Hybrid Model of Computer-Aided Breast Cancer Diagnosis from Digital Mammograms

Mugahed A. Al-antari[1], Mohammed A. Al-masni[1], Yasser M. Kadah[2]*

[1]Department of Biomedical Engineering, College of Electronics and Information, Kyung Hee University, Yongin, Republic of Korea
[2]Electrical and Computer Engineering Dept., King Abdulaziz University, Jeddah, Saudi Arabia

Abstract Computer-aided diagnosis (CAD) system is developed to assist radiologists to interpret digital mammographic images. The system learns the nature of different tissues in digital mammograms and uses this information to diagnose abnormalities. In this study, we develop a hybrid CAD system for digital mammograms combining several algorithms for selecting the significant features. The impact of quantization level of gray-level co-occurrence matrix(GLCM) on the performance of the system is analyzed. The proposed technique starts with peripheral equalization method that is a dedicated preprocessing technique for mammograms enhancement. Then regions of interest (ROIs) are excerpted by utilizing centered region of 32×32 pixels. A set of 422 quantitative attributes are extracted and normalized from each ROI. The features selection is performed using t-test, Kolmogorov-Smirnov test, Wilcoxon signed rank test, Sequential Backward, Sequential Forward, Sequential Floating Forward and Branch and Bound Selection algorithms. Voting k-Nearest Neighbor, Support Vector machine, Linear Discriminant Analysis, and Quadratic Discriminant Analysis classifiers are applied for CAD recognition. The proposed system is evaluated using quantitative metrics including sensitivity, specificity, positive predictive value, negative predictive value, overall accuracy, Cohen-k factor and area under ROC curves. The results show that the Sequential Forward algorithm offers potential for high performance with all classifiers especially with quantization level equal size of ROI.

Keywords Digital Mammography, computer-aided diagnosis, feature extraction, texture classification

Introduction

Uncontrolled growth of abnormal cells is called breast cancer which is one of the most common disease that spreads rapidly. It also considered as a second women fatal disease after lung cancer. The breast cancer usually starts inside the milk ducts. The genetic makeup and aggressiveness are two types of breast cancer [1]. There is no knowledge trend for preventing breast cancer arising. The detection in early stage allows taking enough health precaution before spreading. Detecting and diagnosing breast cancer is carried out by mammography machine which is safe and less harmful tool compared to other ways [2]. Mammography is an accurate tool for diagnosing which is important to avoid any further possible risks. American Cancer Society (ACS), American College of Radiology (ACR) and American Congress of Obstetricians and Gynecologists (ACOG) motivate women at age of 40 years to take annual mammograms [2-4]. For women ages 40 to 49 years, the National Cancer Institute (NCI) encourages them to take breast screening one or two times in a year [5].

CAD system provides the radiologists with a second reader opinion to make a better diagnosis of abnormalities. CAD system prompts the radiologist to review the suspicious regions in a mammogram by specialized computer algorithms [6]. The implementation of CAD system involves several techniques from image processing, statistics, physics, and mathematics. The objective of CAD is to enhance the overall accuracy and diagnostic

performance [7]. The result of CAD system is very helpful because a radiologist may miss lesions during diagnosis process such as microcalcifications and small masses in mammograms [8].

There are many previous studies that targeted to detect masses in the mammograms. Vállez *et al*. [9] improved a CAD system to reduce the false positives (FP) in breast density classification. They developed automated CAD system and compared many classification techniques. Also they proposed hierarchical classification with linear discriminant analysis (LDA) as a novel classifier. They used 1459 images from Mammographic Image Analysis Society (mini-MIAS) database and 298 features are extracted. The accuracy of their proposed CAD was 99.75%. Also the results showed 91.58% agreement when they used 1137 full-field digital mammograms (FFDM) dataset. Bueno *et al*. [10] developed a CAD system for automatic breast parenchymal density classification. They used many classifiers and applied them on screen-film mammography (SFM) and mini-MIAS databases to develop their CAD system. Their results reached to 84% as accuracy. In [11], Pohlman *et al*. presented a new technique to segment mass from the breast. The sensitivity for 51 dataset that used was 97%. Wei *et al*. [12] used Stepwise LDA to reduce dimension and select the features. They used wavelet transform to get these features. Oliver *et al*. [13] presented a technique to reduce FP in the mass detection. They used Principal Component Analysis (PCA) technique to get the features. For classification stage they used decision tree and k-Nearest Neighbor (KNN) together. By using Receiver Operating Characteristics (ROC) they evaluated their system. Also Akram *et al*. [14] used wavelet technique to obtain the detailed coefficients and extract the features from these coefficients to distinguish between normal and abnormal masses. For classification they used minimum distance and KNN classifiers independently. Mudigonda *et al*. [15] classified the mass region if it is a true or FP. They used new features that rely on flow direction in adaptive ribbons of pixels through the mass region. They calculated the features based on 2D histogram which called gray-level co-occurrence matrix (GLCM). They segmented and classified the mass as benign or malignant harmful disease. Dheeba *et al*. [16] built a CAD system to detect the breast cancer by using neural network (NN) classifier that was optimized by using wavelet. They extracted Laws texture attributes from breast lesions. They collected 216 mammograms (54 patients) from different centers of screening. Their result showed that the sensitivity and specificity were 94.167% and 92.105%, respectively. And the area under the ROC curve was 96.853%.

In this paper, hybrid CAD system for digital mammograms combining several algorithms for selecting the significant features is presented. Also, we explore the impact of quantization level of GLCM on the CAD system performance. The results from several classifiers to distinguish between different tissue abnormalities are presented and compared.
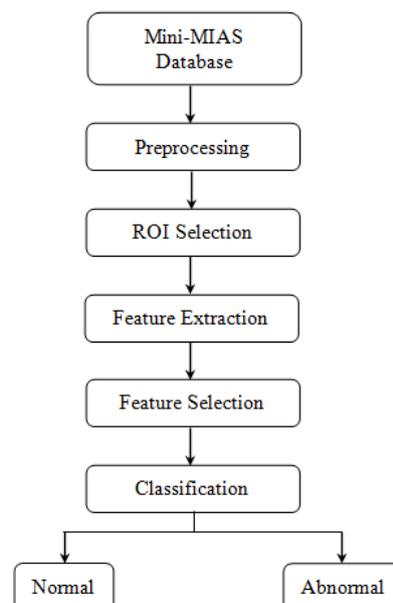
Materials and Methods

Figure 1: Block diagram of proposed CAD system

The proposed system consists of multiple stages to distinguish between the different tissue types. These stages include preprocessing, extraction of the Region of Interest (ROI), feature extraction, feature selection and classification stages. Figure 1 shows the proposed block diagram of CAD system.

