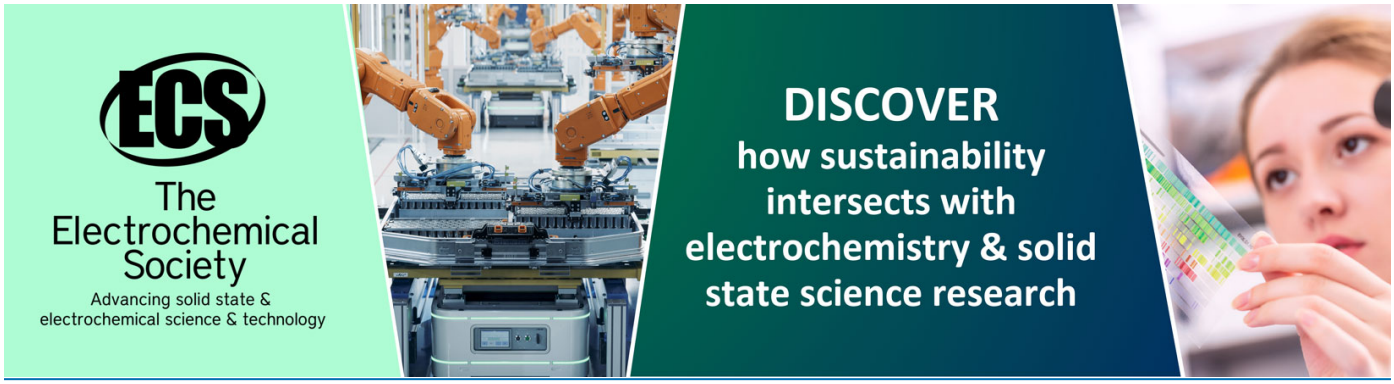
Analysis of Tissue Abnormality in Mammography Images Using Gray Level Co-occurrence Matrix Method



Analysis of Tissue Abnormality in Mammography Images Using Gray Level Co-occurrence Matrix Method

Mohammed Y. Kamil * **and Abdul-Lateef A. Jassam**

College of Sciences, Mustansiriyah University, Baghdad, Iraq.
* Email: m80y98@uomustansiriyah.edu.iq, ORCID: orcid.org/0000-0001-5709-2549

Abstract. One of the dangerous diseases is breast cancer, which threatens women and men to the same extent. But women are more affected by this disease. Computer-Aided Diagnosis (CAD) is the optimal method for the early detection of breast cancer. It can reduce the false positives in radiologist diagnosis, which leads to reduce the death-rate. This paper presents a feature extraction technique with mammography images to breast mass recognition. Then, distinguishing normal tissue and abnormal breast masses. The mini-MIAS database of mammograms was used in this paper. Gray Level Co-occurrence Matrix is the method that was used to extract features from the region of interest. The best sensitivity, specificity, and accuracy are observed with a k nearest neighbor classifier.

1. Introduction

Breast cancer is the second most common type of cancer in the world, according to the World Health Organization. It caused half a million deaths in the year, with more than 1.5 million diagnosed cases in 2010[1]. One of the common risk factors for breast cancer is genetics, breast density, family profile, and age[2]. The strongest risk factor for breast cancer is breast density (the amount of dense tissue in the breast). Breast cancer is more likely to develop as breast density increases. The relationship between the age of women and breast density where older women usually have less dense breasts than younger women[3]. Breast cancer is formed in the tissues of the breast and has a base in the channels (tubes holding the milk in the nipple) and lobes (glands resulting in milk). It occurs in both individuals, although breast cancer is not common in males [4]. Mammography is an X-ray carrying a low-dose. It depicts the inside body of the breast. Mammography is accurate, but it is very similar to most therapeutic tests. Screening after menopause is more accurate before menopause [5]. Medical centers around the world are having difficulty analyzing the growing volume of X-ray mammography[6]. As a result, many computer-aided diagnostic systems (CAD) have been proposed to assist the radiologist in analyzing the mammogram or reduce the number of diagnostic errors [7].

It is no secret that the world of medical imaging has evolved beyond all expectations, especially with the introduction of modern technology and computer to this area until the emergence of digital devices that rely mainly on the computer [8]. A Mammography is a vital tool for the early diagnosis of breast cancer [9]. The radiograph is an x-ray and creates detailed images of the breast. X-rays do not see all breast cancers[10]. Even when high-resolution mammograms are present, there are hidden parts of the intense tissue, making the process of image interpretation difficult for a radiologist.[11]Mammography is the key to knowing breast cancer. It is a low-energy x-ray that passes through the compact breast. The image is classified by two types of x-rays according to the first view method: the “Medio-Lateral Oblique” (MLO) view and the “Crania-Caudal" (CC) view [3].

