we apply these eight measures for each value of d ϵ [1, L/2], we have (8*L/2) features. This is very large data, most of which are redundant and not useful. So we use the following functions for each measure - co-occurrence feature – over all values of d, these functions are:

1- **Mean**:

$$x_{n+8} = \frac{1}{L/2}\sum_{d=1}^{L/2} x_n^{d}, \qquad for\ all\ n \in \{1,2,3,4,5,6,7,8\} \quad 3.10$$

2- Mean absolute deviation:

$$x_{n+16} = \frac{1}{L/2}\sum_{d=1}^{L/2} |x_n^{d} - x_{n+8}|, \qquad for\ all\ n \in \{1,2,3,4,5,6,7,8\} \quad 3.11$$

3- Minimum:

$$x_{n+24} = Min_{d=1}^{L/2} x_n^{d}, \qquad for\ all\ n \in \{1,2,3,4,5,6,7,8\} \quad 3.12$$

4- Maximum:

$$x_{n+32} = Max_{d=1}^{L/2} x_n^{d}, \qquad for\ all\ n \in \{1,2,3,4,5,6,7,8\} \quad 3.13$$

5- Variance:

$$x_{n+40} = \frac{1}{L/2}\sum_{d=1}^{L/2}(x_n^d - x_{n+8})^2, \qquad for\ all\ n \in \{1,2,3,4,5,6,7,8\} \tag{3.14}$$

Where x_{n+8} is the mean of feature x_n^d over all d ϵ {1, 2, .. L/2}.

6- Skewness:

$$x_{n+48} = \frac{\frac{1}{L/2}\sum_{d=1}^{L/2}(x_n^d - x_{n+8})^3}{\left(\frac{1}{L/2}\sum_{d=1}^{L/2}(x_n^d - x_{n+8})^2\right)^{\frac{3}{2}}}, \qquad for\ all\ n \in \{1,2,3,4,5,6,7,8\} \tag{3.15}$$

The pseudo code description of Spatial Gray Level Dependence Method is given in Algorithm 1.

In steps 7 – 10 of the algorithm, we create eight images, each of which represents the origin image of the mammogram rotated by θ ∈ {0°, π/8, 2π/8, 3π/8, 4π/8, 5π/8, 6π/8, and 7π/8}. The purpose of these steps is to create the eight SGLDM matrices in one direction (θ=0).

Algorithm 1: The Pseudo Code of SGLDM Technique

Purpose:

1. Extracting 56 features from the mammogram images.

Input:

2. Enhanced segmented objects with black background (I).
3. The used direction for co-occurrence θ = {0°, π/8, 2π/8, 3π/8, 4π/8, 5π/8, 6π/8, 7π/8}
4. Features To be Computed *FTC*={ correlation, homogeneity, sum average, cluster prominence, maximum probability, variance, energy, contrast}

Output:

5. The ***SGLDM*** extracted features vector of the input image I.

Procedure:

6. *begin*
7. *for each angle in θ do*
8. $create\ I(\theta), which\ is\ the\ image\ rotated\ by\ angle\ \theta.$
9. $I(\theta) = crop\ I(\theta)\ to\ get\ the\ inscribed\ squares\ within\ the\ I(\theta).$
10. *end loop;*
11. $L = image\ side;$
12. *for all distances d ∈ {1, 2, 3.., L/2} do*
13. *Cooc=0;//Cooc is the sum of co-occurrence matrices*
14. *for each angle in θ do*
15. $Cooc = Cooc + cooccurrence\ matrix\ of\ I(\theta)$
16. *end loop;*
17. *for i =1 to length(FTC)do*

18. $S_i(d) = computeFeatures(Cooc, FTC(i));$
19. *end loop;*
20. *end loop;*
21. *n= length(FTC);//n is the number of measures to be computed features statistics:*
22. *for i =1 to length(FTC)do*
23. $S_{i+n} = mean(S_i);$
24. $S_{i+2n} = MAD(S_i);$
25. $S_{i+3n} = minimum(S_i);$
26. $S_{i+4n} = maximum(S_i);$
27. $S_{i+5n} = variance(S_i);$
28. $S_{i+6n} = skewness(S_i);$
29. *end loop;*
30. $features = [S];$
31. *return features;*
32. *end*

In step 18, we extract the co-occurrence features from Cooc matrix which is the sum of the eight SGLDM matrices computed at step 15.

3.1.2 Run Difference Matrix (RDM): RDM is based on the estimation of the probability density function of the gray level differences in an image. The run difference matrix comprised the gray-level difference along with a distance between the pixels, when the displacement vector between two pixels is given [20].

With respect to Figure Figure 2, we can compute the displacement vector $\overline{D}$ using the following equation:

$$\overline{D} = [\Delta x, \Delta y] = [x_a, y_b] - [x_n, y_m] \quad 3.16$$

By using Figure 2, RDM could be defined as a total number of pixels pair in ROI with distance r and gray level difference g_{dif} with the given direction θ:

$$RDM(r, g_{dif}|\theta) = \#\{((x_a, y_b), (x_n, y_m)): ((x_a, y_b), (x_n, y_m)) \in ROI, |G(x_n, y_m) - G(x_a, y_b)| = g_{dif}\}/N \quad 3.17$$

Where G(x, y) is the gray level value of the pixel (x, y),

$$r = \sqrt{\Delta x^2 + \Delta y^2}, \qquad \theta = \tan^{-1}\left(\frac{\Delta y}{\Delta x}\right) \quad 3.18$$

And N is used for normalization and it is the total number of all pixels pairs in ROI:

$$N = \{\#((x_a, y_b), (x_n, y_m)): ((x_a, y_b), (x_n, y_m)) \in ROI\} \quad 3.19$$

In cancer cases, the shape and margins are non-uniform. So we propose to extend RDM matrices for eight different angles θ ∈ {0°, π/8, 2π/8, 3π/8, 4π/8, 5π/8, 6π/8, and 7π/8}.