Preprocessing

During the acquisition of mammogram data in all mammography machines, the whole breast is compressed in the specific tool in mammography machine. The deformation of breast will be happened due to this compression. The peripheral area of breast is affected by this compression which impacts on the grey level values of breast tissue at these regions [17]. The intensity values of peripheral area are always lower than central area. So, to diagnose an image correctly, physician must use certain settings of the window level during inspection on the suspicious regions. But this process may take long time especially with a huge number of patients and it is inconvenient at the same time.

Peripheral equalization (PE) method is a specific image processing algorithm improved for mammogram enhancement. It is used to enhance the ability to make both central and peripheral regions to be more visible with one window level settings [18]. Tao Wu *et al.* [19] technique is used in this study to enhance the peripheral area of mammogram. PE technique consists of five sequential stages. Segmentation of the breast region by using adaptive threshold, that calculated by utilizing Otsu thresholding, is the first stage as illustrated in Figure 2(b). The label of mammogram is omitted in this step as well. Then, 2D Gaussian low pass filter (GLPF) in frequency domain is applied to the original mammogram to get a blurred one (BI) as shown in Figure 2(c). After that, the BI is multiplied by the segmented image (SI) to eliminate the pixels that are placed outside the breast as depicted in Figure 2(d). The normalized thickness profile (NTP) of the mammogram is estimated as shown in Figure 2(e). The NTP is obtained as mean value from the BI after using five-threshold values T_n [19]. Each threshold value is computed as follows,

$$T_n = I_{ave} \times F_n \; ; \; n = 1,2,\dots 5, \tag{1}$$

where, I_{ave} is the average intensity of BI and F_n equals to 0.8,0.9,1.0,1.1 and 1.2 [19]. So, BI image is rescaled according to each threshold value as follows,

$$BI(i,j) = \begin{cases} \dfrac{BI(i,j)}{T_n} \; ; & BI(i,j) \leq T_n \\[2em] 1 \; ; & otherwise, \end{cases} \tag{2}$$

then, $\quad NTP = \dfrac{1}{V}\sum_{n=1}^{5} BI(n),$ (3)

where, $i = 1,2,3,\dots\dots,M$ and $J = 1,2,\dots\dots,N$ and $M \times N$ is the size of BI image and V is the normalized factor. Finally, the peripheral equalization (PE) of mammogram is achieved as the following formula,

$$PE = \frac{AI}{(NTP)^r} \, , \tag{4}$$

where, AI is an attenuation image (i.e. mammogram) that converted from x-ray projection. The peripheral equalization of mammogram is illustrated in Figure 2(f), and r is constant value where belongs to the range of [0.70 – 1.0] as in [19]. The ratio of Signal to noise (SNR) for both images in Figure 2(a) and (f), with take label in account, is computed. The SNR value at r = 0.7 and 1.0 is 0.4075 and 0.7537 dB for this data, respectively. So, in this study $r = 1.0$ for all dataset is verified and used.

ROI Extraction

The database used to train the CAD system in the mini-MIAS database [20]. This database contains 322 mammograms which are normal, benign and malignant tissues. 144 mammograms are used to accomplish this study with considering 72 are normal and others 72 are benign and malignant (i.e. abnormal). The abnormal mammograms, that are used, include different types of lesion such as circumscribed, speculated, ill-defined, architectural distortion and asymmetry. From each mammogram, ROIs are excerpted around the center of the mass with size of 32×32 pixels.

Feature Extraction

This stage is a key step in CAD system development because these features represent the texture of different tissue types and hence will affect directly on the system performance. The attributes are quantitative measures of texture that are used to explain the silent characteristics of the image texture. These features are extracted from each ROI. Three categories of features are used as described below to collect 422 features for achieving this study.
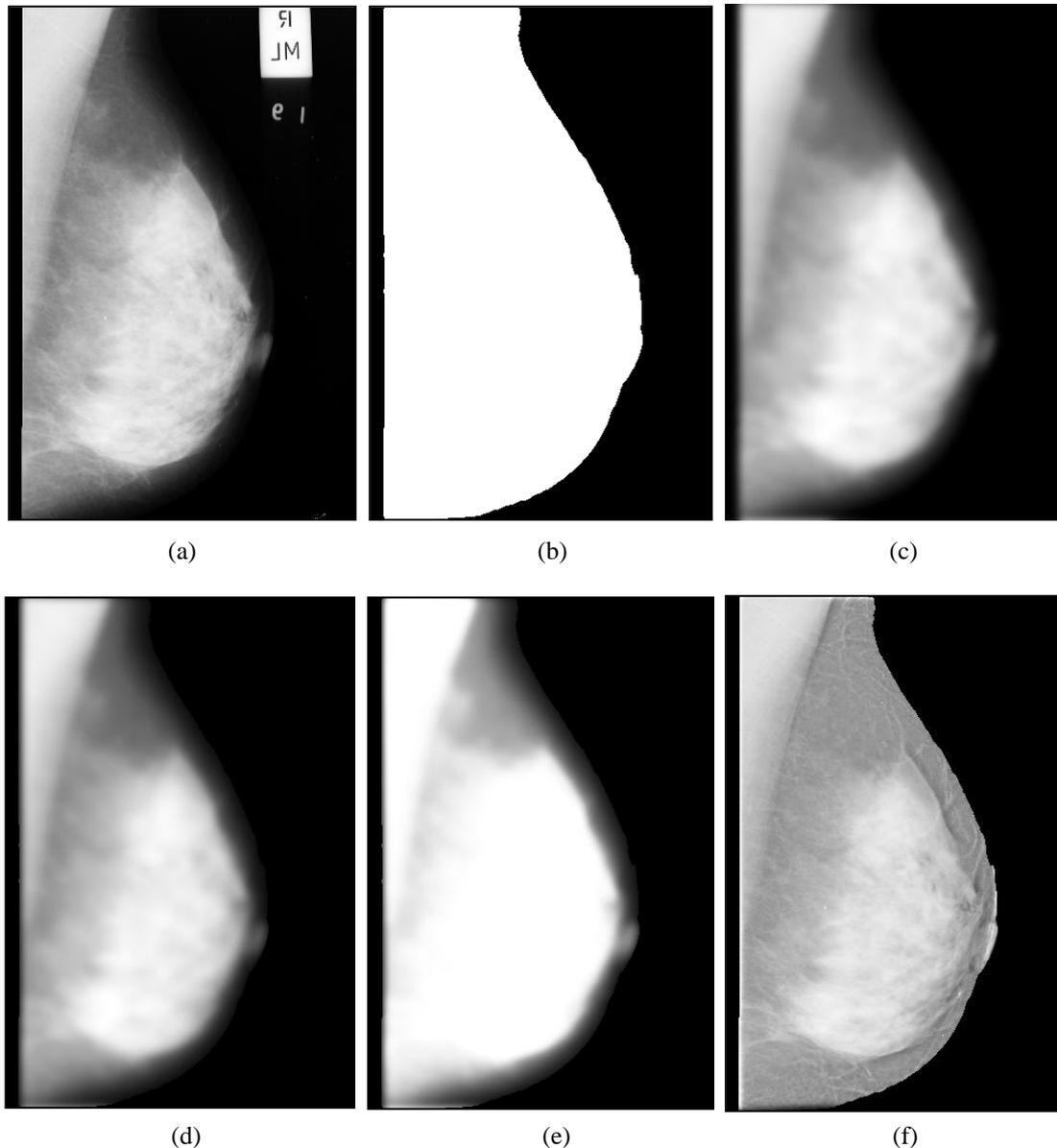

