UNDER PEER REVIEW

Breast Cancer Computer Aided Diagnosis System from Digital Mammograms

ABSTRACT

Recently, breast cancer is one of the most popular cancers that women could suffer from. The gravity and seriousness of breast cancer can be evidenced from the fact that the mortality rates associated with it are the second highest after lung cancer. For the treatment of breast cancer, Mammography has emerged as the one whose modality when it comes to the defection of this cancer is most effective despite the challenges posed by dense breast parenchyma. In this regard, computer-aided diagnosis (CADe) leverages the mammography systems' output to facilitate the radiologist's decision. It can be defined as a system that makes a similar diagnosis to the one done by a radiologist who relies for his/her interpretation on the suggestions generated by a computer after it analyzed a set of patient radiological images when making. Against this backdrop, the current paper examines different ways of utilizing known image processing and techniques of machine learning detection of breast cancer using CAD – more specifically, using mammogram images. This in turn helps pathologist in their decision-making process. For an effective implementation of this methodology, CADe system was developed and tested on the public and freely available mammographic databases named MIAS database. CADe system is developed to differentiate between normal and abnormal tissues, and it assists radiologists to avoid missing breast abnormalities. The performance of all classifiers is the best by using sequential forward selection (SFS) method. Also, we can conclude that the quantization gray level of (gray-level co-occurrence matrices) GLCM is a very significant factor to get robust high order features where the results are better with L equal to the size of ROI. Using an enormous number of several features assist the CADe system to be strong enough to distinguish between the different tissues.

Keywords: CAD, Breast Cancer, Medical Image Processing, Feature Extraction, Digital Mammography, Feature Selection, Classifications, Computer Applications in Medicine.

1. Introduction

In the recent past, many researchers have developed CADe systems to classify and detect abnormalities in the breast. In many systems, certain some common stages can be achieved in order to find the suspicious lesions. Figure 1 depicts these stages.

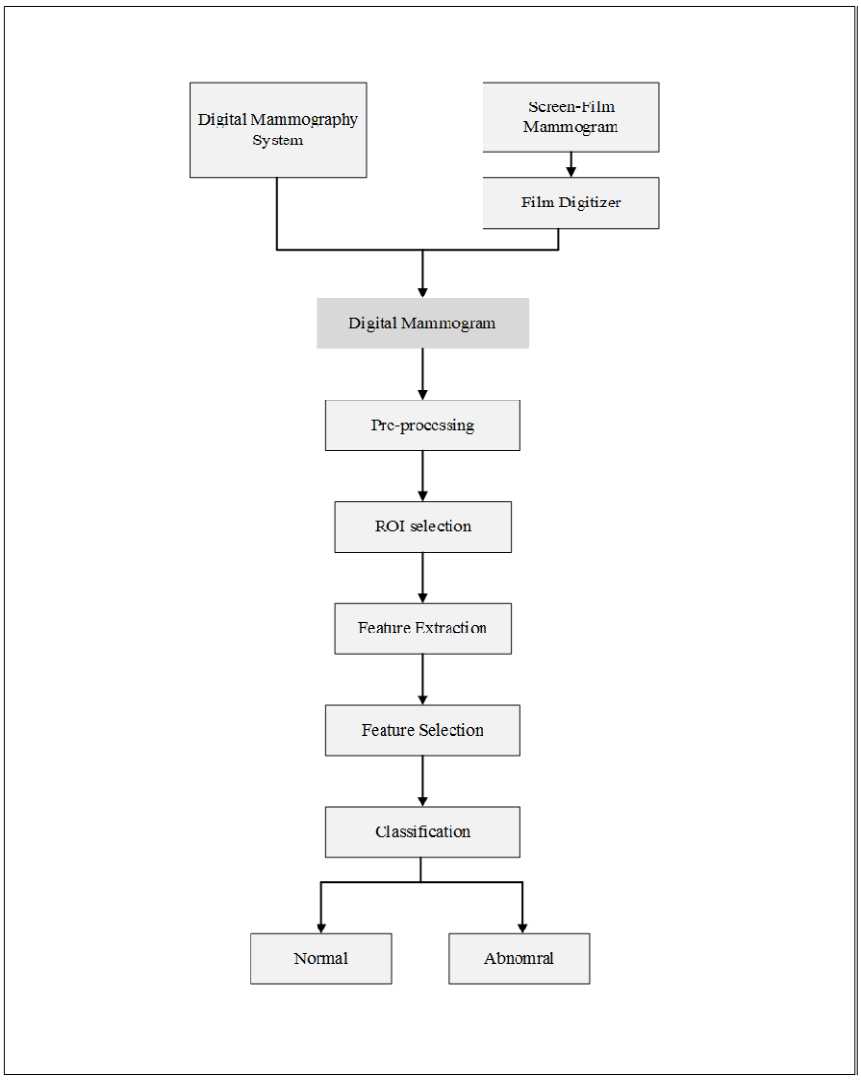


Figure 1 Visual representation of a generic CADe system.

Digitized mammography databases were used for developing our CADe system in various stages, commencing with the pre-processing phase where the region surrounding the breast was segmented by applying techniques of image processing for lowering the mammograms' noise ratio. Meanwhile the next stage entailed the selection of the region of interest (ROI). Here, several suspicious ROIs are selected to identify them as either abnormal or normal lesions. The following stage encompasses the feature extraction, whose objective is to ensure the lesions' characterization and distinguish actual lesions from their falsely detected counterparts; several features were calculated for the selected ROI. After that, we performed feature selection analysis, in which denotes a vital step in developing the classification system. In order to have a successful classification scheme, it was paramount to select the appropriate method and integrate them effectively into the model. Meanwhile classification was performed in the last stage. Here,

we fed the selected features into the classification system to train it to differentiate between normal and abnormal tissue.