Many researchers studied the breast tissue via mammograms. M. M. Fathima et al. (2013) [12] regions of interest are determined by fragmentation of the threshold. Class-grade text features are derived, and

gray-level interference features are grayed out. Support vector machine (SVM) was used Classified results of the query image were obtained based on the trained structure. Mammography data from the (MAIS) mammography association were used with 322 images available with information on the diagnosis of the doctor. J. Sharma et al. (2014)[13] propose a method to mass detect based on the statistical features extracted from the GLCM of the mammogram. Based on mass treatment, the undesirable details are excluded and the features are extracted using GLCM. The features are calculated and analyzed to identify the affected areas. V. Nguyen et al. (2015) [14] apply to GLCM features are extracted from detected regions of interest (ROI). The average sequential sequence defines a typical subset of 8 features of the full feature. Also, the SVM is utilized to classify normal or abnormal regions. At the expense of the set of databases investment discovered from the MIAS database, the proposed method achieves AZ = 0.938. V. Gaike et al. (2015) [15] use GLCM in remote sensing images to detect breast cancer. They used GLCM first and second features only. Higher ranks have not been used to detect malignant diseases in breast tissue images. S. Ray et al. (2016)[16] propose the algorithm consists of two components: (1) the adaptive selection of the topical areas of interest; and (2) the advantage of extracting the Haralick tissue through Gray- Level Co-Occurrence Matrices (GLCM) The highest ROC performance value for AUC = 0.77. M. Abdel-Nasser et al. (2017) [17] propose a CAD system consists of four stages feature extraction, classification, extraction of the region of interest and super-resolution computation. Evaluated the performance of five texture methods with the proposed CAD system: local binary, histogram of oriented gradients, phase congruency based local binary pattern and gray level co-occurrence matrix features. M. A. Berber (2018) [18] use seven texture features for the GLCM method and applied them to sub-images to improve their performance.

Mammograms are actually one of the most efficient methods for early diagnosis of breast cancer. The main aim of this project is improving the performance of breast mass detection for reducing the false positives to decrease the number of needless biopsies — this leads to helping radiologists to detect abnormalities in breast screening images. The key contribution is to use a region of interest inside the region suspicion, which bounded by the radiologist.

2. Materials And Methods

The texture is an important distinction for the understanding of image processing. The texture holds essential information that is used for the explanation and understands of images. Texture refers to an arrangement of the basic elements of an image and the spatial interrelationships [19]. The GLCM method is a very robust statistical descriptor in medical image processing. One important application of image texture is used to classify images to regions that provide information in the spatial mode of intensities or colors. The texture is a repeating pattern of local variations in image intensity [20].

The GLCM method included some of the spatial measures, as energy, contrast, correlation, and homogeneity [21].

The extracted Haralick textual features are as follows [22]: Energy: “Energy which is low when all elements in the GLCM are close to either 0 or 1 and high when the GLCM has equal values or pixels are similar”. Energy is defined as:

$$\text{Energy} = \sum_{i} \sum_{j} p(i,j) \tag{1}$$

Contrast: “Contrast measures the spatial frequency of an image. It is the difference between the highest and the lowest values of the contiguous set of pixels”. Contrast is defined as:

$$\text{Contrast} = \sum_{i} \sum_{j} (i-j)^2 p(i,j) \tag{2}$$

Correlation: “Correlation feature is a measure of gray tone linear dependencies in the image”. Correlation is defined as:

$$\text{Correlation} = \frac{\sum_i \sum_j (ij)p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y} \tag{3}$$

where μ_x, μ_y $\sigma_x$and σ_y are the means and standard deviations of p_x and p_Y.

Homogeneity: This statistic is also called an Inverse Difference moment. It measures "the image homogeneity and assumes larger values for smaller gray tone differences in pair elements". It has a maximum value when all the elements in the image are the same. Homogeneity is defined as:

$$\text{Homogeneity} = \sum_i \sum_j \frac{1}{1+(i-j)^2} p(i,j) \quad (4)$$

The mammographic images have used from the MIAS database, which consists of 322 images. That is possessed from 161 different ladies; inside this: 209 of them are of healthy breasts, 51 of them are malignant mass, and 62 are demonstrated as benign masses. Every mammogram has an image size of 1024×1024 pixels. The breast tissue classifies the digital images (dense-glandular, fatty-glandular, and fatty) and the kind of the masses (ill-define, normal, asymmetry, architectural distortion, speculated, well defined, and calcification).