Figure 2: Peripheral density correction using Tao Wu et al. algorithm for mini-MIAS database (mdb004). (a) Original mammogram, (b) Segmented image (SI) with adaptive Otsu thresholding, (c) Blurred image (BI), (d) Blurred image after multiply by SI, (e) Normalized thickness profile (NTP) of mammogram and (f) The peripheral equalized (PE) of mammogram.

First order statistical feature

In this section, 28 features are extracted. Nine attributes are extracted from the ROI's histogram such as entropy, modified entropy, standard deviation (SD), modified standard deviation (MSD), energy, modified energy, asymmetry, modified skewness and range value. Other features are extracted from ROI directly such as mean,

SD, smoothness, 3rd moment, entropy, skewness, kurtosis, variance, mode, interquartile range, and Percentiles or quintiles at levels 0.1 to 0.9.

Higher Order Statistical Features

Higher order statistics features are very useful because they take into account the spatial inter-relationships of the pixels, as well as their gray level. Second order values are obtained by performing a statistical analysis on GLCM that proposed by Haralick *et al*. [21]. 2D histogram of gray level intensity for a pair of pixels is called GLCM. In this paper, we studied three important factors for GLCM. These factors are quantization gray level *L*, angle of orientation *θ* and displacement vector or distance value *d*, where (d = 1, 3, 5 and 9 with θ = 0°, 45°, 90° and 135°) are selected. Two GLCM sizes determined by selecting two different values of L at 8 and 32 are used. From each value of d we estimated four GLCMs, each one at a different θ. The extracted features from GLCM are energy, contrast, correlation, homogeneity, entropy, maximum probability, inverse different moment (IDM), variance, sum average, sum entropy, sum variance, difference entropy, difference variance, autocorrelation, dissimilarity, cluster shade, cluster prominence, correlation information #1 and correlation information #2. Thus, we collected 304 features from 16 different GLCMs at each quantization level.

Wavelet Transform Features

The wavelet transform estimates approximation, horizontal, vertical, and diagonal coefficients matrices LL, LH, HL, and HH, respectively from the input matrix (i.e. ROI) by using Daubechies (db1)[22]. Here, we use only one level of wavelet transform to obtain LH, HL and HH matrices and exclude LL matrix. From each coefficient matrix LH, HL or HH we compute two averaged GLCMs. First one is computed at d = 1 with θ = 0°, 45°, 90° and 135°. Second one is computed at d = 2 with θ = 0°, 45°, 90° and 135°. At each value of d we obtained four different GLCMs at different angles. After that we compute the averaged GLCM from these four GLCMs. Finally, six averaged GLCMs are collected. Thereafter, we extract some features from each averaged GLCM. These features are entropy, maximum probability, homogeneity, IDM, variance, uniformity, correlation information#1, Correlation information#2 and invariant moment (7 features). From wavelet transform section we extract or collect 90 features.

After extracting the feature set is carried out, rescaling them in the same range as [0, 1] or [−1, 1] is very important to get the powerful meaning for all of them. Selecting the target range depends on the nature of the data [23]. The general formula for normalization or scaling each feature is given as,

$$value_{scaled} = \frac{value - \min(value)}{\max(value) - \min(value)}, \tag{5}$$

where, $value$ is an original value of feature and $value_{scaled}$ is the normalized feature value. The scaling process is used to facilitate the coefficient values to avoid any statistical bias in classification stage is occurred.

Feature Selection

Many features may contain redundant information which affect the classifier performance. So, the reducing of the extracted features dimension and selecting the more powerful features is the main goal of this section. The performance of CAD system depends critically on the selected features. All extracted features are used as the input to the selection methods. Seven selection methods are used in this study. T-test, Kolmogorov and Smirnov (KS-test) and Wilcoxon signed rank (W-test) algorithms are used by Matlab Statistics Toolbox [24-25]. Sequential Backward (SBS), Sequential Forward (SFS), Sequential Floating Forward (SFFS) and Branch and Bound Selection (BBS) algorithms are also used by another Matlab Toolbox called PRTools4 [26]. The value of significance level selected to be 0.05 for all statistical selection methods. The most powerful selected features depend on how the selection method is good enough to determine these features. KS-test and W-test selection methods have exactly similar features. So, the performance of both will be the same as we will see in result and discussion part.

Classification Stage

The final stage of the proposed CAD system is classification phase which uses pattern recognition techniques to distinguish between the breast tissues. The most powerful selected features pass through CAD system to the

classification stage. Voting k-Nearest Neighbour (KNN) at K = 1, 3 and 5, Support Vector Machine (SVM), Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA) are used to accomplish this approach.

Evaluation of CAD System

There are several metrics or indices are used to evaluate our proposed CAD system performance. These indices are sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), overall accuracy and Cohen-k factor. Confusion Matrix or contingency table for two different classes is used to obtain all of these metrics. Table 1 reports the definitions with mathematical formulas of these indices.

Table 1: Metrics definition with their mathematical formulas

Index	Definition	Formula	
Sensitivity	Capability to measure the disease presence	$\dfrac{TP}{TP + FN}$	(6)
Specificity	Capability to measure the disease absence	$\dfrac{TN}{TN + FP}$	(7)
Positive Predictive Value (PPV)	Reliability of the positive result	$\dfrac{TP}{TP + FP}$	(8)
Negative Predictive Value (NPV)	Reliability of the negative result	$\dfrac{TN}{TN + FN}$	(9)
Overall accuracy	Global reliability	$\dfrac{TP}{TP + TN + FP + FN}$	(10)

where, TP, TN, FP and FN indicate true positive, true negative, false positive and false negative, respectively. Receiver operator characteristic (ROC) curve with its AUC to evaluate our proposed system is used too. This curve is created by graphical plotting, trapezoidal numerical integration for curve data fitting, between sensitivity and 1-specificity for the different possible cut-points of a diagnostic test. Cohen-k factor is a quantitative measurement that evaluates the CAD system performance. It's a statistical measure of intra- and inter observer agreement for qualitative items. This factor is estimated by using confusion matrix too as follows,

$$Cohen\text{-}k = \frac{P_o - P_a}{1 - P_a} \tag{11}$$

where, P_o and P_a are overall and expected agreement, respectively. These variables are calculated as follows,

$$P_o = \frac{TP + TN}{Total} \tag{12}$$

$$P_a = \frac{PEF + NEF}{Total} \tag{13}$$

where, $Total = TP + TN + FP + FN$, PEF and NEF are positive and negative expected frequency. In general, Kohen-k factor varies in the range of [0, 1]. Total absence of agreement between the observers (i.e. radiologist and CAD system) refer to 0 and the perfect agreement refer to 1 [27].