Two stages were followed in this study, training stage followed by the testing stage. 50% of the database was used in the testing stage. In the training stage, normal images and cancerous images were used to train the system to differentiate between them. The second stage encompassed a testing stage where we introduced the system to a new image and ascertain the accuracy of detection results.

2. Literature Review

This section includes some extant literature studies which emphasized the use of CADe systems to classify problematic areas of the breast. These studies were intended to make improvements in the diagnostic performances of radiologists by classifying the aforementioned regions. Although there is a vast body of research papers to ascertain optimal performances of the CADe system, not many studies have been conducted pertaining to the subject we are intending to address.

False positives (FP) has been reduced in breast density classification by the CAD system enhanced by N. Vállez et al. [3] who devised a system of automating the classification of density of the breast and facilitate timely identification and evaluation of lesions. They proposed the use of CAD system for grouping the mammograms into various classes of tissues (BIRADS), basing this scheme of classification on as many as 298 features. Using some intuitive algorithms for detection, they undertook the testing with nearly 1460 images. According to the findings, 322 mammograms of the dataset (MIAS) demonstrated the classification of 99.76% of all samples.

In another study, [4], S. Pohlman et al. were able to successfully detect sensitivity of 97% from as many as 51 mammographic images using their ingenious method for intelligent region-expansion in order to group cancerous clusters from the normal background.

Similarly, Wei et al. [6] examined the possibility of differentiating between clusters and normal tissue in mammograms by analyzing textures using many different resolutions. Digitized mammograms regions of interest (ROIs) were broken down into different scales using wavelet transform. Optimal features as well as linear discriminant classifier were selected using stepwise linear discrimination technique.

Meanwhile with a view to reduce instances of inaccurate positives when detecting masses on the breast, Oliver et al. [7] suggested a technique that extracted features using the 2DPCA (or Two-Dimensional Principal Component Analysis) algorithm.. For assessment purposes, they used ROC (Receiver Operating Characteristics) evaluation.

Akram I. Omara et al. [8] meanwhile made use of k-nearest (voting) neighbor as well as MDC (minimum distance classifier) in order to extract 28 levels of details of wavelet coefficients by applying wavelet decomposition on the locally processed image and used them as features to differentiate between abnormal lesions and normal tissue.

In a similar vein, flow-like textural information for analysis methods in mammography were suggested by Mudigonda et al. [9]. For the purpose of identifying detected regions as false positives or accurate regions of masses, they successfully identified and segmented the mass regions, following which they went on to classify them as either benign or malignant areas. They accomplished this task by leveraging the method of logistic regression features of computing texture using GLCMs (gray-level co-occurrence matrices).

Li et al. [10] made an algorithm to trace masses by following a couple of steps. In step one - preliminary segmentation of suspicion areas was obtained using adaptive threshold of grey levels. In the subsequent step, features based on contrast, regions, size and shape were used across the selected areas in order to classify them as either normal tissue or as masses.

J. Dheeba et al. developed a CAD-bases system to trace cases of breast cancer [10] using a classifier called PSOWNN (particle swarm optimized wavelet neural network). It is noteworthy that the algorithm for detecting proposed abnormality facilitated the classification of potentially problematic areas close to the breast through the application of a pattern-based classifier. The database included 54 patients and 216 mammograms (collated from screening centers. According to the finding, this algorithm helped identify specificity of 92.105% and sensitivity of 94.167%.

3. Methodology

In this chapter, we developed a CADe system by using MIAS database [1]. This study includes some procedures to achieve the system. The first step is the preprocessing step (peripheral enhancement of breast) which is discussed in detail in chapter 3. Subsequently, we excerpted a set of ROIs from mammograms. Thereafter, we extracted some features from these ROIs using as many as seven methods of selection and making certain comparisons. These methods are Unpaired Student test, KS test (or Kolmogorov-Smirnov test), W-test (or Wilcoxon signed rank test), SFS/SBS test, (Sequential Forward and Backward Selection), BBS test (or Branch and Bound Selection) and SFSS (or Sequential Floating Forward Selection). Then we used some classifiers to classify between lesions using classifiers such as KNN (K-Nearest Neighbor) classifier, “SVM (Support Vector Machine) classifier, LDA (Linear Discriminant) classifier, QDA (Quadtradic Discriminant Analysis) classifier, NB (Naïve Bayes) classifier as well as ANN (Artificial Neural Networks) classifier. All these classifiers can be utilized for the purpose of performing a hard classification, where the output is a binary class label, therefore, the (0) label for normal and (1) label for abnormal or (cancer) breast. Finally, we evaluated the CADe system performance using several indices such as specificity, sensitivity, positive and negative predictive value, respectively, and Cohen-k factor, among others.

The entire stimulation was performed in MATLAB® (R2016b) software in i7-4500M CPU (Intel® Core™) @ 2.40GHz system with 16GB RAM memory. Meanwhile the operating system was Windows 7 Home Premium 64-bit.

The proposed system was divided into the following blocks

Preprocessing

The first stage in the CAD system is called the preprocessing in order to augment the projected breast peripheral area's (uncompressed portion). We made use of the processing technique in order to reach this stage that is credited to Tao Wu et al. [17].

ROI Extraction

Using the information provided by the MIAS dataset for each mammogram, we used 72 normal and 72 abnormal mammograms. In addition, we extracted 144 centered ROIs using a window whose size was 32×32 pixels. Figure.2 illustrates some types of the masses which are extracted from the MIAS database mammograms with the white circle surrounding the masses in each mammogram.

