

Diagnosis of Breast Cancer in Digital Mammograms by Using GLCM and Fuzzy Logic

Shaker K. Ali
Thi-Qar University,
Computer and mathematic collage,
Computer department,
Shalynar@hotmail.com

Abstract: *In this paper we suggest algorithm to diagnosis the breast cancer for three cases (normal, cancer and benign) by using GLCM (grey level Co-occurrence Matrix) and fuzzy logic. The algorithm is used digital mammograms pictures for diagnosis the breast cancer in two steps; first check the pictures if it is normal or abnormal, second step when the picture is abnormal it must be decide if it is cancer or benign) by extracting the features for each case then these features will be input to fuzzy logic to give the final decision, there are six features used as six input for fuzzy logic which give the result for the first step (normal and abnormal) and there are two features which used in second step (cancer or benign) the result from our algorithm gives more accuracy and it is benefit for diagnosis .Our data base are collected from University of South Florida, USA.*

المخلص: في هذا البحث صممنا خوارزميه لتمييز انواع الاورام التي تصيب الصدر والتي استخدمت فيها وكذلك التقنية الضبابية للتحليل والتمييز. GLCM الصور المأخوذة بواسطة الاشعة السينية. استخدمنا مصفوفة الخوارزمية هي تتكون من مرحلتين , الاولى هي لمعرفة ان الصورة مصابة ام لا , والثانية هي اذا كانت مصابة تقوم باستخلاص الخواص حيث تم تعيين 6 خواص في GLCM هل هي مصابة بمرض سرطان حميد ام خبيث . المرحلة الاولى أي لمعرفة فيما اذا كانت الصور مصابة ام لا , اما المرحلة الثانية فتم استخلاص صفتين لتحديد فيما اذا كانت الصورة مصابة بمرض سرطان خبيث ام حميد. النتائج التي حصلنا عليها كانت عالية الدقة وهي مفيدة جدا في حقل الطب , حيث تم اعتماد الصور القياسية من قاعدة بيانات جامعة جنوب فلوريدا الامريكية .

Keywords: *Diseases diagnosis, Mammograms, GLCM, Fuzzy logic, Feature Extraction.*

1. Introduction

Cancer detection has become a hot cake of research in pattern recognition community and disease diagnoses. Breast cancer is one of the leading causes of cancer mortality among women in the Iraq and the world. Breast cancer is a potentially fatal condition; early diagnosis of disease can lead to successful treatment [1], and save many lives.

There are various methods for breast cancer screening like clinical and self-breast exams, genetic screening, ultrasound, magnetic resonance imaging and mammography. Mammography is the best and most efficient method for detecting breast cancer at the early stage. Mammography is a specific type of imaging that uses a low dose x-ray system to examine breast [2].

In our paper we first classification if the case is normal or abnormal, if the case is abnormal which are classified into two types: masses and calcification. The calcification appears as small calcium deposit. Calcification consists of two classes: micro calcification and macro calcification. Macro calcifications are large calcium deposit associated with non-cancerous pattern. Micro calcifications are small calcium deposit. Micro calcification is a sign of breast diseases. Masses are different in shape and appearance [3].

Analysis by using Gray Level Co-occurrence Matrix (GLCM) is needed to find the features that will be the key for classification. GLCM is the most classical second-order statistical method for texture analysis [4].

2. Gray Level Co-occurrence Matrix (GLCM)

Haralick et al first introduced the use of co-occurrence probabilities using GLCM for extracting various texture features. GLCM is also called as Gray level Dependency Matrix. It is defined as “A two dimensional histogram of gray levels for a pair of pixels, which are separated by a fixed spatial relationship.” GLCM of an image is computed using a displacement vector d , defined by its radius δ and orientation θ [5]. Consider a 4x4 image represented by figure (1-a) with four gray-tone values 0 through 3. A generalized GLCM for that image is shown in figure (1-b) where $\#(i,j)$ stands for number of times gray tones i and j have been neighbors satisfying the condition stated by displacement vector d [6]