As a result, eight RDM matrices are generated; these matrices are summed to obtain a rotation invariant matrix MRDM:

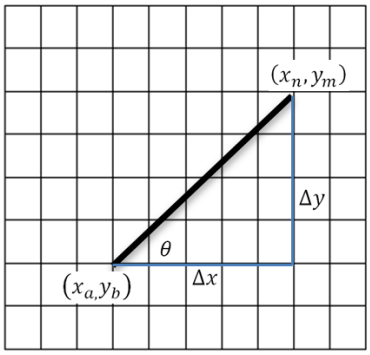

Figure 2. Run Difference Matrix (RDM) is a Function of r and Level Difference with the Given Direction θ

$$\mathrm{M}_{RDM} = \sum_{\theta} RDM(r, g_{dif}|\theta) \tag{3.20}$$

The obtained matrix MRDM is not designed for the extraction of features [20], but we can use the three characteristic vectors defined along with the original run difference matrix to extract features. These characteristics are the DGD (distribution of gray level difference), the DOD vector (distribution of average difference), and the DAD (distribution of average distance) vector:

$$DGD_{gdif} = \sum_{r=1}^{\frac{L}{2}} M_{RDM}, \tag{3.21}$$

$$DOD_r = \sum_{g_{dif}=0}^{G-1} M_{RDM}.g_{dif}, \tag{3.22}$$

$$DAD_{gdif} = \sum_{r=1}^{L/2} M_{RDM}.r \tag{3.23}$$

From these three characteristics we can extract the following five features:

1- Large difference emphasis (LDE), which measures the predominance of large gray level differences.

$$x_{57} = LDE = \sum_{gdif=0}^{G-1} DGD(g_{dif}).\ln(2/g_{dif}), \tag{3.24}$$

2- Sharpness, which measures the contrast and definition in an image.

$$x_{58} = \text{Sharpness} = \sum_{gdif=0}^{G-1} DGD(g_{dif}).(g_{dif})^3, \tag{3.25}$$

3- Second Moment of DGD (SMG), which measures the variation of gray level differences.

$$x_{59} = SMG = \sum_{gdif=0}^{G-1} \left(DGD(g_{dif})\right)^2, \tag{3.26}$$

4- Second Moment of DOD (SMO), which measures the variation of average gray level differences.

$$x_{60} = SMO = \sum_{r=1}^{L/2} (DOD(r))^2, \quad 3.27$$

5- Long distance emphasis for large difference (LDEL), which measures the prominence of large differences a long distance from each other.

$$x_{61} = LDEL = \sum_{gdif=0}^{G-1} DAD(g_{dif}).(g_{dif})^2. \quad 3.28$$

The pseudo code description of the Run Difference Matrix is given in Algorithm 2.

As in SGLDM algorithm, in steps 6-9 we create eight images of the origin ROI in order to extract features from each of which at one direction (θ-0).

In step 16 we create the RDM matrix, and then the measure DGD, DOD and DAD which are computed in steps 21, 22, and 23. From the computed measures, we can then extract the five features (LDE; Sharpness; SMG; SMO; LDEL) as shown in step 24.

Algorithm 2: The Pseudo Code of RDM Technique

Purpose:

1. Extracting 5 features from the mammogram images.

Input:

2. Enhanced segmented objects with black background (I).
3. The used direction for co-occurrence θ = {0°, π/8, 2π/8, 3π/8, 4π/8, 5π/8, 6π/8, 7π/8}.

Output:

4. The ***RDM*** extracted features vector of the input image I.

Procedure:

5. *begin*
6. *for each angle θ* do
7. $create\ I_\theta, which\ is\ the\ image\ rotated\ by\ angle\ \theta.$
8. $I_\theta = crop\ I_\theta\ to\ get\ the\ inscribed\ squares\ within\ the\ I_\theta.$
9. *end loop;*
10. $L = image\ side;$
11. *for each angle θ do*
12. *for distances d=1 to L/2 do*
13. *for height h=1 to L do*
14. *for pixel p=1 to (L-d) do*
15. $G_{dif} = |I_\theta(p,h) - I_\theta(p+d,h)| + 1;$
16. $RDM_\theta(d, G_{dif}) = RDM_\theta(d, G_{dif}) + 1;$
17. *end loop;*
18. *end loop;*
19. *end loop;*
20. *end loop;*
21. $DGD = computeDGD;$//see equation 3.21
22. $DOD = computeDOD;$//see equation 3.22
23. $DAD = computeDAD;$//see equation 3.23
24. *RDMfeatures* = [LDE; Sharpness; SMG; SMO; LDEL];
25. *return RDMfeatures*
26. *end*

3.1.3 Human Features: In addition to the computational extracting features, we use human extracting features, which are given by DDSM[14]. The interpretation description of these human features are based on Breast Imaging Reporting and Data System (BI-RADS) [3], which are used as a guide for standardizing mammographic reporting. In the following, we describe -in brief using BI-RADS- the used human extracted features:

$$x_{62} = patient\ age, \tag{3.29}$$

$$x_{63} = Density, \tag{3.30}$$

We use values between 1 and 4 to describe the breast density, the difficulties increased as the value increased. As shown in Table 1, density values are interpreted with respect to BI-RADS [3] definition.

Table 1. BI-RADS Interpretation for Breast Density Values

Density value	BI-RADS interpretation
1	The breast is almost entirely fat.
2	There are scattered fibroglandular densities that could obscure a lesion on mammography.
3	The breast is heterogeneously dense. This may lower the sensitivity of mammography.
4	The breast tissue is extremely dense, which lowers the sensitivity of mammography.

$$x_{64} = mass\ shape, \tag{3.31}$$

The mass shape is described by either one or more of the following five keywords:

Round: a mass that is spherical, ball-shaped, circular or globular in shape.

Oval: a mass that is elliptical or egg-shaped.

Lobulated: a mass that has contours with undulations.

Irregular: the lesion's shape cannot be characterized by any of the above.

Architectural distortion: The normal architecture is distorted with no definite mass visible.

$$and\ x_{65} = mass\ margins. \tag{3.32}$$

The mass margin is described by either one or more of the following five keywords:

Circumscribed: the margins are sharply demarcated with an abrupt transition between the lesion and the surrounding tissue. Without additional modifiers there is nothing to suggest infiltration.

Microlobulated: the margins undulate with short cycles producing small undulations.

Obscured: one which is hidden by superimposed or adjacent normal tissue and cannot be assessed any further.

Ill defined: The poor definition of the margins raises concern that there may be infiltration by the lesion and this is not likely due to superimposed normal breast tissue.

Spiculated: The lesion is characterized by lines radiating from the margins of a mass.

3.2 Feature Selection

The main goal of feature selection is to reduce the dimensionality by eliminating irrelevant features and selecting the best discriminative features. Many search methods are proposed for feature selection. These methods could be categorized into sequential or randomized feature selection. Sequential methods are simple and fast but they could not backtrack, which means that they are candidate to fall into local minima. The problem of local minima is solved in randomized feature selection methods, but with randomized methods it is difficult to choose proper parameters.

To avoid problems of local minima and choosing proper parameters, we propose a new technique for feature selection. We use two methods one classified as sequential and the others as randomized, in order to cover the disadvantages of each one by the advantages of the other. To this end, we use the sequential forward feature selection and the genetic algorithm (GA) feature selection. We use all extracted features as input for both sequential feature selection and GA feature selection individually.