Result and Discussion

The work in this study is divided into two sections. Quantization levels (L) of GLCM equal to 8 and 32 are used for both sections. In each section, an integrated CAD system is achieved with all evaluation parameters such as positive predictive value (PPV), negative predictive value (NPV), sensitivity or true positive rate, specificity or true negative rate, overall accuracy, Cohen-k factor and ROC curves with their AUCs. Confusion Matrix for two classes is used to get all of these indices. Then the comparison between each classifier performance with each selection method is accomplished independently.

The performance metrics of CAD system with both values of quantization level (L) are reported independently for each selection method in Table 2 to Table 7. Table 2 reports the evaluated metrics of CAD system

performance with t-test. The SVM classifier at L=8 has the highest performance where AUC and Cohen-k factor represent as 95.77% and 91.67%, respectively. KNN with K=1 at L=32 has better performance than that when L=8. But KNN (K=3 and 5) has the same performance with both value of L where the overall accuracy and Cohen-k factor are 93.06%, 90.28% and 86.11%, 80.56%, respectively. As known the AUC is affected directly by both sensitivity and specificity which control the shape of ROC curve. All classifier results with t-test method present high sensitivity when L=8 compared with L=32 except KNN at K=1. The performance of most classifiers sound to be better at L=8 except KNN at K=1 which is better when L=32.

The performance of all classifiers for both KS-test and W-test is similar as listed in Table 3. Herein, KNN (K=3), SVM with L=8 and SVM when L=32 have the highest performance where the overall accuracy and Cohen-k factor equal to 94.44% and 88.89%, respectively. Due to the different values of sensitivity and specificity of classifiers, we obtained the different shapes of ROC curves. This means that AUC values vary corresponding to trade-off between these indices. Then, AUC represents as 95.73%, 94.14% and 94.32% with the same highest performance, KNN (K=3) and SVM when L=8 and SVM (L=32), respectively. So, the performance when L=8 of most classifiers seem also to be better except KNN (K=1) too.

From previous comparison with statistical methods, we can summarize that the performance of all classifiers seem to be better with L=8 except KNN when K=1. In general, the performance of statistical selection methods is extremely similar with all different classifiers.

The performance of all classifiers with SBS method is presented in Table 4. We can clearly observe that the overall accuracy of KNN (K=1, 3 and 5) is increased from 97.22% 95.83% and 93.06% when L=8 to 98.61%, 97.22% and 95.83% when L=32, respectively. With both values of L, the accuracy and Cohen-k factor of SVM are not changed, while ROC curves have different shapes and AUC values corresponding to the sensitivity and specificity values. These indices are 88.89% and 100% with L=8 and 94.44% when L=32. On the other hand, the performance of LDA and QDA classifiers at L=8 is better than those when L=32, where the accuracy is decreased from 97.22% for both to 95.83% for LDA and 93.06% for QDA, as well as Cohen-k factor is decreased from 94.44% for both of them to 91.67% for LDA and 86.11% for QDA. KNN (K=1) has the highest performance at L=32 with AUC equal to 98.89%. In this method, at L=32 QDA classifier has the same performance as KNN (K=3) with t-test method for all indices. Also SVM (L=32) has the same performance with SVM (L=8) in KS-test and W-test methods. The performance of all classifiers demonstrate better performance at L=32 except LDA and QDA which are better when L=8.

The performance of all classifiers is increased corresponding to the increasing value of L with SFS method as reported in Table 5. KNN (K=3 at L=8) and KNN (K=1 and 3 at L=32) has the optimal performance. On the other side, QDA classifier has better performance when L=8 where overall accuracy and AUC are equal to 97.22% and 94.44% respectively. SFS selection method sounds to be more robust to select the most powerful features that represent the silent texture characteristic of breast tissue. So, we can get high performance of all classifiers, during CAD system development, with this technique.

Also, classifier performance is increased from 97.22%, 97.22%, 95.83%, 97.22% and 94.44% when L=8 to become 98.61%, 100%, 98.61%, 98.61%, and 95.83% when L=32, respectively with SFFS method as demonstrated in Table 6. But QDA has better performance when L =8. Also we can consider the SFFS method to be encouraging choice during CAD system development to obtain the highest performance as in KNN (K=3 at L=32). At L=32 in this method, KNN (K=1 and 5), KNN (K=5) with SFS method and KNN (K=1) with SBS method have the similar performance with all metrics.

In BBS method, QDA classifier become has a better performance that is increased from 96.03% to 95.83%, when L=32 as reported in Table 7. In this method, at L=32 KNN (K=5) has the same performance comparing with KNN (K=5) with SBS method and LDA with SFS and SFFS methods as well.

The comparison between all classifiers with all selection methods, for both values of L by using Cohen-k factors and overall accuracy is also achieved as it's depicted in Figure 3 and Figure 4, respectively. The SFS method has the best behavior with both value of L, but when L=32 its result is better as it's obviously illustrates in these figures. Also, as in these figures demonstration, the results present better performance of all classifiers with SBS, SFS, SFFS and BBS than those in statistical selection methods. For the statistical selection methods, the performance of most classifiers is better at L = 8 except with KNN when k=1. All classifiers have better