Haralick extracted thirteen texture features from GLCM for an image. These features are as follows [6,7]:

1. **Contrast** is a measure of intensity or gray-level variations between the reference pixel and its neighbor. In the visual perception of the real world, contrast is determined by the difference in the color and brightness of the object and other objects within the same field of view.

$$f_1 = \sum_{n=0}^{N-1} n^2 \{ \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} p_{d,\theta}(i,j)^2 \} \dots\dots\dots(1)$$

2. **Correlation** feature shows the linear dependency of gray level values in the co-occurrence matrix. It presents how a reference pixel is related to its neighbor, 0 is uncorrelated, 1 is perfectly correlated.

$$f_2 = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} p_{d,\theta}(i,j) \frac{(i-\mu_x)(j-\mu_y)}{\sigma_x\sigma_y} \dots\dots\dots(2)$$

Where μ_x, μ_y and σ_x, σ_y are the means and standard deviations of p_x and p_y .

3. **Energy**. It measures the uniformity of an image. When pixels are very similar, the ASM value will be large.

$$f_3 = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} p_{d,\theta}(i,j)^2 \dots\dots\dots(3)$$

$$\mu_x = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} i * p_{d,\theta}(i,j) \quad \mu_y = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} j * p_{d,\theta}(i,j) \dots\dots\dots(4)$$

$$\sigma_x = \sqrt{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i - \mu)^2 * p_{d,\theta}(i,j)} \quad \sigma_y = \sqrt{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (j - \mu)^2 * p_{d,\theta}(i,j)} \dots\dots\dots(5)$$

4. **Homogeneity**: that measures the local homogeneity of an image. IDM feature obtains the measures of the closeness of the distribution of the GLCM elements to the GLCM diagonal.

$$f_4 = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{1}{1+(i-j)^2} p_{d,\theta}(i,j) \dots\dots\dots(6)$$

0	0	1	1
0	0	1	1
0	2	2	2
2	2	3	3

Test image

Gray tone	0	1	2	3
0	#(0,0)	#(0,1)	#(0,2)	#(0,3)
1	#(1,0)	#(1,1)	#(1,2)	#(1,3)
2	#(2,0)	#(2,1)	#(2,2)	#(2,3)
3	#(3,0)	#(3,1)	#(3,2)	#(3,3)

General form of GLCM

Figure (1-a) the test image for GLCM

4	2	1	0
2	4	0	0
1	0	6	1
0	0	1	2

GLCM for $\delta=1$ and $\theta=0^\circ$

4	1	0	0
1	2	2	0
0	2	4	1
0	0	1	0

GLCM for $\delta=1$ and $\theta=45^\circ$

6	0	2	0
0	4	2	0
2	2	2	2
0	0	2	0

GLCM for $\delta=1$ and $\theta=90^\circ$

2	1	3	0
1	2	1	0
3	1	0	2
0	0	2	0

GLCM for $\delta=1$ and $\theta=135^\circ$

Figure (1-b): The four GLCM for angles equal to 0° , 45° , 90° and 135°

3. Algorithm

Our algorithm contain the following steps and as shown in figure (2)

1. Input pictures step: the picture that we collected from the (University of South Florida Digital Mammography) by taken 100 samples for training and 20 sample for testing.
2. Pre- processing step: we need to make all the pictures has the same feature and has the same environments to be get the real result for all the samples.
3. Extract the features steps: Extract the features by using GLCM (Grey Level Co-occurrence Matrix) to extract the features that obtained from 16 vector that is contain (Homogenous, Energy, Contrast, correlation) in (0,45,90,135) degrees.
4. Fuzzy logic step: in Fuzzy logic step we need to decide the kind of the disease (Normal, Cancer and Benign), but we do that in two steps, first detect if it normal or abnormal, second if it is abnormal then find new feature to detect if it (Cancer and Benign).
5. Final decision: to give the final decision about the disease if normal or abnormal (Cancer or Benign).