Validation of the test is necessary for the computer detection system to improve an acceptable degree. Accuracy, sensitivity, and specificity are used as performance evaluation statistical measures. Sensitivity: "measures the proportion of images that contain a cancerous mass that has been classified correctly"[23].

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (5)$$

Specificity: "measures the proportion of images containing a cancerous mass that has been incorrectly classified".

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (6)$$

Accuracy: "measures the ratio of correctly classified pixels to the entire area of the region of interest".

$$\text{Accuracy} = \frac{TP+TN}{TN+FN+TP+FP} \quad (7)$$

3. Results and Discussion

The GLCM algorithm is an important algorithm in extracting features that can be summarized in this work with the two steps:

Step 1: We extract the region of interest (ROI) (e.g.25x25) pixels by hand from all breast cancer radiographs. Choose the area of interest just behind the nipple from the central breast area, as shown in figure (1). Because they are denser and the areas most likely to develop breast cancer [24]. The pectoral muscles are not taken in the region of interest (ROI) because they are an obstacle in the detection. The algorithm used did not include the entire image but only the region of interest (ROI). Therefore no segmentation step was included in the proposed method.

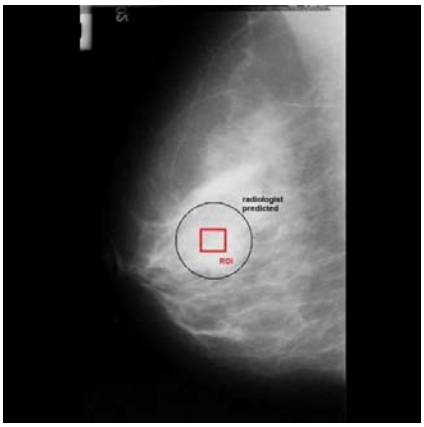


Figure 1. Example of an ROI (25 x 25) extracted from a breast image.

Step 2: Features are extracted from each area of interest. To get a superior rating rate and see the best common features. The features were obtained from the gray level image. The spatial relationship of the

pixel takes the ROI and the pixel to the right. For features, the features selected were the GLCM features applied to each area of interest. We will come up with GLCM features as follows:
1. Create a matrix by specifying the gray plane number as the size of the matrix.
2. Determine the relationship between reference and adjacent pixels used. The distance between pixels (1, 2, 3, …., 30) and angles were 0, 45, 90, 135.
3. Compute the occurrences and fill the matrix.
4. Get an asymmetrical matrix by adding the matrix to its transpose.
5. Transform the normalized matrix to its probabilities.
6. Compute texture features that are proposed by Haralick.
Two different images are selected, one is normal, and the other is abnormal. These two images are chosen as a sample to show the algorithm results and extract the features from it. Figure (2) shows two images are presented, one diagnosed with breast cancer (mdb165) and other normal (mdb054). We have calculated the features (energy, homogeneity, contrast, correlation), by four angles (0˚, 45˚, 90˚, 135˚) and a region of interest (10 × 10), (15 × 15), (20 × 20), (25 × 25) and (30 × 30). also, the distance changes according to the area of interest (1,2, ..., 30).

(a) (b)

Figure 2. Samples of (a) abnormal image (mdb165), (b) normal image (mdb054).

In this paragraph, a sample is shown of the results obtained. It displays a comparison of the results of extracted features for the region of interest (30×30). Figure (3) is represented the relationship between the energy feature and the distance to the region of interest at different angles.
Figure (3) is illustrated the energy can use to distinguish between abnormal tissue (mdb165) and normal tissue (mdb054). It is observed the values of the abnormal tissue differed from the values of the normal tissue which indicates. Therefore, the energy can be used to distinguish between them. Where the energy values range between (0 - 1), and if the energy value is equal to (1), this means that the image is constant. Also, it is observed that when the angle was 135˚, the distinction was clearly without a change of distance.
Figure (4) is represented the relationship between the contrast feature and the distance to the region of interest (30 × 30) at different angles. Figure (4) is illustrated the contrast can use to distinguish between abnormal tissue and normal tissue. So the values of the abnormal tissue differed from the values of the normal tissue. If the value of the contrast is equal to (0), this means that the image is constant. It is observed when the value of the distance is greater than 10, the difference between the abnormal and normal tissues is clearly. Figure (5) is represented the relationship between the correlation feature and the distance to the region of interest (30 × 30) at four different angles. Figure (5) is observed the correlation can use to distinguish between the abnormal tissue (mdb165) and normal tissue (mdb054). The energy values range between (-1, 1), and if the energy value is equal to (NaN), this means that the image is constant. Figure (6) is shown the relationship between the homogeneity feature and the distances taken to the region of interest (30×30) at four different angles. Figure (6) is illustrated the homogeneity can use to distinguish between abnormal tissue and normal tissue. it is observed the distance is greater than (10), so the distinction is very clear.