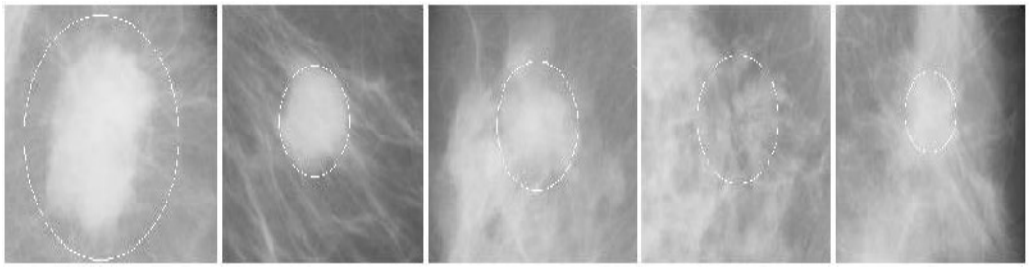


Figure 2 : An Example of masses of some MIAS mammogram types. The first one from the left: Spiculated mass (mdb148). The second: Circumscribed mass (mdb028). Third: ill-defined mass (mdb265). Fourth: Architectural distortion mass (mdb125). Fifth: Asymmetry mass (mdb102).

Features Extraction

This section explains one of the most important stages of CAD system that directly affect the performance of the system. Features refer to the texture's quantitative measures for elucidating an image's salient features. We can express these characteristics as mathematical descriptors to help in distinguishing between different tissues.

In this study, we extracted a range of different features for each selected ROI from spatial as well as transform domains. In addition, we obtained a total of seven hundred (700) different features from each ROI as following: First, Second and Higher Order Statistical Features, Zernike moment features and Wavelet Transform Features.

Wavelet Transform Features:

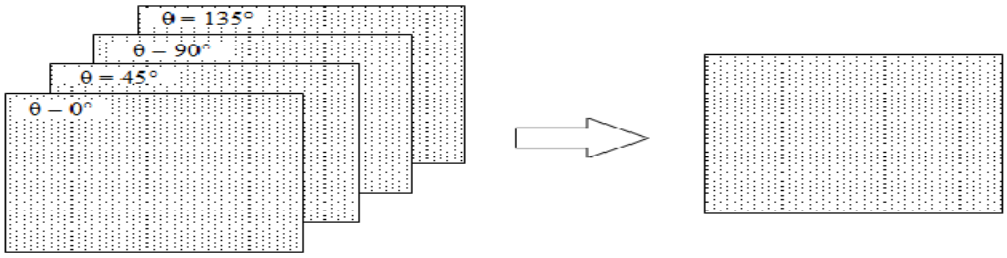
In today's day and age, wavelet transformation represents one of most effective transformations involving time-frequency. In our study, we implemented the wavelet decomposition on the

problem area of interest via the MATLAB (Wavelet Toolbox). The wavelet transform calculates the matrix of approximation coefficients (LL) as well as other coefficients matrices - LH, HL, and HH from the input matrix using the wavelet Daubechies [14]. Here the input matrix is specific to each ROI. Several research papers used wavelet transform to get specific features and used them in CAD system development.

In this study, we used only one level of wavelet transform using the MATLAB Wavelet Toolbox. Furthermore, we implemented the wavelet decomposition on each ROI as the input image to procure the aforementioned matrices. LL matrix meanwhile was not part of this study. At a subsequent stage, we used these matrices during the feature-extraction stage of the planned CADe system.

In our study, we used the idea proposed by Dhanashree Gadkari for calculating averaged GLCM from each wavelet coefficient matrix of the input image. However, there is a difference between our work and theirs. While they conducted their study on the Radar database, we computed averaged GLCMs for ROIs of mammograms in our study. It is notable that we computed two GLCMs on an average for every coefficient matrix. Four varied GLCMs were observed at divergent angles every time 'd' value was assigned. Subsequently, the average of these four GLCMs was used to calculate the total average. Under this step, we got a total of six (6) averaged GLCMs for each ROI.

Thereafter, the six resulting averaged GLCMs were used to extract as many as 16 features from each single GLCM as following: Entropy, Maximum probability, Homogeneity, Inverse Different Moment (Homogeneity2), Variance, Energy (Uniformity), autocorrelation, Correlation information1, Correlation information2 and seven (7) invariant moments. The mathematical formulas of these features are elucidated in Table 3 with the exception of the formulas of invariant features. Figure.3 shows how we obtained the averaged GLCM. The total features extracted from the wavelet transform part are ninety-six (96) features for each ROI.



(a) (b)

Figure. 3 (a) A separated GLCMs at different angles with constant value of d for all coefficient matrices. (b) Average GLCM.

The mathematical formulas of the invariant features are described as following:

An order's (p + q) 2D movement of digital image f(x,y) of the size M×N can be calculated as:

$$m_{pq} = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} x^p \, y^q \, f(x,y) \quad (1)$$

where p = 0, 1, 2, … are integers. Meanwhile the other order movement (central) (p + q) is calculated as:

$$\mu_{pq} = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} (x - \bar{x})^p \, (y - \bar{y})^q \, f(x,y) \quad (2)$$

$$\text{where,} \qquad \bar{x} = \frac{m_{10}}{m_{00}}, \qquad \bar{y} = \frac{m_{01}}{m_{00}}, \quad (3)$$

η_{pq}denotes the moments' normalized central. It is calculated as:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu^{\gamma}_{00}}, \quad (4)$$

$$\gamma = \frac{p+q}{2} + 1 \qquad \text{for } p + q = 2,3 \ldots \quad (5)$$

We can derive seven moments (invariant) from the second and third moments as follows:

$$\phi_1 = \eta_{20} + \eta_{02} \quad (6)$$

$$\phi_2 = (\eta_{20} + \eta_{02})^2 + 4\eta^2_{11} \quad (8)$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \quad (9)$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (3\eta_{21} + \eta_{03})^2 \quad (10)$$

$$\phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + 3\eta_{12})\left\{(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2\right\} + (3\eta_{21} - \eta_{03})(3\eta_{21} + \eta_{03})\left\{3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right\}. \quad (11)$$

$$\phi_6 = (\eta_{20} - \eta_{02})\left\{(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right\} + 4\eta_{11}(\eta_{30} + \eta_{12})\,(\eta_{21} + \eta_{03}) \quad (12)$$

$$\phi_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})\left\{(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2\right\} + (3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{03})\left\{3(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2\right\}. \quad (13)$$

The final feature set is composed of 700 features for each ROI. Table.4 as shows the number of features extracted from each feature type.