In forward sequential method, we start with empty list of selected feature, and successively we add one useful feature to the list until no useful feature remains in the extracted input list. The selection of useful feature is based on misclassification error, which is defined as follows:

$$E = \frac{1}{N}\sum_{d=1}^{N} I(o_d \neq t_d) \qquad 3.33$$

Where I is the image to be classified (training example), N is the number of training examples, od and td are the output value and the target value for training example d.

We choose SVM classification method in criteria function, that SVM training always finds a global minimum.

In GA algorithm, we use a fitness function based on the principle of Max-Relevance and Min-Redundancy (mRMR) [32] to have a discriminant power feature with minimum redundancy.

As shown in Figure 3, we apply feature selection methods 10 times, each time the results are slightly changed from previous. Then, the results are combined and ordered with respect to both the number of occurrences and the accuracy values that are obtained by criteria or fitness function. Algorithm 3.4 describes the steps of feature selection.

As shown in Algorithm 3.4, the selected features are represented by four sets:

1- Sequential Forward Features (SFF): which are the features selected by Sequential Forward Selection technique.

2- Genetic Algorithm Features (GAF): which are the features selected by Genetic Algorithm technique.

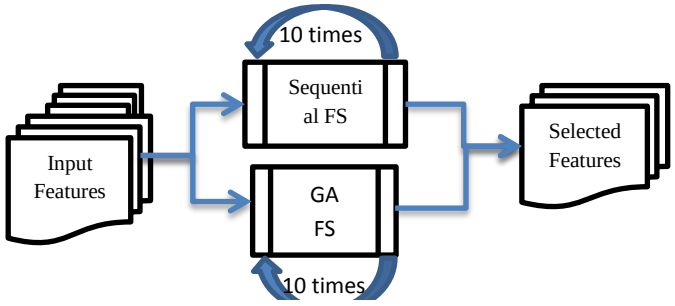

Figure 3. Feature Selection Technique

3- Union of Selected Features (USF): USF contains the union of features of SFF and GAF.

4- Intersection of Selected Features (ISF): represents the shared features between SFF and GAF.

Algorithm 3: The Pseudo Code of the Proposed Feature Selection Technique

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

Input:
2. The 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. $SFF = UNION\begin{pmatrix} SFF, Features\ selected\ by \\ Sequential\ Forward \end{pmatrix};$
7. $GAF = UNION(GAF, Features\ selected\ by\ GA);$
8. end loop;
9. $USF = UNION(SFF, GAF);$
10. $ISF = INTERSECT(SFF, GAF);$
11. $return\ SFF, GAF, USF, ISF;$
12. $end;$

4. Experiments

We download 610 mammographic images from DDSM for three datasets which are described in detail in Table 2.

Table 2. 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 4Figure 4 shows three cases representing three types of mammogram: normal, benign and cancer, downloaded from DDSM.

To avoid loss of image information, DDSM use the Lossless JPEG (LJPEG) image format – lossless means that there is no quality loss due to compression - to represent the mammography images.

LJPEG format is not readable by MATLAB, so we have to convert LJPEG mammographic images obtained from the DDSM into another lossless compression format that is readable by MATLAB. One of the most popular lossless compression image formats is TIF –Tag Image File format.

4.1 Implementation Environment

To convert LJPEG to TIF format we use heathusf software version 1.1.0 with gcc version 4.4.3. We run the conversion software using Oracle VM VirtualBox Version 4.0.8 r71778, and Ubuntu operating system version 11.04 - the Natty Narwhal.

The mammogram computer-aided diagnosis is implemented on windows 7 operating system using MATLAB R2011, with Matlab image processing tools and statistical tools. All experiments are implemented on a dell laptop of Intel Core 2 Due Processing power of 2.4 GHz CPU with 4GB RAM.

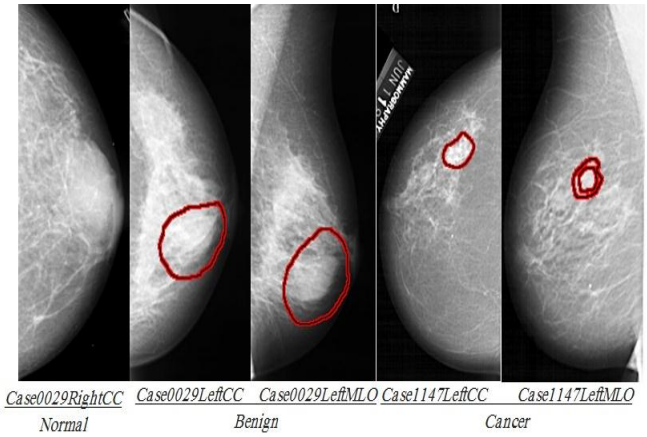

Figure 4. An Example for Each Normal, Benign and Cancer

4.2 Manual Segmentation

We apply segmentation in two stages, the first stage is done manually and before enhancement step, the second stage is done after enhancement using gray level growing algorithm. There are two objectives for using manual segmentation:

Decreasing the image size since TIF files of mammography images are generally pretty large, so this step leads to decrease the time required for both enhancement and auto segmentation steps. Decreasing the size of ROI for mass detection, this makes the task of auto-segmentation step easy.

As shown in Figure 4, the suspicious regions –benign or malignant- are marked by experienced radiologists. We use Adop-Photoshop to locate and cut the smallest square that contains the marked suspicious region. Figure 5 shows some examples of manual segmentation results.

4.3 Enhancement

It is important to do enhancement on mammogram images before applying auto segmentation for mass objects. The parenchyma density hides the mass objects and it will be so difficult to apply auto segmentation algorithms, so we use histogram equalization to enhance the mammogram images.

The main idea of histogram equalization is to use all available gray levels as shown in Figure 6Figure the histogram of original mammogram - case "C109LeftCC"- not using all available gray levels. To perform histogram equalization, we first create Probability Density Function (PDF) and the Cumulative Density Function (CDF) of mammogram image. Then

we use PDF as a match to the output enhancement image. The output is shown in Figure 6; it is clear that after enhancement all gray levels are used.