performance at L=32 with PR-Toolbox methods. But BBS method has slightly different manner, comparing with other PR-Toolbox algorithms, especially with KNN and QDA classifiers where its performance is better when L=8. On the other hand, it has better performance when L=32 with SVM and LDA classifiers. In general, with all selection methods QDA classifier always provides better performance at L = 8 except with BBS method which is better with L=32. And LDA classifier mostly has the consistency behaviour especially when L=8 with SFS, SFFS and BBS methods with all metrics. This comparison by using overall accuracy and Cohen-k factor is similar in statistical quantitative measurements but it's different in the scientific concept as we mentioned before.
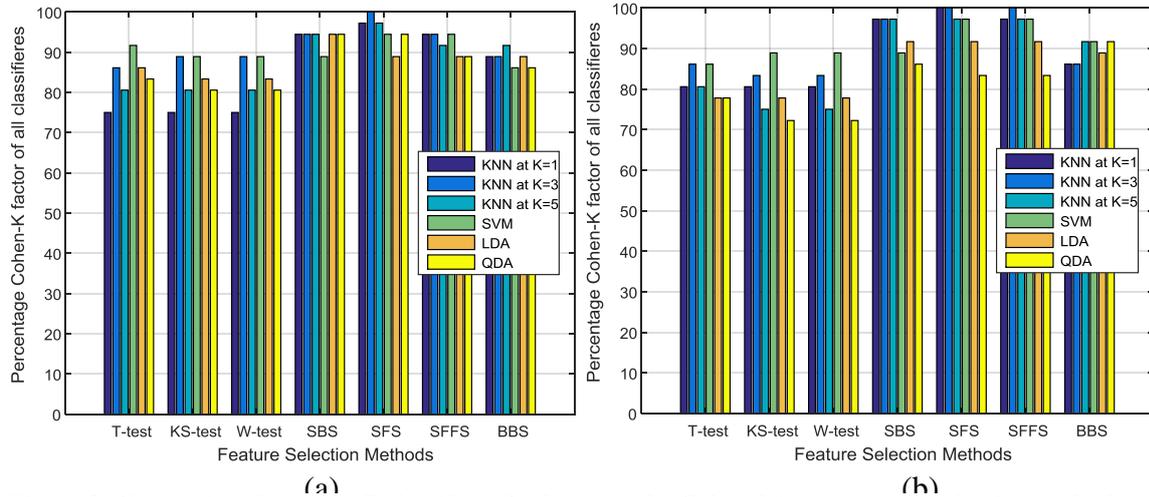

Figure 3: Comparison between all classifiers Performance by Cohen-k factor at each selection method with L = 8 (a) and L = 32 (b).
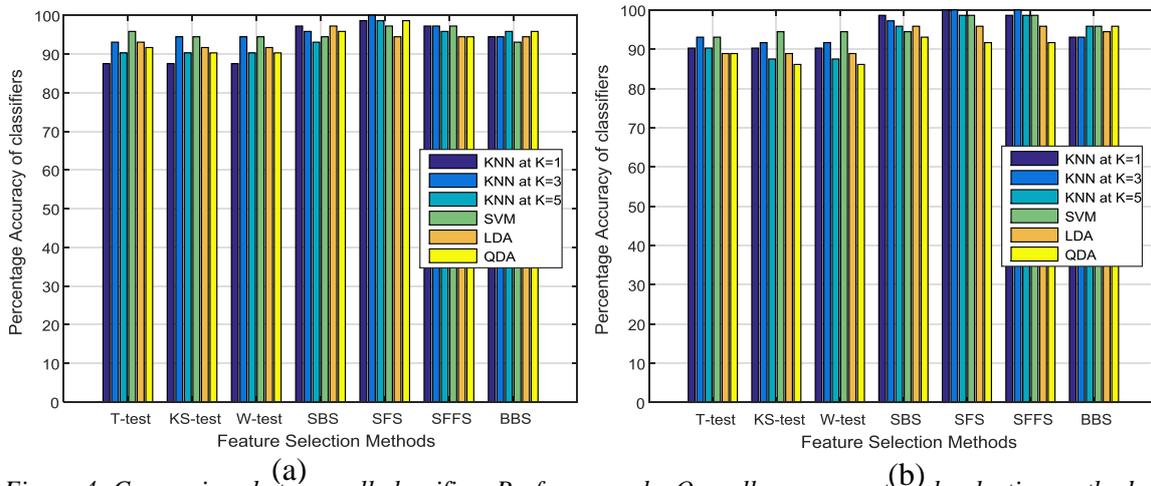
Figure 4: Comparison between all classifiers Performance by Overall accuracy at each selection method with L = 8 (a) and L = 32 (b).

Table 2: Indices for CAD system performance with all classifiers by using t-test

Indices (%)	L=8			SVM	LDA	QDA	L=32			SVM	LDA	QDA
	KNN						KNN					
	K=1	K=3	K= 5				K=1	K=3	K= 5			
Sensitivity	94.44	97.22	97.22	94.44	97.22	97.22	97.22	91.67	91.67	94.44	91.67	91.67
Specificity	80.55	88.89	83.33	97.22	88.89	86.11	83.33	94.44	88.89	91.67	86.11	86.11
PPV	82.92	98.74	85.37	97.15	89.74	87.5	85.37	94.29	89.19	91.89	86.84	86.84
NPV	93.55	96.96	96.77	94.60	96.97	96.88	96.77	91.89	91.43	94.29	91.18	91.18
Accuracy	**87.55**	**93.06**	**90.28**	**95.83**	**93.06**	**91.67**	**90.27**	**93.06**	**90.28**	**93.06**	**88.89**	**88.89**
AUC	91.17	95.17	92.34	95.77	95.17	93.99	92.34	92.94	91.01	92.40	89.67	89.67
Cohen-k	**75.00**	**86.11**	**80.56**	**91.67**	**86.11**	**83.33**	**80.56**	**86.11**	**80.56**	**86.11**	**77.78**	**77.78**

Table 3: Indices for CAD system performance with all classifiers by using KS and W test

Indices (%)	L=8						L=32					
	KNN			SVM	LDA	QDA	KNN			SVM	LDA	QDA
	K=1	K=3	K=5				K=1	K=3	K=5			
Sensitivity	91.67	97.22	97.23	94.44	97.22	97.22	94.44	91.67	88.89	100	91.67	94.44
Specificity	83.33	91.67	83.33	94.44	86.11	83.33	86.11	91.67	86.11	88.89	86.11	77.78
PPV	84.62	92.11	85.37	94.44	87.5	85.37	87.18	91.67	86.49	90	86.84	80.60
NPV	90.91	97.10	96.77	94.44	96.88	96.77	93.94	91.67	88.57	100	91.18	93.33
Accuracy	**87.5**	**94.44**	**90.22**	**94.44**	**91.67**	**90.27**	**90.28**	**91.67**	**87.50**	**94.44**	**88.89**	**86.11**
AUC	91.01	95.73	92.34	94.14	93.97	92.34	93.06	91.04	88.08	94.32	89.67	91.27
Cohen-k	**75.00**	**88.89**	**80.56**	**88.89**	**83.33**	**80.56**	**80.56**	**83.33**	**75.00**	**88.89**	**77.78**	**72.22**