Table. 1 Summary of the extracted features

Features categories	Number of extracted features
First order statistical features	39
Second order statistical features	400
Higher order statistical features	44
Wavelet transform features	96
Zernike moment features	121
Total number of features	700

Normalization of Extracted Features

Following the extraction of the feature set, we are required to rescale the features in the range of [0, 1] or [−1, 1] to make them independent of each other. In this regard, selection of the target range is predicated on the data's nature and scope. Features normalization is an important step to simplify the coefficients value to avoid any statistic bias in classification stage [15]. For this purpose, we applied a common approach to normalization referred to as Min-Max scaling wherein the data is scaled to a range [0,1] using the following formula:

$$x_{\grave{n}orm} = \frac{(x - min(x))}{max(x) - min(x)} \quad (14)$$

where x denotes the original value of the feature, $max(x)$ refers to the maximum value in the features vector, $xmin$ signifies the minimum value in the features vector and x_{norm}denotes the feature's normalized feature.

Features Selection

The stage of selecting features denoted a vital component of all classification schemes since it aims to select from the extracted feature set a number of features that are most relevant to the predictive modeling problem and yield minimum classification error. A CAD system's performance is predicated on the efficacy with which selection of the features is undertaken [16]. In addition, methods of selecting feature enable us to lower computation time, make improvements in performance of prediction, and facilitate well-informed decisions relating to applications that entail the use of pattern recognition or machine learning in order to decipher data in a better manner.

Unpaired t-test is performed for the hypothesis as per which data is divided into varying distributions (e.g. Here normal and abnormal lesions) with random samples sourced from equal variances as well as means, as opposed to the other alternative wherein there is no equality in the means. The t-test determines the amount of overlap between the two distributions. The ability of differentiation is determined by knowing this amount. In this study, we determined the significance level as following: $\alpha = 0.05$. In case there is a difference between p- value and

significance level wherein the former is less than the latter, we can then differentiate between these two sets which come from two different distributions by this feature [15].

Classification Stage

The classification is the final stage in any CAD system's development and entails the identification of categories to which novel observations belong, premised on the data's training set [18]. Features selected in the previous stage are mad to pass the classifier in both the phases. During the training phase, the selected features of training dataset which have already been labeled as normal or as abnormal are passed to the classifier, and the classifier is trained. In this stage, we used the training dataset which consists of 36 normal ROIs and 36 abnormal ROIs from the MIAS database. For this stage, we used the testing dataset containing 36 normal ROIs and 36 abnormal ROIs from the same dataset. This study witnessed the use of multiple classifiers (7) that have already been mentioned before.

4. Results and Discussion

The total number of ROIs used for the proposed CADe system from the MIAS database was 144 mammogram images (including 72 mammograms for normal case and 72 mammograms for abnormal lesions are (41 benign and 31 malignant)). The outcome of the feature selection resulted in seven hundred various features from each ROI of size 32x32 pixels selected from each mammogram. These features are 38 statistical features of the first order, 400 statistical features of the second order, 44 statistical features of a high order, and 96 features extricated from wavelet transform and also, one hundred and twenty-one features extracted from the Zernike moment of order twenty. Following this selection, we utilized as many as seven methods of features selection from the statistics and Pattern Recognition Toolbox. The statistics toolbox which are T-test, W-test, and KS-test, as well as pattern recognition toolboxes: SFFS, SFS, SBS, and BBS. We applied KNN with K=1 and 3, SVM, LDA, QDA, NB and ANN for classification of each selection method and also, we studied withal classifiers' behavioral pattern using all aforementioned methods of selection. We evaluated all classifiers' performances with each selection method by calculating PPV, NPV, sensitivity/specificity, Cohen-K factor, sensitivity, ROC curve and overall accuracy. The both classes' confusion matrices are used for obtaining the indices. We and make comparisons between different classifier performances wherein the selection of each method was done in an independent manner.

In this study, we used two different values of (L) of the GLCM equal at 8 and 32. The training and testing of KNN, SVM, LDA, QDA, NB and ANN classifiers were based on independent half and half training and testing sets with randomly chosen of each from the MIAS database is selected according to the available number of the samples. The normalization min-max method was utilized to make each feature selection values range between the zero and one. It can avoid the numerical instabilities in the operation of training the classifiers and let differences in various features to be well represented equally with no dominating features that occur to have broader numeric ranges.

Tables 1 and 2 summarize the confusion matrix entries computed for the aforementioned classifiers through the use of aforementioned methods of selecting features at the quantization gray levels' two levels (L=8 and 32 respectively).

Table. 1 Confusion Matrices in CADe system of quantization gray level at (L=8)

		T-test		W-test and KS-test				SBS		SFS		SFFS		BBS	
		N	A	N	A	N	A	N	A	N	A	N	A	N	A
KNN-1	N	35	3	35	4	35	4	35	1	36	3	36	3	34	3
	A	1	33	1	32	1	32	1	35	0	33	0	33	2	33
KNN-3	N	33	3	34	3	34	3	36	1	36	1	36	1	33	1
	A	3	33	2	33	2	33	0	35	0	35	0	35	3	35
SVM	N	35	0	35	0	35	0	35	0	36	0	36	0	34	2
	A	1	36	1	36	1	36	1	36	0	36	0	36	2	34
LDA	N	34	4	34	5	34	5	34	2	35	2	35	3	33	2
	A	2	32	2	31	2	31	2	34	1	34	1	33	3	34
QDA	N	33	3	33	3	33	3	34	1	35	5	35	4	33	2
	A	3	33	3	33	3	33	2	35	1	31	1	32	3	34
NB	N	33	3	33	3	33	3	34	1	35	5	35	4	33	2
	A	3	33	3	33	3	33	2	35	1	31	1	32	3	34
ANN	N	31	2	36	8	36	8	34	1	36	4	36	2	34	1
	A	5	34	0	28	0	28	2	35	0	32	0	34	2	35