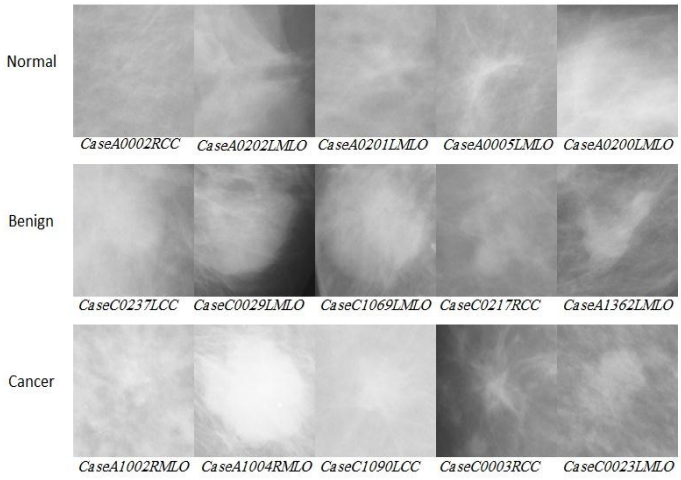

Figure 5. Some Examples of Manual Segmentation

Figure 7 shows some examples of mammogram images enhancements, as shown, after enhancements the mass margin are obvious and it became easy to be auto-segmented.

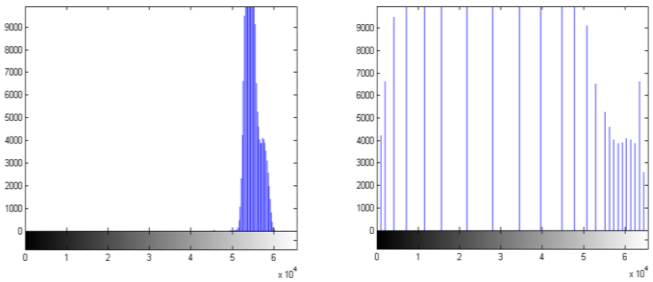

Figure 6. Histogram of Original Mammogram (left) and After Enhancement (right)

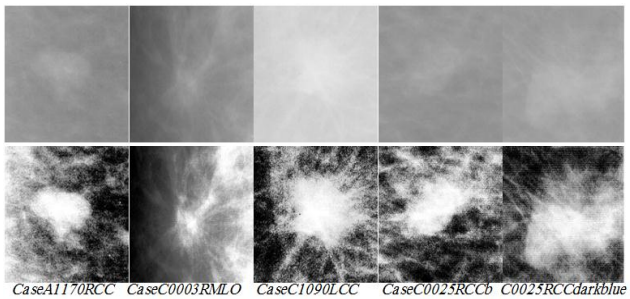

Figure 7. Mammogram Images Enhancements

4.4 Segmentation

After mammogram enhancement, it became easy to start auto segmentation for the suspicious regions, which has an observed contour. We use region grow algorithm based on

active contour segmentation. The basic idea in active contour models or snakes is to deform an initial curve so that it separates foreground from background based on the means of the two regions in order to detect objects in the processed image. For instance, starting with a curve around the object to be detected, the curve moves toward its interior normal and has to stop on the boundary of the object [2].

This technique is very robust and gives very nice results as there is a difference between the foreground and background means. The main problem we are faced with is in selecting the seed or the curve location, at which the algorithm starts deforming the curve until locating the mass as to be a segment. This problem is solved for a while by using the curve around the center of the suspicious region, but this solution is not good if we have multiple separated masses in one region. So, we use multiple seed, distributed on a different location in the mammogram image.

Another problem we are faced with is the number of iterations required for deforming the curve in order to locate the mass. We use large numbers for iterations with possibility to stop when no changes occurred on the curve shape.

As a result, we have for each mammogram a binary mask: zero for black and one for white. By multiplying the mask image with the origin image we obtain the segmented mass as shown in Figure Figure 8.

4.5 Feature Extraction and Selection

We extract 65 features, some of these features are obtained from DDSM database –human features- and the others are extracted mathematically.

The humanity features of some observations are shown in Table 3 as an example; the values of these features are described in Table 4.

For feature selection, 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.

In MATLAB, SVM functions support only two classes based on the basic algorithm of SVM which is designed as a binary classifier. So, we have here a problem that in our datasets we have three classes: cancer, benign and normal, so how could we extend a binary classifier SVM to multiclass? We use the idea demonstrated in Figure Figure 9. We divide the three classes into two classes: “Cancer” and “Not Cancer”.

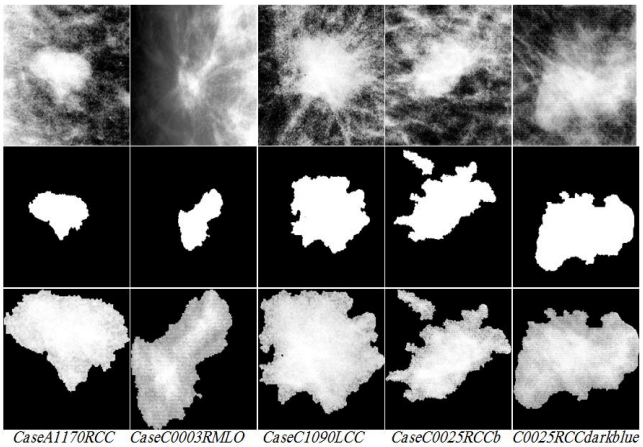

Figure 8. By Multiplying the Enhanced Image (first raw) with the Mask (second raw) we Obtain the Segmented Mass Shown in the Third Raw

If the decision of SVM choose the class "Cancer" then stop and store the result. Otherwise, we use the result as input to SVM classifier to classify it into "Benign" or "Normal".

Table 3. Humanity Features for Some Observations

		Humanity features			
Case	**Class**	**Age**	**Density**	**Shape**	**Margins**
C0235RCC	2	42	2	2	2
C0237LCC	2	62	3	2	1
C0029LCC	2	66	3	2	4
C0033RMLO	2	60	3	2	2
C0217RMLO	2	56	2	1	3
C1069LCC	1	70	2	1	5
A1357RMLO	2	76	4	2	4
A1361LCC	2	54	4	4	3
A1362LeftCC	2	58	4	4	4
A1363RMLO	2	43	4	2	1

Table 4. Class, Margin and Shape Values BI-RADS Description

Class	
Value	**Description**
1	CANCER
2	BENIGN
3	NORMAL
Margin	
Value	**Description**
1	CIRCUMSCRIBED
2	MICROLOBULATED
3	OBSCURED
4	ILL_DEFINED
5	SPICULATED
Shape	
Value	**Description**
1	ROUND
2	OVAL
3	LOBULATED
4	IRREGULAR
5	ARCHITECTURAL_DISTORTION

In our experiments, we use 10-fold cross-validation, by which the data set is divided into 10 subsets, one of them is used as a test and the remaining subsets are used for training. Repeatedly, we use a criterion function based on misclassification error for each training subset to train or fit a model, which is used next to predict the target of test subset.