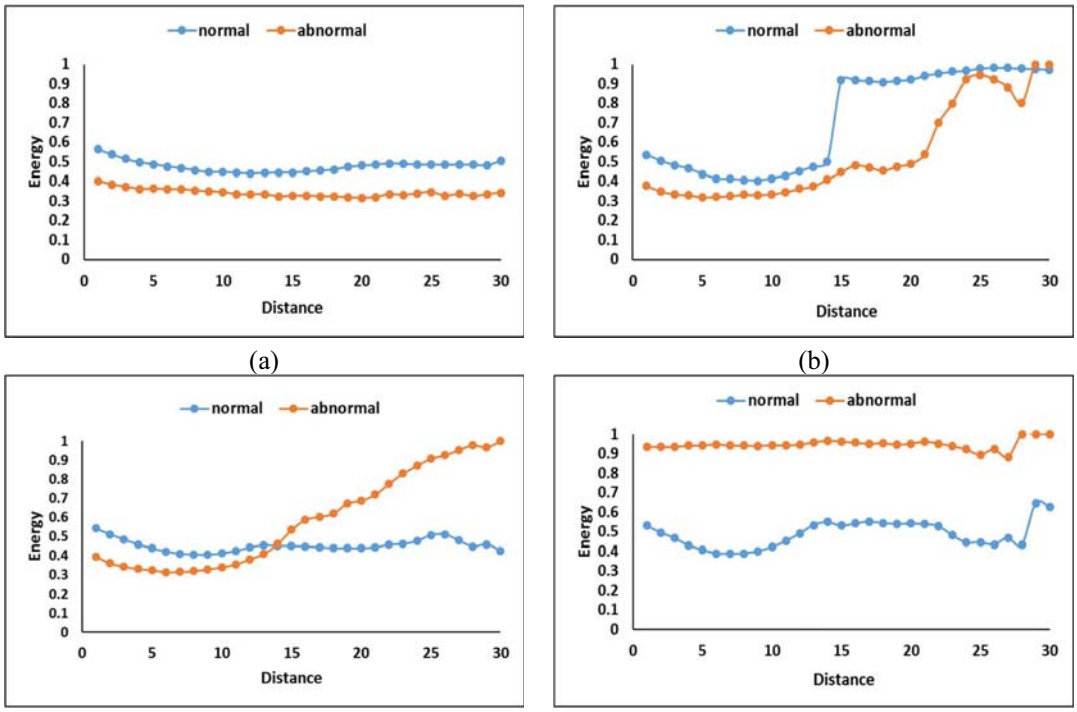


Figure 3. Relationship between Energy feature and distance for ROI (30×30): (a) angle = 0˚. (b) angle = 45˚. (c) angle = 90˚. (d) angle = 135˚.