Table. 2 Confusion Matrices in CADe system of quantization gray level at (L=32)

		T-test		W-test and KS-test				SBS		SFS		SFFS		BBS	
		N	A	N	A	N	A	N	A	N	A	N	A	N	A
KNN-1	N	36	2	36	4	36	4	36	1	36	0	36	0	34	3
	A	0	34	0	32	0	32	0	35	0	36	0	36	2	33
KNN-3	N	33	3	33	4	33	4	34	3	36	1	36	0	33	1
	A	3	33	3	32	3	32	2	33	0	35	0	36	3	35
SVM	N	35	0	36	0	36	0	35	2	36	0	36	0	34	2
	A	1	36	0	36	0	36	1	34	0	36	0	36	2	34
LDA	N	34	5	34	5	34	5	34	4	34	2	34	2	33	2
	A	2	31	2	31	2	31	2	32	2	34	2	34	3	34
QDA	N	34	4	35	6	35	6	33	4	35	11	33	2	33	2
	A	2	32	1	30	1	30	3	32	1	25	3	34	3	34
NB	N	34	4	35	6	35	6	33	4	35	11	33	2	33	2
	A	2	32	1	30	1	30	3	32	1	25	3	34	3	34
ANN	N	33	3	33	0	33	0	32	1	36	1	36	2	32	1
	A	3	33	3	36	3	36	4	35	0	35	0	34	4	35

Tables 3, 4, 5, 6, 7, and 8 summarize the different measures calculated from the findings observes in Tables 1 and 2, respectively in order to better decipher the outcomes of classifiers such as sensitivity, specificity, PPV, NPV, accuracy, AUC, Cohen-K factor, where each

independently denotes the performance indices of CAD system at different selection methods with the following quantization levels: L = 8 and L = 32. There is a compelling case for comparing the performances of all classifiers with each selection method, but we will try to accommodate as much information as possible.

Figure 4 illustrates ROC plots concerning all classifiers (KNN with K=1 and K= 3, SVM, LDA, QDA, NB and ANN) with all selection methods at L=8 which that T-test (a), KS and W-test (b), SBS (c), SFS (d), SFFS (e), and BBS (f).

Figure.5 depicts a group of ROC plots concerning all classifiers (KNN with K=1 and K= 3, SVM, LDA, QDA, NB and ANN) with all selection methods at L=32 which that T-test (a), KS and W-test (b), SBS (c), SFS (d), SFFS (e), and BBS (f).

Table. 3 CADe performance indices during the T-test involving the classifiers.

Levels	**L = 8**							**L = 32**						
Indices (%)	KNN		SVM	LDA	QDA	NB	ANN	KNN		SVM	LDA	QDA	NB	ANN
	K=1	K=3						K=1	K=3					
Sensitivity	97.22	91.66	97.22	94.44	91.66	91.66	**86.11**	100	91.66	97.22	94.44	94.44	94.44	91.66
Specificity	91.66	91.66	100	88.88	91.66	91.66	94.44	94.44	91.66	100	86.11	88.88	88.88	91.66
PPV	92.10	91.66	100	89.47	91.66	91.66	93.93	94.73	91.66	100	87.17	89.47	89.47	91.66
NPV	97.05	91.66	97.29	94.11	91.66	91.66	87.17	100	91.66	97.29	93.93	94.11	94.11	91.66
Accuracy	94.44	91.66	98.61	91.66	91.66	91.66	90.27	97.22	91.66	98.61	90.27	91.66	91.66	91.66
AUC	94.93	91.99	98.57	92.73	91.62	91.62	90.27	96.80	92.13	98.63	91.57	94	94	92.09
Cohen-K	88.88	83.33	97.22	83.33	83.33	83.33	80.55	94.44	83.33	97.22	80.55	83.33	83.33	83.33

Table. 4 CADe performance indices during the KS and W test involving the classifiers.

Levels	**L = 8**							**L = 32**						
Indices (%)	KNN		SVM	LDA	QDA	NB	ANN	KNN		SVM	LDA	QDA	NB	ANN
	K=1	K=3						K=1	K=3					
Sensitivity	97.22	94.44	97.22	94.44	91.66	91.66	100	100	91.66	100	94.44	97.22	97.22	91.66
Specifi	88.	91.	100	86.	91.	91.	**77.**	88.	88.	100	86.	83.	83.	100

city	88	66		11	66	66	77	88	88		11	33	33	
PPV	89.74	91.89	100	87.17	91.66	91.66	81.81	90	89.18	100	87.17	85.36	85.36	100
NPV	96.96	94.28	97.29	93.93	91.66	91.66	100	100	91.42	100	93.93	96.77	96.77	92.30
Accuracy	93.05	93.05	98.61	90.27	91.66	91.66	88.88	94.44	90.27	100	90.27	90.27	90.27	95.83
AUC	92.93	93.42	98.57	91.57	91.62	91.62	90.89	93.38	90.61	99.69	91.57	92.59	92.59	97.79
Cohen-K	86.11	86.11	97.22	80.55	83.33	83.33	77.77	88.88	80.55	100	80.55	80.55	80.55	91.66

Table. 5 CADe performance indices on the usage of SBS.