The criterion function returns the sum of loss predicted targets of the test subset. Then we implement cross-validation calculation for a given candidate feature set by summing the values returned by criterion function and divide that sum by the total number of test observations. We use that mean value to evaluate each candidate feature subset.

We run forward sequential feature selection for ten times. As mentioned before we use GA as a randomized feature selection method, which prevents falling in local minima. We use 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:

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

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 5 shows the number of occurrences of each feature in the 10 rounds.

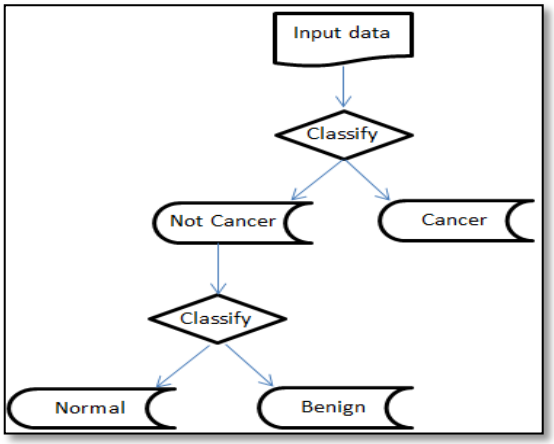

Figure 9. Use SVM for Three Classes

Table 5. The Selected Features and the Number of Times of it's Selection by Each Technique

	Forward Sequential	Genetic Algorithm
x_{65}	10	10
x_{63}	9	10
x_{59}	9	10
x_{62}	5	10
x_{61}	3	10
x_{43}	1	10
x_{18}	2	0
x_{10}	2	0
x_{17}	2	0
x_{32}	2	0
x_{16}	1	0
x_{31}	0	10
x_{29}	0	10
x_{33}	0	10

4.6 Classification

In our experiment we use 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.

In hidden layer, we faced the problem of what is the perfect number of nodes we have to use in the hidden layer? Too few nodes for hidden layer lead to under-fitting and too large number of nodes lead to over-fitting. There is no specific rule to determine the number of neurons -nodes- of hidden layer, so selecting the number of hidden layers neurons come down to trial and error. By trial and error we found that the best number of hidden layer neurons is the number of input + 2.

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

$$logsig(n) = 1/(1 + exp^{(-n)}) \quad 5.2$$

Since we are using a FFANN, which is a gradient descent search, it will always suffer from local minima and the result will heavily depend on the initial values.

We try to solve this problem and to get a better estimate of the performance by 10-fold cross validation. 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 11 features; we use these features as input to the FFANN with 13 neurons in the hidden layer and three neurons at the output layer.

Training stage: Figure 10 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: From 126 cancer cases, 112 are classified truly as cancer while the remained 14 are classified as benign, From 105 benign cases, 81 are classified truly as benign, and 24 cases are classified in false as a cancer, And 179 normal cases are classified truly as normal. The overall classification accuracy is 90.7%, with 88.9% sensitivity and 77.1% specificity.

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 83% as shown in the confusion matrix in Figure 11, and with dataset-2 the accuracy is 85% as shown in the confusion matrix in Figure 12.

2- Classification using GAF:

GA method selects 10 features as the best for the classification; we use these features as input to the FFANN with 12 neurons in the hidden layer and three neurons at the output layer.

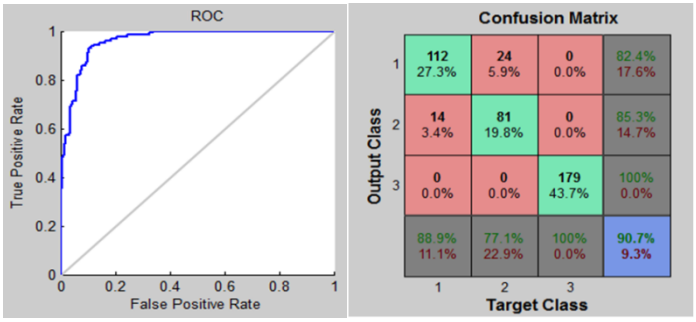

Figure 10. The ROC Curve and the Confusion Matrix Analysis of FFANN Classifier for Traning using SFF Features

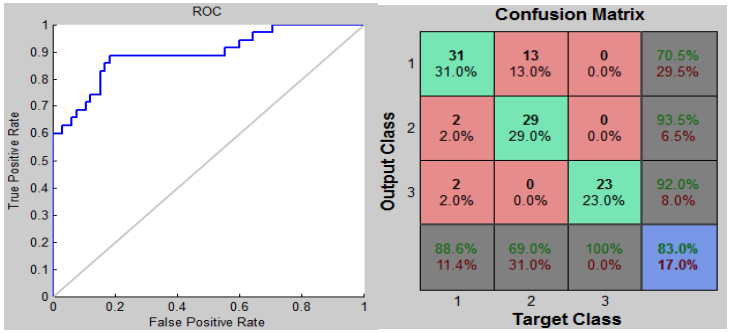

Figure 11. The ROC Curve Analysis and the Confusion Matrix of Testing Dataset-1 using SFF Features

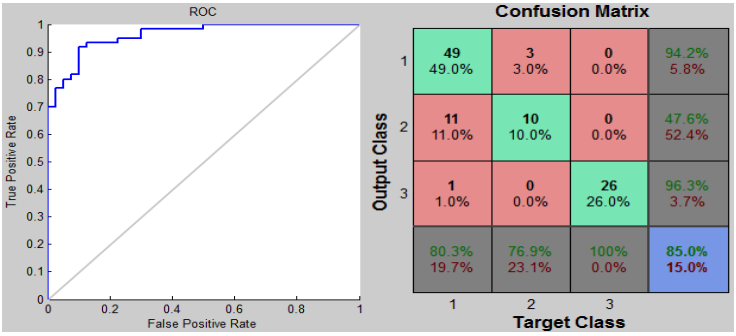

Figure 12. The ROC Curve Analysis and the Confusion Matrix of Testing Dataset-2 using SFF Features

Training stage: The classification performance is demonstrated by the ROC curve and the Confusion matrix shown In Figure 13, as shown the area under ROC curve is large like that in Figure 10, which means that the performance here will be near that in case 1. The Confusion matrix could be interpreted as follow: From 126 cancer cases, 110 are classified truly as cancer while the remained 16 are classified as benign, From 105 benign cases, 85 are classified truly as benign and 20 cases are classified in false as a cancer, And 179 normal cases are classified truly as normal. The overall classification accuracy is 91%, with 87.3% sensitivity and 81% specificity.