4. Fuzzy Modelling Systems Structure

For multi-inputs single-output (MISO) fuzzy modeling systems, the description of fuzzy rules can be illustrated as follows: $R(i)$: if X is x_j then Y is y_i , $i = \{1,2,\dots,m\}$ where $X = (x_1, x_2, \dots, x_n)$ are linguistic variables, represents the i th rule, x_j are fuzzy sets of the input vector(x) in the premise part and y_i is a real number in the consequent part. m is the number of total rules, the definition of the fuzzy set is described by the following membership function [8].

$$ME_i = \exp\left(-\left(\frac{(x_i - a_{i1})^2}{b_{i1}^2} + \dots + \frac{(x_n - a_{in})^2}{b_{in}^2}\right)\right) \dots\dots\dots(7)$$

This can be considered that there is a hyper-ellipsoid in an n-dimensional space. where (ai1,ai2,...,ain) is the center of the hyper-ellipsoid and (bij) is the length of the jth principal axis of the hyper-ellipsoid.

The simple weighted average defuzzier is taken to convert fuzzy domain into real value. If the firing strength of the premise part in ith rule is deserved, the actual output Y of the fuzzy system can be calculated by

$$Y = \frac{\sum_{i=1}^m ME_i(x) \cdot y_i}{\sum_{i=1}^m ME_i(x)} \dots\dots\dots(8)$$

There are many types of fuzzy neural network, for the forward multi-layer fuzzy neural network, the typical ones are Mamdani model, Sugeno model and Tsukamoto model f. The simple and widely used Sugeno model is applied in this paper. For simplicity, we assume that there are two rules in the fuzzy reasoning system of Sugeno model to know if the pictures normal or abnormal as following:

R1: IF G0(h)[212-438] & G1(h)[90-251] & G2(co)[0.485-0.576] & G2(cr)[0.0104-0.0196] & G3(cr)[0.0121- 0.0217] then output normal

R2:IF G0(h)[77-230] & G1(h)[36-68] & G2(co)[0.589-0.663] & G2(cr)[0.055- 0.234] & G3(co)[0.589 - 0.663] & G3(cr)[0.0555 - 0.2732] then output abnormal

The following rules depend on the result in table (1 and 2) respectively.

And there are also two rules to know the case of the abnormal picture whither it cancer or benign as following:

R1:IF G0(h)[57 - 127]&G1(h)[134 - 127] then bening

R2:IF G0(h)[131 - 355]&G2(h)[130 - 353]then cancer

The following rules depend on the result in table (3 and4) respectively.