Levels	**L = 8**							**L = 32**						
Indices (%)	KNN		SVM	LDA	QDA	NB	ANN	KNN		SVM	LDA	QDA	NB	ANN
	K=1	K=3						K=1	K=3					
Sensitivity	97.22	100	97.22	94.44	94.44	94.44	94.44	100	94.44	97.22	94.44	91.66	91.66	88.88
Specificity	97.22	97.22	100	94.44	97.22	97.22	97.22	97.22	91.66	94.44	88.88	88.88	88.88	97.22
PPV	97.22	97.29	100	94.44	97.14	97.14	97.14	97.29	91.89	94.59	89.47	89.18	89.18	96.96
NPV	97.22	100	97.29	94.44	94.59	94.59	94.59	100	94.28	97.14	94.11	91.42	91.42	89.74
Accuracy	97.22	98.61	98.61	94.44	95.83	95.83	95.83	98.61	93.05	95.83	91.66	90.27	90.27	93.05
AUC	97.08	98	98.57	95.67	95.39	95.39	96.60	98	95.19	95.85	93.40	91.06	91.06	92.29
Cohen-K	94.44	97.22	97.22	88.88	91.66	91.66	91.66	97.22	86.11	91.66	83.33	80.55	80.55	86.11

Table. 6 CADe performance indices on the usage of SFS.

Levels	**L = 8**							**L = 32**						
Indices (%)	KNN		SVM	LDA	QDA	NB	ANN	KNN		SVM	LDA	QDA	NB	ANN
	K=1	K=3						K=1	K=3					
Sensitivity	100	100	100	97.22	97.22	97.22	100	100	100	100	94.44	97.22	97.22	100

Specificity	91.66	97.22	100	94.44	86.11	86.11	88.88	100	97.22	100	94.44	69.44	69.44	97.22
PPV	92.30	97.29	100	94.59	87.50	87.50	90	100	97.29	100	94.44	76.08	76.08	97.29
NPV	100	100	100	97.14	96.87	100	100	100	100	100	94.44	96.15	96.15	100
Accuracy	95.83	98.61	100	95.83	91.66	91.66	94.44	100	98.61	100	94.44	83.33	83.33	98.61
AUC	95.60	98.23	99.69	96.50	93.31	93.31	95.70	99.69	98.72	99.69	95.67	85.16	85.16	98.72
Cohen-K	91.66	97.22	100	91.66	83.33	83.33	88.88	100	97.22	100	88.88	66.66	66.66	97.22

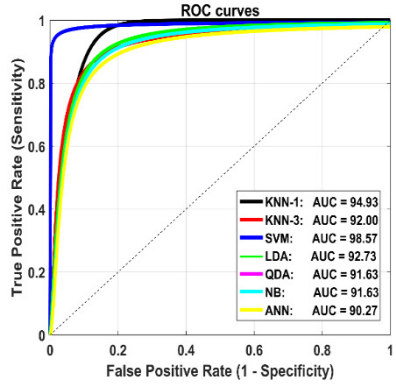
Table. 7 CADe performance indices on the usage of SSFSS.

Levels	**L = 8**							**L = 32**						
Indices (%)	KNN		SVM	LDA	QDA	NB	ANN	KNN		SVM	LDA	QDA	NB	ANN
	K=1	K=3						K=1	K=3					
Sensitivity	100	100	100	97.22	97.22	97.22	100	100	100	100	94.44	91.66	91.66	100
Specificity	91.66	97.22	100	91.66	88.88	88.88	94.44	100	100	100	94.44	94.44	94.44	94.44
PPV	92.30	97.29	100	92.10	89.74	89.74	94.73	100	100	100	94.44	94.28	94.28	94.73
NPV	100	100	100	97.05	96.96	96.96	100	100	100	100	94.44	91.89	91.89	100
Accuracy	95.83	98.61	100	94.44	93.05	93.05	97.22	100	100	100	94.44	93.05	93.05	97.22
AUC	95.81	98	99.69	96.27	95.21	95.21	97.89	99.69	99.69	99.69	95.09	93.79	93.79	97.00
Cohen-K	91.66	97.22	100	88.88	86.11	86.11	94.44	100	100	100	88.88	86.11	86.11	94.44

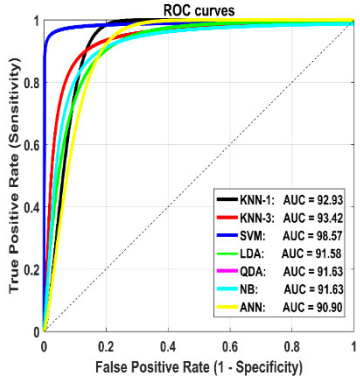
Table. 8 CADe performance indices on the usage of BBS.

Levels	L = 8							L = 32						
Indices (%)	KNN		SVM	LDA	QDA	NB	ANN	KNN		SVM	LDA	QDA	NB	ANN
	K=	K=3						K=1	K=3					
Sensiti	94.	91.	94.	91.	91.	91.	94.	94.	91.	94.	91.	91.	91.	88.

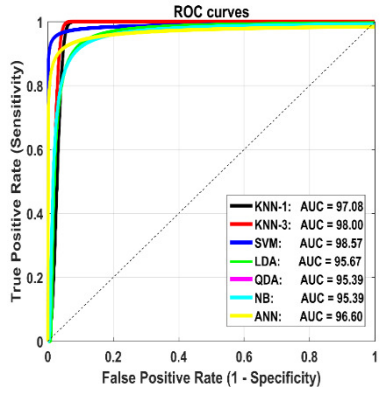
vity	44	66	44	66	66	66	44	44	66	44	66	66	66	88
Specifi city	91. 66	97. 22	94. 44	94. 44	94. 44	94. 44	97. 22	91. 66	97. 22	94. 44	94. 44	94. 44	94. 44	97. 22
PPV	91. 89	97. 05	94. 44	94. 28	94. 28	94. 28	97. 14	91. 89	97. 05	94. 44	94. 28	94. 28	94. 28	96. 96
NPV	94. 28	92. 10	94. 44	91. 89	91. 89	91. 89	94. 59	94. 28	92. 10	94. 44	91. 89	91. 89	91. 89	89. 74
Accura cy	93. 05	94. 44	94. 44	93. 05	93. 05	93. 05	95. 83	93. 05	94. 44	94. 44	93. 05	93. 05	93. 05	93. 05
AUC	95. 36	95. 27	95. 09	94. 11	94. 11	94. 11	96. 60	95. 36	95. 27	95. 09	94. 11	94. 11	94. 11	92. 29
Cohen-K	86. 11	88. 88	88. 88	86. 11	86. 11	86. 11	91. 66	86. 11	88. 88	88. 88	86. 11	86. 11	86. 11	86. 11