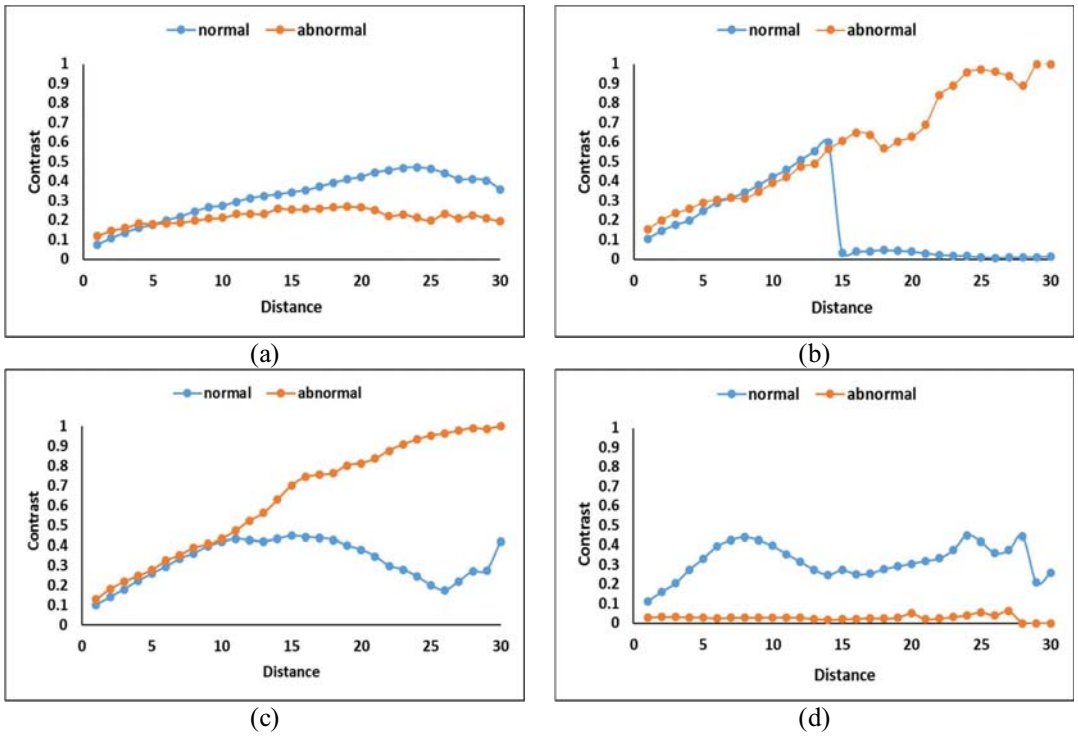


Figure 4. Relationship between contrast feature and distance for ROI (30×30): (a) angle = 0˚. (b) angle = 45˚. (c) angle = 90˚. (d) angle = 135˚.

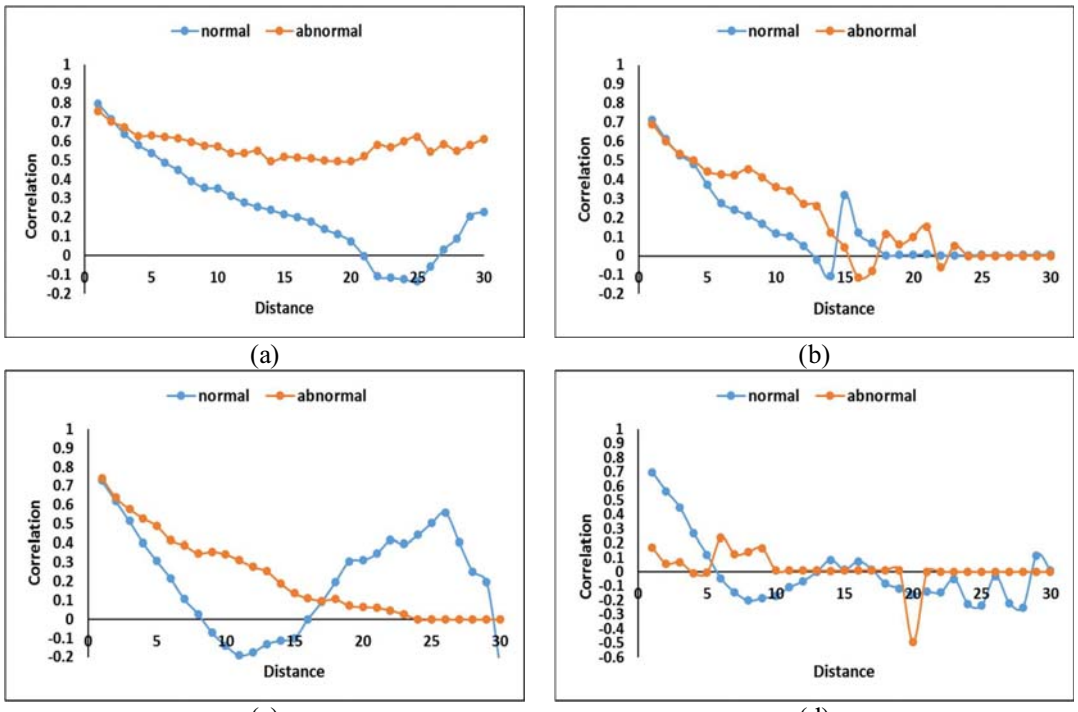


Figure 5. Relationship between correlation feature and distance for ROI (30×30): (a) angle = 0˚. (b) angle = 45˚. (c) angle = 90˚. (d) angle = 135˚.