Table 4: Indices for CAD system performance with all classifiers by using SBS

Indices (%)	L=8						L=32					
	KNN			SVM	LDA	QDA	KNN			SVM	LDA	QDA
	K=1	K=3	K=5				K=1	K=3	K=5			
Sensitivity	100	97.22	97.22	88.89	97.22	94.44	100	100	100	94.44	97.22	91.67
Specificity	94.44	94.44	88.89	100	97.22	100	97.22	94.44	91.67	94.44	94.44	94.44
PPV	94.74	94.73	94.74	100	97.22	100	97.30	97.30	97.30	94.44	94.60	94.29
NPV	100	100	100	90	97.22	94.74	100	100	100	94.44	97.15	91.89
Accuracy	**97.22**	**95.83**	**93.06**	**94.44**	**97.22**	**97.22**	**98.61**	**97.22**	**95.83**	**94.44**	**95.83**	**93.06**
AUC	98.00	97.10	94.13	93.96	98.82	96.15	98.89	97.89	97.51	94.14	96.71	92.94
Cohen-k	**94.44**	**94.44**	**94.44**	**88.89**	**94.44**	**94.44**	**97.22**	**97.22**	**91.67**	**88.89**	**91.67**	**86.11**

Table 5: Indices for CAD system performance with all classifiers by using SFS

Indices (%)	L=8						L=32					
	KNN			SVM	LDA	QDA	KNN			SVM	LDA	QDA
	K=1	K=3	K=5				K=1	K=3	K=5			
Sensitivity	100	100	97.22	97.22	94.44	94.44	100	100	100	100	100	83.33
Specificity	97.22	100	100	97.22	94.44	100	100	100	97.22	97.22	91.97	100
PPV	97.30	100	100	97.22	94.44	100	100	100	97.30	97.30	92.31	100
NPV	100	100	97.30	97.22	94.44	94.74	100	100	100	100	100	85.71
Accuracy	**98.61**	**100**	**98.61**	**97.22**	**94.44**	**97.22**	**100**	**100**	**98.61**	**98.61**	**95.83**	**91.67**
AUC	98.89	99.70	98.43	97.41	95.41	96.15	99.70	99.70	98.89	98.72	97.51	91.07
Cohen-k	**97.22**	**100**	**97.22**	**94.44**	**88.89**	**94.44**	**100**	**100**	**97.22**	**97.22**	**91.67**	**83.33**

Table 6: Indices for CAD system performance with all classifiers by using SFFS

Indices (%)	L=8						L=32					
	KNN			SVM	LDA	QDA	KNN			SVM	LDA	QDA
	K=1	K=3	K=5				K=1	K=3	K=5			
Sensitivity	100	100	100	94.44	94.44	94.44	100	100	100	97.22	100	94.44
Specificity	94.44	94.44	91.67	100	94.44	94.44	97.22	100	97.22	100	91.67	88.89
PPV	94.74	94.73	92.31	100	94.44	94.44	97.30	100	97.30	100	92.31	89.74
NPV	100	100	100	94.74	94.44	94.44	100	100	100	97.30	100	94.11
Accuracy	**97.22**	**97.22**	**95.83**	**97.22**	**94.44**	**94.44**	**98.61**	**100**	**98.61**	**98.61**	**95.83**	**91.67**
AUC	98.44	98.44	97.61	96.15	95.41	59.41	98.72	99.70	98.72	97.48	97.61	92.29
Cohen-k	**94.44**	**94.44**	**91.67**	**94.44**	**88.89**	**88.89**	**97.22**	**100**	**97.22**	**97.22**	**91.67**	**83.33**

Table 7: Indices for CAD system performance with all classifiers by using BBS

Indices (%)	L=8						L=32					
	KNN			SVM	LDA	QDA	KNN			SVM	LDA	QDA
	K=1	K=3	K=5				K=1	K=3	K=5			
Sensitivity	97.22	91.67	94.44	91.67	94.44	91.67	100	100	100	97.22	97.22	97.22
Specificity	91.67	97.22	97.22	94.44	94.44	94.44	86.11	86.11	91.67	94.44	91.67	94.44
PPV	92.11	97.06	97.15	94.29	94.44	94.29	87.80	87.80	92.31	94.60	92.11	94.60
NPV	97.10	92.11	94.60	91.89	94.44	91.89	100	100	100	97.15	97.10	97.15
Accuracy	**94.44**	**94.44**	**95.83**	**93.06**	**94.44**	**93.06**	**93.06**	**93.06**	**95.83**	**95.83**	**94.44**	**95.83**
AUC	96.41	95.28	95.72	92.94	95.41	92.94	93.39	93.82	97.61	96.71	96.41	96.71
Cohen-k	**88.89**	**88.89**	**91.67**	**86.11**	**88.89**	**86.11**	**86.11**	**86.11**	**91.67**	**91.67**	**88.89**	**91.67**

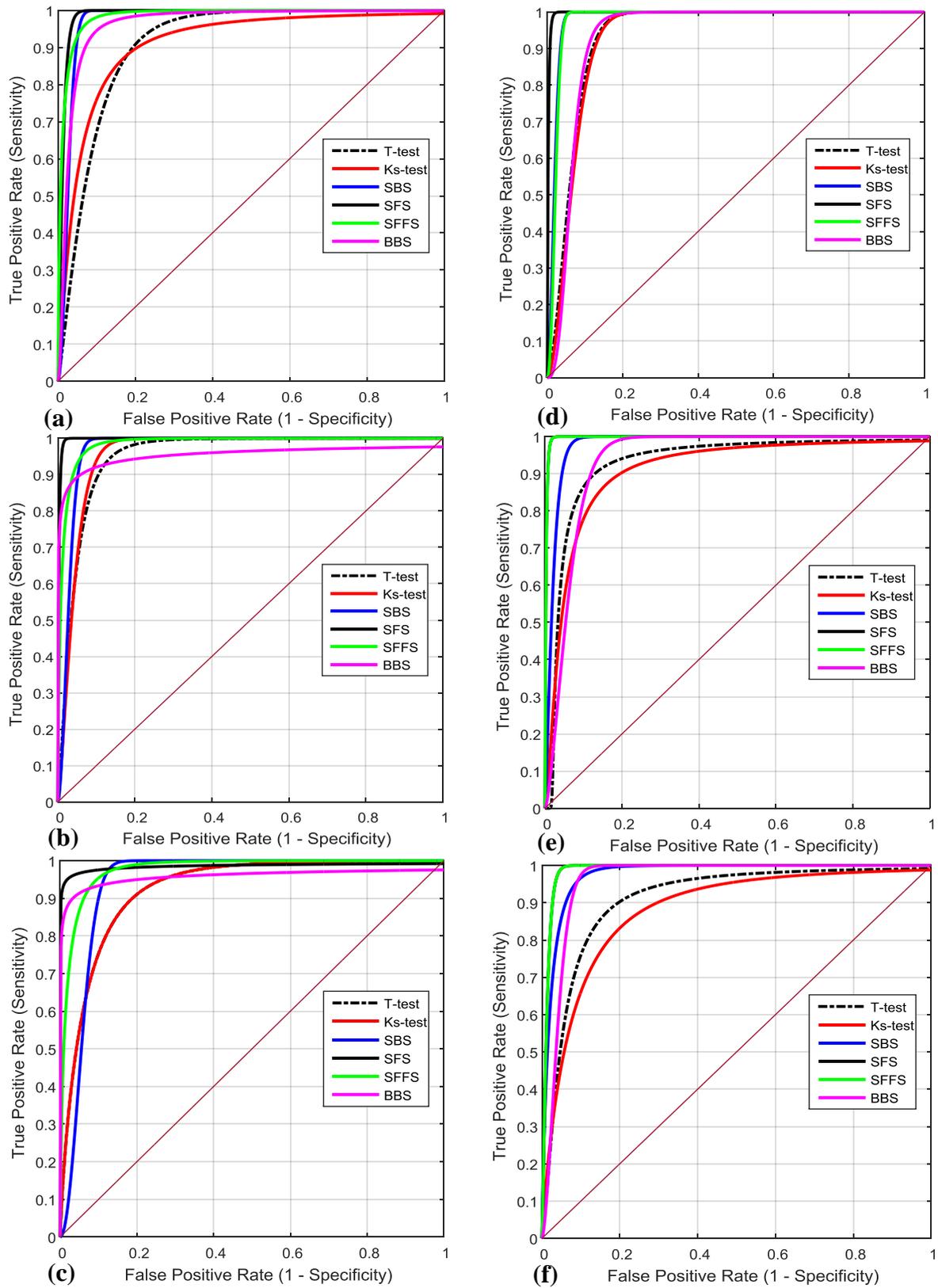*Figure 5: ROC curves for KNN classifier at K=1 (a, d), K =3 (b, e) and K = 5 (c,f). Left side is for L=8 and right side for L=32.*