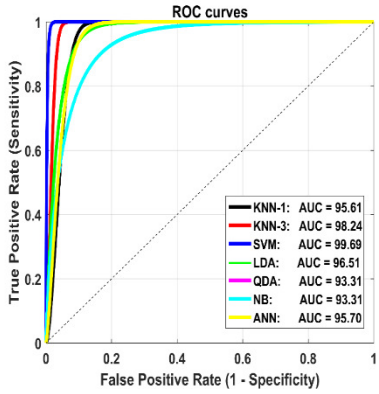
(a)



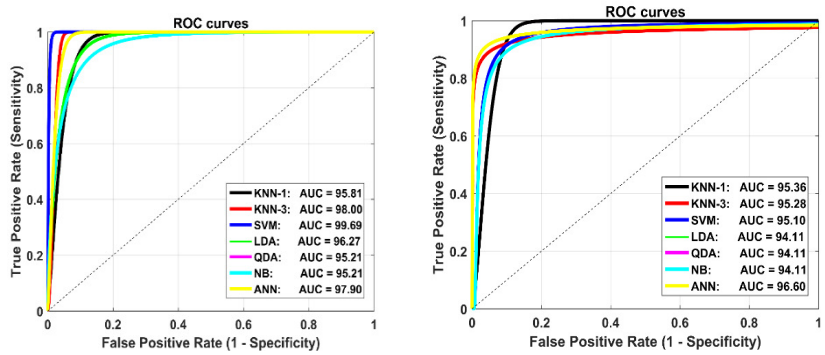
(b)



(b)

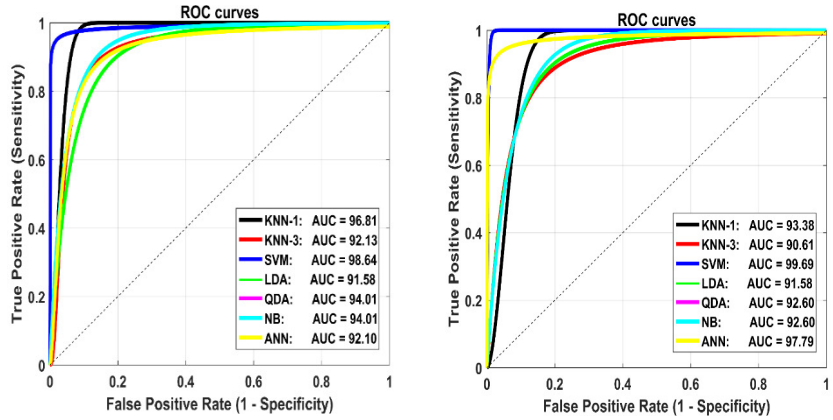


(d)

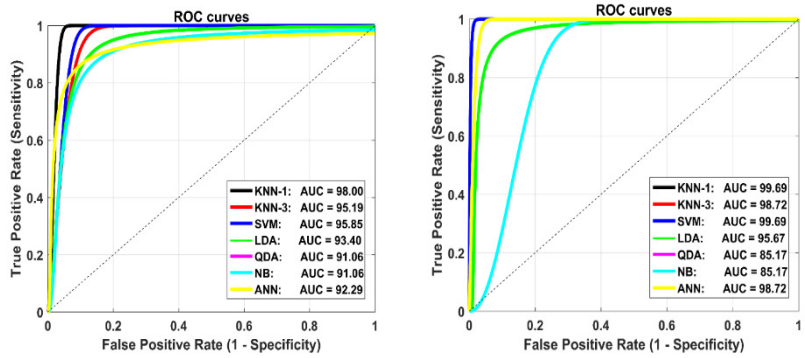


(e) (f)

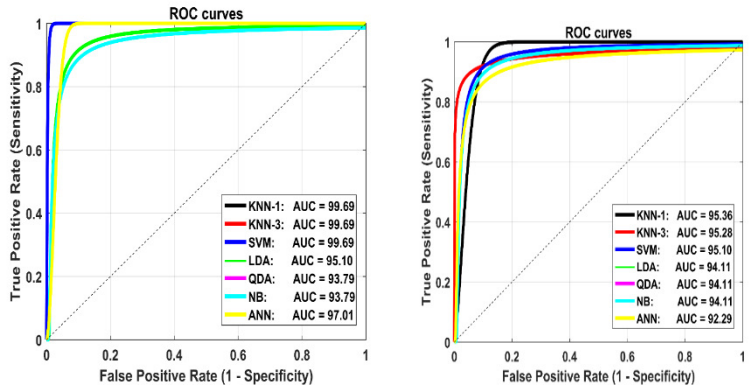
Figure. 4 ROC curves concerning all classifiers and involving all selection methods at L=8 which that (a): T-test, (b): KS and W-test, (c): SBS, (d): SFS, (e): SFFS, and (f): BBS.



(a) **(b)**



(b) **(d)**



(e) **(f)**

Figure. 5 ROC curves concerning all classifiers and involving all selection methods at L=32 which that (a): T-test, (b): KS and W-test, (c): SBS, (d): SFS, (e): SFFS, and (f): BBS.

Our findings show an Az value of (99.7%) with Accuracy Ac (100%), a sensitivity of (100%) and a specificity of (100%) for SVM classifier in both SFS/SFFS selection methods with both of quantization gray level (L) of GLCM at L=8 and L=32.