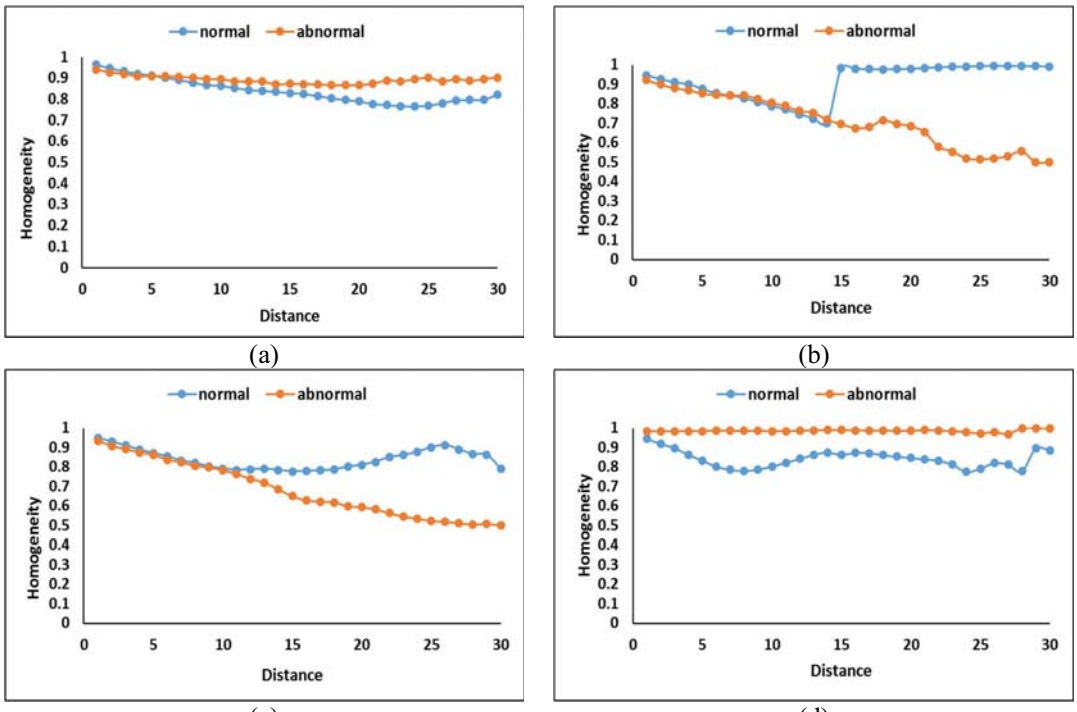


Figure 6. Relationship between Homogeneity feature and distance for ROI (30×30): (a) angle = 0˚ (b) angle = 45˚. (c) angle = 90˚. (d) angle = 135˚.

After obtaining the results of the four features of all the images used in the database. The classification is used for these features to determine the type of tumor (benign and malignant). then, to know the best features that affect the classification of tissues.
A classification method is used k nearest neighbor (KNN), the number of neighbors is equal to 10, and a distance metric is a Euclidean method.
Table (1) shows accuracy, sensitivity and specificity coefficients for all regions calculated for all images. Table (1) is observed that the highest accuracy value is (86.1) at the region of interest (30) at the angle (0˚). While the lowest value of the accuracy obtained is (66.5) at a region of interest (10) at an angle (135˚). Therefore, the higher values for accuracy are at the (0˚) and (90˚) angles better than the values at (45˚) and (135˚).
Also, it is observed that the highest value of the sensitivity is (92) at the region of interest (15) and the angle (0˚). The lowest sensitivity value obtained is (63) at a region of interest (20) at an angle (135˚). The highest value of the specificity is at a region of interest (20) and for the angle (90˚), the value is (86). While the lowest value is at (15) for the angle (45˚), the value is (51).

Table 1. Accuracy, sensitivity and specificity of breast mass detection with GLCM, using KNN classifier.