Figure 5 shows independently a set of ROC curves for KNN classifier (K=1, 3 and 5) with all selection methods at L=8 and 32 corresponding to aforementioned discussion. Because KS-test and W-test methods selected the same and similar number of features, the performance behaviour of all classifiers exactly is similar. So,they have the same ROC curves (red curves).Due to this similarity, a coincidence between some ROC curves, in shape, is happened with some selection methods. For example, we can see that clearly in Figure 5(d) with SBS and SFFS selection methods. The ROC curves for SVM, LDA and QDA classifiers with all selection methods are depicted in Figure 6with both values of L.

Table 8 presents the comparison between the previous works of literature review and our results for the proposed CAD system. Our results are different corresponding to the selection method as we mentioned before. The results of CAD system performance also depend on the type of database that is used. There are many databases available online to use in the field of academic work such as mini-MIAS and digital database for screening mammography (DDSM) [28]. Our results show high performance with SFS method with KNN (K=1 and 3 at L=32) as well as SFFS method with KNN (K=3), where AUC is equal 99.70%.

Table 8: Comparison between our results and previousresult in the literature

Author	Database Type	No. of image	Sensitivity (%)	Specificity (%)
Dheeba *et al.*,2014 [16]	Mini-MIAS	216	94.167	92.105
Elbially *et al.*, 2013 [29]	Mini-MIAS	147	94	95
Kurt *et al.*, 2014 [30]	Mini-MIAS	96	93.2	80.6
Elmanna *et al.*, 2015 [31]	DDSM	40	94	98
S. Sharma et.al., 2015 [32]	DDSM	200	97	96
Vállez *et al.*, 2014 [9]	Mini-MIAS&	322	Accuracy (99.75%)	
	FFDM	1137	Accuracy (91.58%)	
Bosch *et al.*, 2006 [33]	Mini-MIAS	322	Accuracy (93.40%)	
Wang *et al.*, 2011 [34]	Mini-MIAS	322	Accuracy (89.00%)	
Subashini *et al.*, 2010 [35]	Mini-MIAS	43	Accuracy (95.44%)	
Nascimento *et al.*, 2013 [36]	DDSM	360	AUC (98.00%)	
Ramos *et al.,* 2012 [37]	DDSM	120	AUC (90.00%)	
Our work	**Mini-MIAS**	**144**	**AUC (99.70%)**	

Conclusions

In this paper, a hybrid CAD system for breast cancer diagnosis is developed. The performances of several classifiers with multiple selection methods at two different values of quantization level of GLCM are compared. The results show that the performance of all classifiers is better with SBS, SFS, SFFS and BBS methods than those of statistical methods. For the statistical selection methods, the performance of most classifiers is better at L = 8 except with KNN when k=1. All classifiers have better performance at L=32 with PR-Toolbox methods. But BBS method has slightly different manner, comparing with other PR-Toolbox algorithms, especially with KNN and QDA classifiers where its performance is better when L=8. On the other hand, it has better performance when L=32 with SVM and LDA classifiers.QDA classifier always provides better performance with all selection methods when L = 8 except with BBS method which is better with L=32. The performance of CAD system is found to be the best with SFS selection method. Also we can consider the SFFS method to be second encouraging choice during CAD system development especially with L=32. LDA classifier mostly has the consistency performance especially when L=8 with SFS, SFFS and BBS methods with all evaluated metrics. Selection of powerful features guarantees high performance of the classifier. This study shows that feature selection is the most important stage to develop CAD system through the comparison between the different selection algorithms. The contribution of this work considered as building two independent models of CAD system with different values of L. The results show better performance of classifiers when GLCM size is equal the size of ROI.

Conflict of Interest Statement

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

[1]. Cheng, M. C. (2004). Mass Lesion Detection with a Fuzzy Neural Network", Pattern Recognition.International Journal of health care, 37(6): 1189-1200.

[2]. "Cancer Facts & Figures", Available:
http://www.cancer.org/research/cancerfactsfigures/cancerfactsfigures, [Accessed: Oct.,-2016].

[3]. Sampat, P. M.,Markey, M. K., &Bovik, B. C. (2005). Computer-aided detection and diagnosis in mammography.Handbook of Image and Video Processing. 2nd ed., 1195-1217.

[4]. Castellino, R. (2005). Computer aided detection (CAD): an overview. Cancer Imaging,5: 17–19.

[5]. "American Cancer Society Guidelines for the Early Detection of Cancer",Available:http://www.cancer.org/healthy/findcancerearly/cancerscreeningguidelines/american-cancer-society-guidelines-for-the-early-detection-of-cancer, [Accessed: Nov.-2016].

[6]. Heath, M., Bowyer, K, Kopans, D., Kegelmeyer, W. P., Moore, R., Chang, K., & Kumaran, S. M. (1998). Current status of the Digital Database for Screening Mammography. Proceedings of the Fourth International Workshop on Digital Mammography, 5(2): 201-209.