From all our results mentioned below, we can conclude some the following important points from for the proposed CADe detection system development such as:

1- SVM classifier achieved the best possible results with all feature selection methods with gray at L=32 except SBS method wherein KNN=1 classifier was the best in this case.

2- The worst performance was achieved by QDA and NB classifiers with most feature selection methods at L=32 except t-test wherein LDA classifier was the worst.

3- Features selected by SFS/SFFS methods yielded the best classification results for most classifiers at L=8 and 32, except QDA and NB, which showed the worst findings.

4- When the number of dataset is increased the performance is improved as well

5- Using SFFS, SBS, BBS and SFS techniques gave better results compared to the statistical methods

6- All classifiers were observed to perform better when the quantization gray level of GLCM was at L=32 as compared to another level at L=8.

7- The time consumption for PR-Tool methods (SBS, SFS, SFFS and BBS) need a long time to select the most powerful features, while statistical methods (T-test, KS-test and W-test) need a short time with both (L=8) and (L=32) in the CAD system.

5. Conclusions

It is a well-known fact that the ability of detecting breast cancer at an early stage assumes great importance in the context of augmenting the rate of survival as well and bringing improvements in the likelihood of ensuring the provision of adequate treatment that yields the

best possible results. In this context, mammography is ubiquitously recognized for being the most acceptable tool that facilitates timely detection; however, its sensitivity and efficacy is not impervious to the expertise of the radiologist of the equality of image.

Against this backdrop and in order to resolve the challenges mentioned above, we proposed a computer-aided detection (CAD) mechanism. The system was designed to perform a recognition task; more specifically, we used the CADe system for the purpose off identifying abnormalities in breast lesions in a timely manner. The MIAS database was used to develop and test the CADe system. A combination of several features extracted from its ROI was used the CAD system. In addition, CADe system made use of relevant classifiers (KNN-1, KNN-3, SVM, LDA, QDA, NB and ANN) for classification stage.

The best performance of the proposed CADe system has been achieved by SVM classifier with most feature selection methods and especially with SFS and SFFS methods where, in this case, all samples were correctly classified. The performance of all classifiers is the best by using SFS method. Also, we can conclude that the quantization gray level of GLCM is a very significant factor to get robust high order features where the results are better with L equal to the size of ROI. Using an enormous number of several features assist the CAD system to be strong enough to distinguish between the different tissues.

6. References

[1] T. m.-M. d. o. mammograms". (march). http://peipa.essex.ac.uk/info/mias.html.

[2] T. Wu, R. H. Moore, and D. B. Kopans, "Multi-threshold peripheral equalization method and apparatus for digital mammography and breast tomosynthesis," ed: Google Patents, 2010.

[3] N. Vállez, G. Bueno, O. Déniz, J. Dorado, J. A. Seoane, A. Pazos, *et al.*, "Breast density classification to reduce false positives in CADe systems," *Computer methods and programs in biomedicine,* vol. 113, pp. 569-584, 2014.

[4] G. Bueno, N. Vállez, O. Déniz, P. Esteve, M. A. Rienda, M. Arias, *et al.*, "Automatic breast parenchymal density classification integrated into a CADe system," *International journal of computer assisted radiology and surgery,* vol. 6, pp. 309-318, 2011.

[5] D. Guliato, R. M. Rangayyan, W. A. Carnielli, and J. L. Desautels, "Segmentation of breast tumors in mammograms using fuzzy sets," *Journal of Electronic Imaging,* vol. 12, pp. 369-378, 2003.

[6] D. Wei, H. P. Chan, M. A. Helvie, B. Sahiner, N. Petrick, D. D. Adler, *et al.*, "Classification of mass and normal breast tissue on digital mammograms: multiresolution texture analysis," *Medical Physics,* vol. 22, pp. 1501-1513, 1995.

[7] A. Oliver, X. Lladó, J. Martí, R. Martí, and J. Freixenet, "False positive reduction in breast mass detection using two-dimensional PCA," *Pattern Recognition and Image Analysis,* pp. 154-161, 2007.

[8] A. Omara, A. S. Mohamed, A.-B. M. Youssef, and Y. M. Kadah, "Computer aided diagnosis in digital mammography," in *the third Cairo International Biomedical Engineering Conference, CIBEC*, 2006.

[9] N. R. Mudigonda, R. M. Rangayyan, and J. L. Desautels, "Detection of breast masses in mammograms by density slicing and texture flow-field analysis," *IEEE Transactions on Medical Imaging,* vol. 20, pp. 1215-1227, 2001.
[10] J. Dheeba, N. A. Singh, and S. T. Selvi, "Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach," *Journal of biomedical informatics,* vol. 49, pp. 45-52, 2014.
[11] A. Bosch, X. Munoz, A. Oliver, and J. Marti, "Modeling and classifying breast tissue density in mammograms," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, 2006, pp. 1552-1558.
[12] J. Haddadnia, M. Ahmadi, and K. Faez, "An efficient feature extraction method with pseudo-Zernike moment in RBF neural network-based human face recognition system," *EURASIP Journal on Advances in Signal Processing,* vol. 2003, p. 267692, 2003.
[13] P. Juszczak, D. Tax, and R. P. Duin, "Feature scaling in support vector data description," in *Proc. ASCI*, 2002, pp. 95-102.
[14] A. Malich, D. R. Fischer, and J. Böttcher, "CAD for mammography: the technique, results, current role and further developments," *European radiology,* vol. 16, p. 1449, 2006.
[15] M. Hollander and D. A. Wolfe, "Nonparametric statistical methods," 1999.
[16] A. W. Whitney, "A direct method of nonparametric measurement selection," *IEEE Transactions on Computers,* vol. 100, pp. 1100-1103, 1971.
[17] T. Wu, R. H. Moore, and D. B. Kopans, "Multi-threshold peripheral equalization method and apparatus for digital mammography and breast tomosynthesis," ed: Google Patents, 2010.