Accuracy				
angle / ROI	0˚	45˚	90˚	135˚
30 x 30	**86.1**	79.7	84.4	82.4
25 x 25	84	80.5	82.7	75.7
20 x 20	81.7	74.7	83.4	73.6
15 x 15	82.6	75	77.7	73.7
10 x 10	80.9	70.3	83.7	66.5
sensitivity				
angle / ROI	0˚	45˚	90˚	135˚
30 x 30	87	81	87	85
25 x 25	87	78	86	80
20 x 20	84	78	81	63
15 x 15	**92**	89	89	85
10 x 10	82	74	88	71
specificity				
angle / ROI	0˚	45˚	90˚	135˚
30 x 30	85	78	80	78
25 x 25	81	83	79	70
20 x 20	79	71	**86**	83
15 x 15	65	51	59	55
10 x 10	80	66	78	61

In this section, the implemented algorithm in this work is compared with related works to performance evaluate the work. Its comparison is based on the measure of accuracy and sensitivity. Also, the papers that used the MIAS database.
It observes table (2), the proposed algorithms have been applied to all dataset images, which gives good results compared with other authors. we have used KNN classifier because it appears the high ability to obtain a successful decision.

Table 2. comparison of the GLCM method in MIAS Dataset images for different classifier

Authors	year	Feature methods	Classifier	Dataset	Sensitivity	accuracy
Singh and Nagarajan [25]	2014	GLCM	SVM	MIAS	94.5%	-
Pratiwi, et al.[26]	2015	GLCM	Back-propagation Neural Network	MIAS	81.6%	79.98%
Sreedevi, et al.[27]	2016	GLCM	SVM	MIAS	-	83.5%
Berbar [18]	2017	GLCM and statistical features	SVM	MIAS	96.12%	97.89%
Proposed method		GLCM	KNN	MISA	92%	86.1%

4. Conclusions

Results of the implementation of the proposed algorithms and evaluation of the tests give some conclusions. Methods based on feature extraction in mammography images are effective in identifying tumors in breast cancer and determine the type of tumor. It observed that the best angle that can be adapted to distinguish between the abnormal tissues and normal tissues is 135˚. When increasing the value of d greater than 10, it observed that the distinction between abnormal tissue and normal tissue is clear up to the value of 30. It observed from the results that the best region of interest is (30×30). Also, it gives good results for the four angles. It observed that the best statistical variable that can be adapted to distinguish between the abnormal tissues and normal tissues is energy. It observed that the worst possible statistical variable in the GLCM feature extraction on mammography images is the correlation. The best classification to find out the type of tumor (benign and malignant) in this work is KNN.

From the advantages and disadvantages that have emerged in the proposed algorithms in this work, can be suggested future works as Breast tissue study in mammography images by using another feature extraction method as a local binary pattern. It is suggested to build a system as a computer-aided diagnosis based on feature extraction methods

References

[1] Jalalian A, Mashohor S B, Mahmud H R, Saripan M I B, Ramli A R B and Karasfi B 2013 Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review *Clinical imaging* **37** 420-6

[2] Arevalo J, González F A, Ramos-Pollán R, Oliveira J L and Lopez M A G 2015 Convolutional neural networks for mammography mass lesion classification. In: *2015 37th Annual international conference of the IEEE engineering in medicine and biology society (EMBC)*: IEEE) pp 797-800

[3] Abdel-Nasser M, Rashwan H A, Puig D and Moreno A 2015 Analysis of tissue abnormality and breast density in mammographic images using a uniform local directional pattern *Expert Systems with Applications* **42** 9499-511

[4] Singh N, Mohapatra A G and Kanungo G 2011 Breast cancer mass detection in mammograms using K-means and fuzzy C-means clustering *International Journal of Computer Applications* **22** 15-21

[5] Michaelson J, Satija S, Moore R, Weber G, Halpern E, Garland A, Puri D and Kopans D B 2002 The pattern of breast cancer screening utilization and its consequences *Cancer* **94** 37-43