[7]. Rangayyan, R. M., Banik, S., & Desautels, J. E. (2010). Computer-aided detection of architectural distortion in prior mammograms of interval cancer. J. Digital Imaging,23(5): 611–31.

[8]. "Quantitative Image Analysis/Computer-aided Diagnosis", Available:
http://radiology.uchicago.edu/page/quantitative-image-analysiscomputer-aided-diagnosis, [Accessed: May-2015].

[9]. Vállez, N., Bueno, G., Déniz, D., Dorado, J., Seoane, J., Pazos, A., & Pastor, C. (2014). Breast density classification to reduce false positives in CADe systems. Comput. Methods Programs Biomed., 113(2): 569–84.

[10]. Bueno, G., Vállez, N., Déniz, O., Esteve, P., Rienda, M., Arias, M., & Pastor, C., (2011). Automatic breast parenchymal density classification integrated into a CADe system. Int. J. Comput. Assist. Radiol. Surg., 6(3): 309–18.

[11]. Guliato, D., Rangayyan, R., Carnelli, W., Zuffo, J., & Desautels, J. (1998). Segmentation of breast tumors in mammograms by fuzzy regiongrowing.Proceedings of 20th Annual International Conference IEEE Engineering in Medicine and Biology,1002-1004.

[12]. Wei, D., Chan, H. P.,Helvie, M., Sahiner, B., Petrick, N., Adler, D., &Goodsitt, M. (1995). Classification of mass and normal breast tissue on digital mammograms: multiresolution texture analysis.Medical Physics., 22: 1501-13.

[13]. Oliver, A. Llad´o, X. Mart´i, J. Mart´i, R. Freixenet .(2007). False positive reduction in breast mass detection using two-dimensional PCA. In: Lect. Not. in Comp. Sc. ,4478: 154–161.

[14]. Omara, A., Mohamed, A., Youssef, A., &Kadah, Y. M.(2006). Computer Aided Diagnosis in Digital Mammography. the third Cairo International Biomedical Engineering Conference, CIBEC 06.

[15]. Mudigonda, N. R., Rangayyan R. M., &Desautels, J. (2001). Detection of breast masses in mammograms by density slicing and texture flow-field analysis. IEEE Trans Med Imaging.

[16]. Dheeba, J., Albert, N., & Tamil s. (2011). Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach. J. Biomed. Inform., 49: 45–52.

[17]. Kallenberg, M., &Karssemeijer, N. (2010). Comparison of Tilt Correction Methods in Full Field Digital Mammograms. Digital Mammography, 10th International Workshop, IWDM 2010.

[18]. Wang, X. H., Good, W. F., Chapman, B. E., Chang, Y. H., Poller, W., Chang, T. S., &Hardesty, L. A. (2003). Automated assessment of the composition of breast tissue revealed on tissue thickness corrected mammography. AJR. Am. J. Roentgenol.,180(1): 257–62.

[19]. Wu, T., Moore, R. H., &Kopans, D. B. (2010). Multi-threshold peripheral equalization method and apparatus for digital mammography and breast Tomosynthesis. US Patent 7-2010.

[20]. "The mini-MIAS database of mammograms", Available: http://peipa.essex.ac.uk/info/mias.html, [Accessed: March-2016].

[21]. Haralick, R. M., Shanmugam, K., &Dinstein, I. (1973). Textural Features for Image Classification. IEEE Trans. Syst. Man. Cybern., 3(6): 610–621.

[22]. Mallat, S. (1989). A theory for multiresolution signal decomposition: the wavelet representation. IEEE Pattern Anal. and Machine Intell., 11(7): 674–693.

[23]. Juszczak, P., Tax, D., &Dui, R. P. (2002). Feature scaling in support vector data descriptions. Proc. 8th Annu. Conf. Adv. School Comput. Imaging.

[24]. Massey, F. J. (1951). The Kolmogorov-Smirnov Test for Goodness of Fit. Journal of the American Statistical Association., 46(253):68–78.

[25]. Gibbons, J.D., &Chakraborti, s. (2011).Nonparametric Statistical Inference. 5th Ed., Boca Raton, FL: Chapman& Hall/CRC Press, Taylor & Francis Group.

[26]. "A matlab toolbox for pattern recognition", Available: http://prtools.org/software/,[Accessed: May-2016].

[27]. Landis, J. R. (1977). The measurement of observer agreement for categorical data. Biometrics, 33:159-174.

[28]. Heath, M., Bowyer, K., Kopans, D., Moore, R., & Philip, W. (2001). The Digital Database for Screening Mammography. Proceedings of the Fifth International Workshop on Digital Mammography, Medical Physics Publishing.

[29]. Elbially, M. S., Ahmed, M. S., Abdelgawwad, M. H., Ali, F. A., Botros, F. S., &Kadah, Y. M. (2013). K8. Hand-Held Computer Aided Diagnosis System with Application in Mammography.Proc. 30th National Radio Science Conference, Cairo, pp. 549-556.

[30]. Kurt, B., Nabiyev, V., &Turhan, T. (2014). A novel automatic suspicious mass regions identification using Havrda & Charvat entropy and Otsu's N thresholding. Comput. Methods Programs Biomed., 114(3): 349–60.

[31]. Elmanna, M. E., & Kadah, Y. M. (2015). Implementation of Practical Computer Aided Diagnosis System for Classification of Masses in Digital Mammograms. Proc. ICCNEEE 2015, Khartoum.

[32]. Bosch, A., Munoz, X., Oliver, A., &Marti, J. (2006). Modeling and classifying breast tissue density in mammogram. Proc.of the 2006 IEEE Computer Society Conference on ComputerVision and Pattern Recognition.

[33]. Wang, J., Li, Y., Zhang, Y., Xie, H., &Wang, C. (2011). Features based classification of breast parenchymal tissue in the mammogram via jointly selecting and weighting visual words. Proc. of the 2011 Sixth Intern. Conference onImage and Graphics.

[34]. Subashini, T. S., Ramalingam, V., &Palanivel, S. (2010). Automated assessment of breast tissue density in digital mammograms. Computer Vision and Image Understanding, 114: 33–43.

[35]. Nascimento, M. N., Martins, A. S., Neves, L., Ramos, R., Flores, E., &Carrijo, G. A. (2013). Classification of masses in mammographic image using wavelet domain features and polynomial classifier. Expert Syst. Appl., 40(15): 6213–622.

[36]. Ramos, R. P., Nascimento, M. Z., & Pereira D.(2012). Texture extraction: An evaluation of ridgelet, wavelet and co-occurrence based methods applied to mammograms. Expert Syst. Appl.,39(12): 11036–11047.

[37]. Sharma, S., &Khanna, P. (2015). Computer-Aided Diagnosis of Malignant Mammograms using Zernike Moments and SVM. J Digit Imaging, 36(5): 115-120.