Testing stage: After training we test the created model using two different and difficult datasets. As shown in Figure 14, the accuracy with dataset-1 is 82%. In cancer cases, 30 of 35 cases are classified truly as cancer, 3 as benign and 2 as normal. In dataset-2, the accuracy is 88%, which is better than that in dataset-1, see Figure 15.

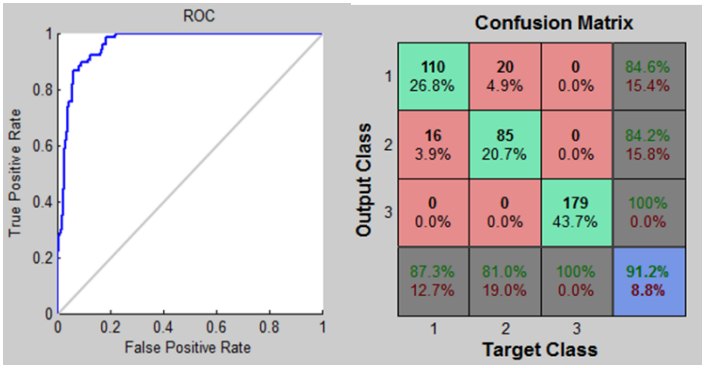

Figure 13. The ROC Curve Analysis and the Confusion Matrix of FFANN Classifier for Traning using GAF Features

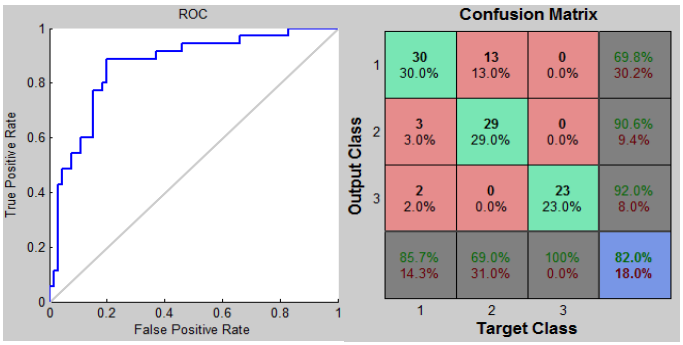

Figure 14. The ROC Curve Analysis and the Confusion Matrix of Testing dataset-1 using GAF Features

3- Classification using ISF

Shared features are the features that are selected by both sequential and GA techniques; the number of these features is 7 features. We create a FFANN with 7 neurons for input, 9 neurons for the hidden layer and 3 neurons for output layer.

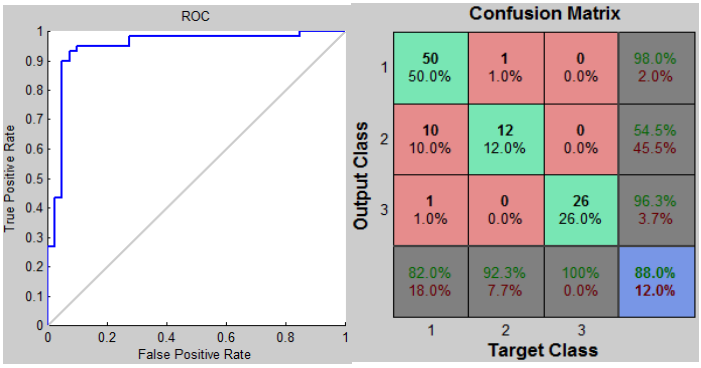

Figure 15. The ROC Curve Analysis and the Confusion Matrix of Testing dataset-2 using GAF Features

Training stage: Figure 16 shows the ROC curve analysis and the confusion matrix for FFANN classifier with ISF features: 112 of 126 cancer cases are classified truly as cancer while the remained 14 cases are classified as benign.

With 105 benign cases, 89 are classified truly as benign and 16 cases are classified in false as a cancer, And 179 normal cases are classified truly as normal.

The overall classification accuracy is 92.7%, with 88.9% sensitivity and 84.8% specificity. As shown the results is better than that obtained with either GAF or SFF.

Test stage: As shown in Figure 17, the accuracy with dataset-1 is 87%. 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 92% as demonstrated in Figure 18. 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 14 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.

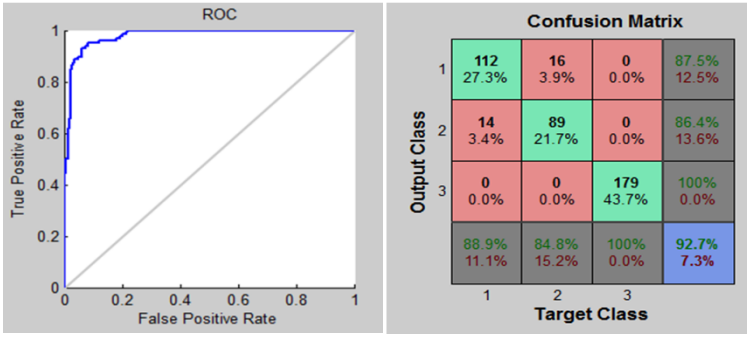

Figure 16. The ROC Curve Analysis and the Confusion Matrix of FFANN Classifier for Training using ISF Features

Training stage: Figure 19 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 94.6% as shown in the confusion matrix in Figure 4-16; in which we have: 120 of 126 cancer cases are classified truly as cancer while the remained 6 cases are classified as benign,

89 of 105 benign cases are classified truly as benign while 16 cases are classified in false as a cancer, And 179 normal cases are classified truly as normal.

So with union features we have the best performance. The overall classification accuracy is 94.6%, with 95.2% sensitivity and 84.8% specificity.

Test stage: As shown in Figure 20, the accuracy of the classification with dataset-1 is 87%. In cancer cases, 31 of 35 cases are classified truly as cancer, 2 as benign and 2 as normal.

Figure 21shows the confusion matrix and ROC curve for the classification with USF, The accuracy of the classification is 91%.