[6] Kamil M Y 2016 Morphological gradient in brain magnetic resonance imaging based on intuitionistic fuzzy approach. In: *2016 Al-Sadeq International Conference on Multidisciplinary in IT and Communication Science and Applications (AIC-MITCSA)*: IEEE) pp 1-3
[7] Kamil M Y and Salih A M 2019 Mammography Images Segmentation via Fuzzy C-mean and K-mean *International Journal of Intelligent Engineering and Systems* **12** 22-9
[8] Samulski M, Hupse R, Boetes C, Mus R D, den Heeten G J and Karssemeijer N 2010 Using computer-aided detection in mammography as a decision support *European radiology* **20** 2323-30
[9] Salih A M and Kamil M Y 2018 Mammography Image Segmentation Based on Fuzzy Morphological Operations. In: *2018 1st Annual International Conference on Information and Sciences (AiCIS)*: IEEE) pp 40-4
[10] Bird R E, Wallace T W and Yankaskas B C 1992 Analysis of cancers missed at screening mammography *Radiology* **184** 613-7
[11] Motakis E, Ivshina A V and Kuznetsov V A 2009 Data-driven approach to predict survival of cancer patients *IEEE Engineering in Medicine and Biology Magazine* **28** 58-66
[12] Fathima M M, Manimegalai D and Thaiyalnayaki S 2013 Automatic detection of tumor subtype in mammograms based On GLCM and DWT features using SVM. In: *2013 International Conference on Information Communication and Embedded Systems (ICICES)*: IEEE) pp 809-13
[13] Sharma J, Rai J and Tewari R 2014 Co-occurrence Matrix and statistical features as an approach for mass classification. In: *Advances in Computing, Communications and Informatics (ICACCI, 2014 International Conference on*: IEEE) pp 2369-73
[14] Nguyen V, Nguyen D, Nguyen T, Phan V and Truong Q 2015 Filter-based feature selection and support vector machine for false positive reduction in computer-aided mass detection in mammograms. In: *Seventh International Conference on Machine Vision (ICMV 2014)*: International Society for Optics and Photonics) p 94451H
[15] Gaike V, Akhter N, Kale K and Deshmukh P 2015 Application of higher order glcm features on mammograms. In: *Electrical, Computer and Communication Technologies (ICECCT), 2015 IEEE International Conference on*: IEEE) pp 1-3
[16] Ray S, Keller B M, Chen J, Conant E F and Kontos D 2016 Parameter optimization of parenchymal texture analysis for prediction of false-positive recalls from screening mammography. In: *Medical Imaging 2016: Computer-Aided Diagnosis*: International Society for Optics and Photonics) p 97851Y
[17] Abdel-Nasser M, Melendez J, Moreno A, Omer O A and Puig D 2017 Breast tumor classification in ultrasound images using texture analysis and super-resolution methods *Engineering Applications of Artificial Intelligence* **59** 84-92
[18] Berbar M A 2018 Hybrid methods for feature extraction for breast masses classification *Egyptian informatics journal* **19** 63-73
[19] Abdel-Nasser M, Moreno A and Puig D 2016 Towards cost reduction of breast cancer diagnosis using mammography texture analysis *Journal of Experimental & Theoretical Artificial Intelligence* **28** 385-402
[20] Saleck M M, ElMoutaouakkil A and Mouçouf M 2017 Tumor Detection in Mammography Images Using Fuzzy C-means and GLCM Texture Features. In: *2017 14th International Conference on Computer Graphics, Imaging and Visualization*: IEEE) pp 122-5
[21] Gaike V, Mhaske R, Sonawane S, Akhter N and Deshmukh P D 2015 Clustering of breast cancer tumor using third order GLCM feature. In: *Green Computing and Internet of Things (ICGCIoT), 2015 International Conference on*: IEEE) pp 318-22
[22] Ohmshankar S and Paul C K C 2014 Haralick fetaures based mammogram classification system. In: *Second International Conference on Current Trends In Engineering and Technology-ICCTET 2014*: IEEE) pp 409-13
[23] Nabeel F. Lattoofi, Israa F. Al-sharuee, Mohammed Y. Kamil, Ayoob H. Obaid, Aya A. Mahidi, Ammar A. Omar and Saleh A k 2019 Melanoma Skin Cancer Detection Based on ABCD Rule *First International Conference of Computer and Applied Sciences (1st CAS2019)*

[24] Li H, Giger M L, Huo Z, Olopade O I, Lan L, Weber B L and Bonta I 2004 Computerized analysis of mammographic parenchymal patterns for assessing breast cancer risk: effect of ROI size and location *Medical Physics* **31** 549-55

[25] Singh W J and Nagarajan B 2013 Automatic diagnosis of mammographic abnormalities based on hybrid features with learning classifier *Computer methods in biomechanics and biomedical engineering* **16** 758-67

[26] Pratiwi M, Harefa J and Nanda S 2015 Mammograms classification using gray-level co-occurrence matrix and radial basis function neural network *Procedia Computer Science* **59** 83-91

[27] Sreedevi S, Mathew T J and Sherly E 2016 Computerized classification of malignant and normal microcalcifications on mammograms: Using soft set theory. In: *Information Science (ICIS), International Conference on*: IEEE) pp 131-7