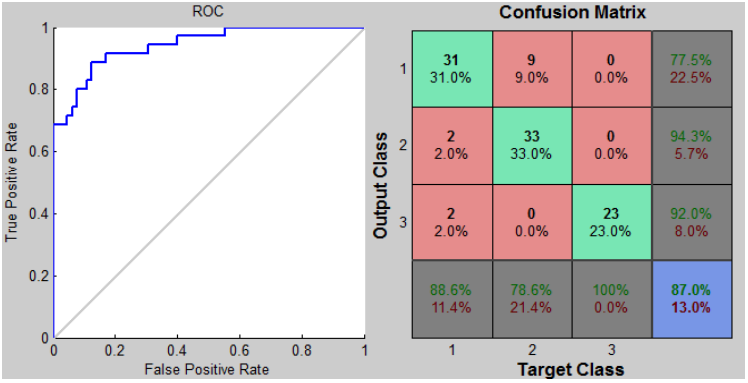

Figure 17. The ROC Curve Analysis and the Confusion Matrix of Testing Dataset-1 using ISF Features

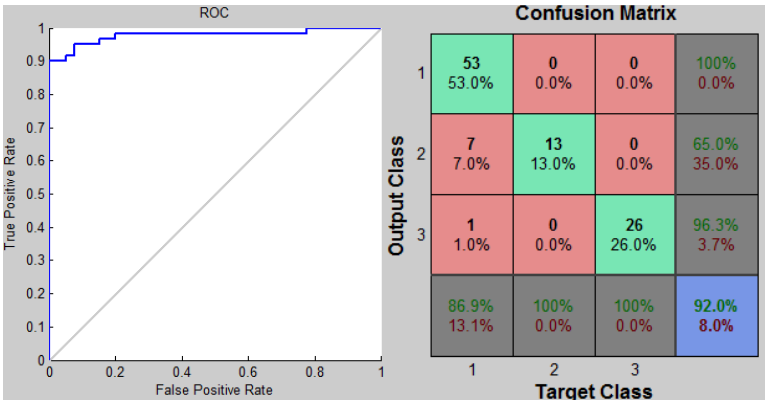

Figure 18. The ROC Curve Analysis and the Confusion Matrix of Testing Dataset-2 using ISF Features

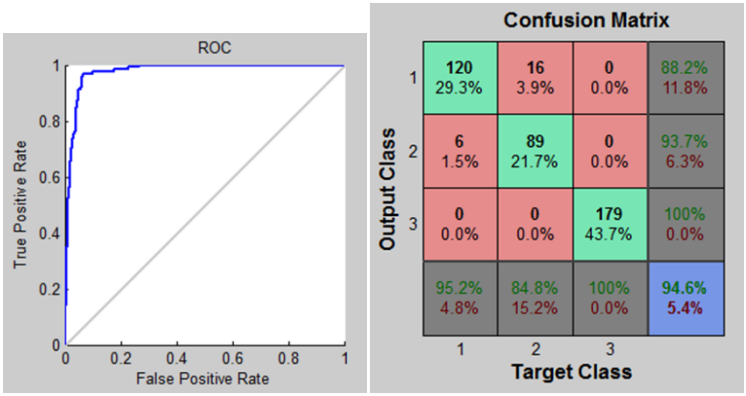

Figure 19. The ROC Curve Analysis and the Confusion Matrix for FF-ANN Classifier using USF Features

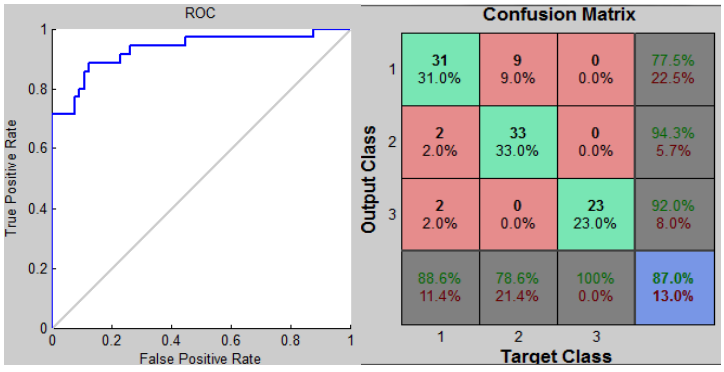

Figure 20. The ROC Curve Analysis and the Confusion Matrix of Testing Dataset-1 using USF Features

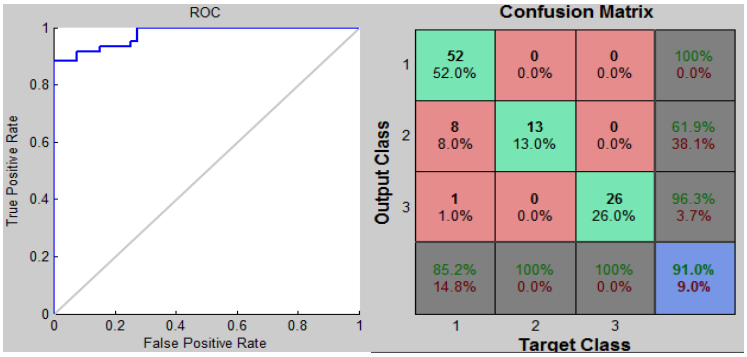

Figure 21. The ROC Curve Analysis and the Confusion Matrix of Testing dataset-2 using USF Features

4.7 Results Discussion

The best accuracy obtained with USF features; with training dataset the accuracy is 96%, but this accuracy decreased to 89% with test datasets. As mentioned before, the main reason of this difference is the difficulties with testing datasets. These difficulties could be described and demonstrated by Figure 22, which shows four misclassified images described as follow:

Image (a) which represents the case "C0125-Right-MLO" is a cancer case in its early stage with an oval shape and ill-defined margins. This image is classified in false as benign.

Image (b) represents the case "C0085-Right-CC" is a cancer case with irregular shape and spiculated margins. This image is classified in false as benign.

Image (c) is a cancer case represents "A1144-Right-MLO", it has an irregular shape and ill-defined margins, and is classified as benign.

Image (d) represents the case "C0029LeftMLO" which is a benign case with oval shape and ill-defined margins.

Images (a), (c) and (d) has ill-defined margins, in which we have unrestrained proliferation of cells in irregular directional pattern. In early stages it is so difficult to differentiate between benign and cancer and also it is so difficult to segment the mass in perfect. In our thesis, 14% of training dataset -56 cases- has ill-defined margins, 49 of them are classified truly. But in testing datasets, we have 50% of the cases with ill-defined margins; most of them are in early stage.

The difficulties of image (b) is in its spiculated margins, which make the segmentation task so difficult, and sure the classification will be influenced by the accuracy of segmentation results.

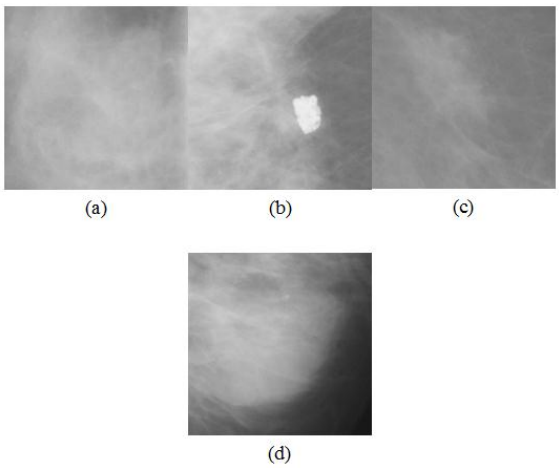

Figure 22. Examples of Misclassified Mammogram Images

In Table 6, 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.

Table 6. A Simple Comparison in Accuracy between our Method and Others

Author	Training dataset	
	size	accuracy
Pohlman [16]	51	76%-93%
Hadjiiski [35]	348	81%
Jiang [15]	211	72.7%
Surendiran [27]	300	87%
Zheng [11]	3000	87%
Proposed method	410	94.6%

5. Conclusions and Future Work

Computer-aided diagnosis (CADx) is a helpful system used to classify the detected regions into malignant or benign categories to help the radiologists in recognizing the next step, biopsy or short-term follow-up mammography.

The performances of different classifying algorithms can come to the same level in detecting mammogram lesions with the same features from the same mammographic database. So the performance of CADx depends more on the optimization of the feature extraction and selection than the classification methods. To obtain this, we concentrate on feature extraction and selection.

We use the Digital Database for Screening Mammography (DDSM) as a source of the mammographic images, which we use in our experiments. We do some enhancement on pre-existing feature extraction methods, which are GLCDM and RDM, by choosing eight different angles with variant pixel distances. The total number of extracted features is 65 features.

After feature extraction, it is important to apply feature selection to avoid the curse of dimensionality and to reduce the complexity of the CADx system. It is very difficult to predict which feature or feature combinations will achieve better in classification rate. In our

thesis we choose two feature selection methods: the first is categorized into the sequential search methods, which is the forward sequential feature selection technique, and the second method is categorized into the randomized search method, which is the Genetic feature selection Algorithm. These two methods select 14 features.

In classification, the best performance of all is obtained when we use all of the selected features by both sequential and GA techniques as input to the classifier.

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.
- 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, http://www.cancer.org/Cancer/BreastCancer/DetailedGuide/breast-cancer-key-statistics, **(2011)**.
[2] T. F. Chan and L. A. Vese, "Active Contours Without Edges", IEEE Transactions on Image Processing, vol. 10, no. 2, **(2001)**, pp. 266-277.
[3] American College of Radiology, "Breast imaging reporting and data system:BI-RADS atlas", American College of Radiology, Reston, atlas **(2003)**.
[4] S. Theodoridis and K. Koutroumbas, Pattern Recognition, 4th ed., Elsevier's Science &Technology, Ed. California, United States of America: Academic Press , **(2009)**.
[5] M. L. Giger, N. Karssemeijer and S. G. Armato, "Computer-aided diagnosis in medical imaging", IEEE, **(2001)**, pp. 1205–1208.
[6] C. J. Vyborny, M. L. Giger and R. M. Nishikawa, "Computer-aided detection and diagnosis of breast cancer", Radiol, **(2000)**, pp. 725–740.
[7] B. Zheng, Y. H. Chang, X. H. Wang and W. F. Good, "Comparison of artificial neural network and Bayesian belief network in a computer assisted diagnosis scheme for mammography", in IEEE International Conference on Neural Networks, **(1999)**, pp. 4181–4185.
[8] M. P. Sampat, M. K. Markey and A. C. Bovik, "Computer-Aided Detection and Diagnosis in Mammography", A.C. Bovik, Ed. Amsterdam: Elsevier Academic Press, **(2005)**.
[9] T. S. Subashini, V. Ramalingam and S. Palanivel, "Automated assessment of breast tissue density in digital mammograms", Elsevier, **(2010)**, pp. 33-43.
[10] J. Marti, J. Freixenet, X. Mu noz and A. Oliver, "Active region segmentation of mammographic masses based on texture, contour and shape features", in Springer-Verlag, Berlin Heidelberg, **(2003)**, pp. 478-485.
[11] 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.
[12] J. L. Viton, M. Rasigni, G. Rasigni and A. Liebaria, "Method for characterizing masses in digital mammograms", Opt. Eng., vol. 35, no. 12, **(1996)**, pp. 3453–3459.
[13] L. H. Li, W. Qian and L. P. Clarke, "Digital mammography: computer-assisted diagnosis method for mass detection with multiorientation and multiresolution wavelet transforms", Acad. Radiol, no. 4, **(1997)**, pp. 724–731.
[14] University of South Florida, "Digital Database for Screening Mammography (DDSM)", http://marathon.csee.usf.edu/Mammography/Database.html, **(2008)**.
[15] 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.
[16] 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.
[17] K. Bovis and S. Singh, "Detection of masses in mammograms using texture features", in 15th International Conference on Pattern Recognition (ICPR'00), **(2000)**, pp. 2:2267.

[18] I. K. Maitra, S. Nag and S. K. Bandyopadhyay, "Identification of Abnormal Masses in Digital Mammography Images", International Journal of Computer Graphics, vol. 3, no. 1, SERSC, **(2011)** May, pp. 17-29.
[19] B. Julesz, "Experiments in the Visual Perception of Texture", Scientific American, vol. 4, no. 232, **(1975)**, pp. 34-43.
[20] 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.
[21] K. Aoki and M. Kudo, "Feature and Classifier Selection in Class Decision Trees", SSPR&SPR, **(2008)**, pp. 562–571.
[22] M. L. Giger, Z. Huo, M. A. Kupinski and C. J. Vyborny, "Computer-aided diagnosis in mammography", J. M. Fitzpatrick M. Sonka, Ed.: SPIE Press, vol. 2, **(2000)**.
[23] 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.
[24] 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)**.
[25] 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.
[26] 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.
[27] 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.
[28] 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.
[29] 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.
[30] 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.
[31] R. Haralick, K. Shanmugam and I. Dinstein, "Textural features for image", SMC 3, **(1973)**, pp. 610–621.
[32] H. Peng, F. Long and C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 8, **(2005)**, pp. 1226-1238.
[33] M. Kallergi, "Computer-aided diagnosis of mammographic microcalcification clusters", Med. Physics, no. 31, **(2004)**, pp. 314–326.
[34] M. Vasantha, V. S. Bharathi and R. Dhamodharan, "Medical Image Feature, Extraction, Selection And Classification", International Journal of Engineering Science and Technology, vol. 2, no. 6, **(2010)**, pp. 2071-2076.
[35] 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.
[36] A. C. Silva, A. C. Paiva, P. C. P. Carvalho and M. Gattass, "Semivariogram and SGLDM Methods Comparison for the Diagnosis of Solitary Lung Nodule", Springer, **(2005)**, pp. 479–486.

Authors

Belal K. Elfarra

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.