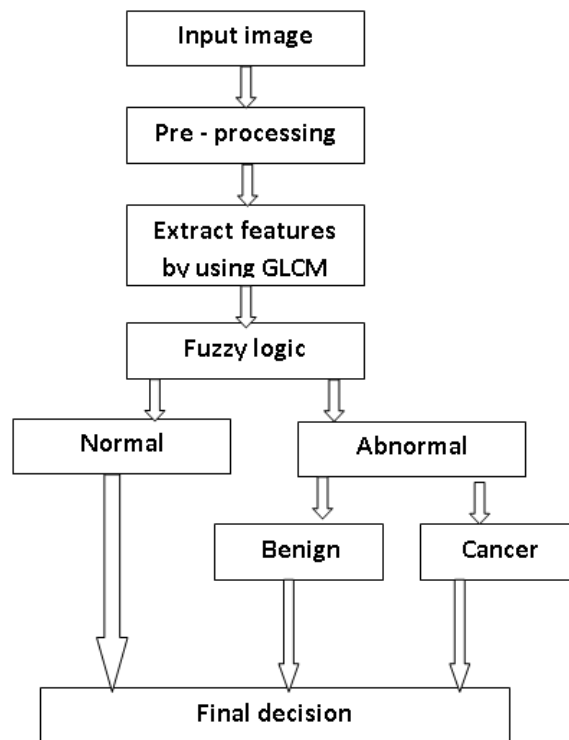


Figure (2). Algorithm

5. Experiment

Our experiment using the pictures which collected from South Florida University data base ,in our paper we collected 100 pictures for training stage and 20 pictures for testing stage to diagnosis if the pictures or cases are normal or abnormal, the abnormal has two cases; benign and cancer as shown in figures (3, 4 and 5) respectively, in first stage there are 5 steps which explain in algorithm section which starting from reading pictures and preprocessing steps then we must analysis the pictures to extract the features by using GLCM, and the output from GLCM are 16 features for each picture these are in 0 degree there are 4 features; Contrast ,Correlation, Energy and Homogeneity and in 45, 90 and 135 degree also there are 4 features for each one , so that is mean in total will be 16 features, but not all these features can used for classification to know if the picture is inflected or not (normal or abnormal) so we found only 6 features can be considered as classification features, but when the pictures are diagnosis as abnormal pictures that is mean they may be benign or cancer.

In first we take 100 pictures which is already know the case of the pictures (normal , cancer and benign) and build data base and made the rules of fuzzy logic by the rules which is mentioned in fuzzy logic section, from the data base we extracted the features which is be the key to diagnosis the pictures, and we found that there are 6 feature in first stage work to diagnosis as shown in table (1,2) and can be see clearly the difference between normal and abnormal as shown in figure (6,7).

Abnormal cases mean there two cases whither it is cancer or benign and we found there are two features works as shown in table (3, 4) and can be see clearly the difference between cancer and benign as shown in figure (8,9).

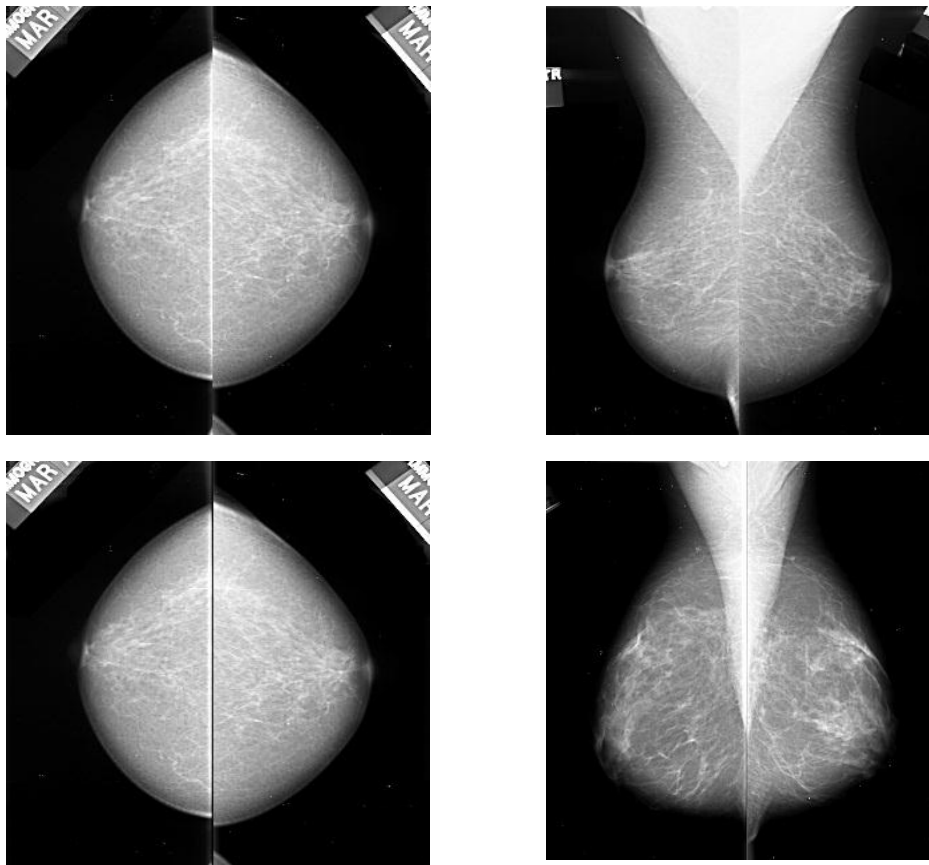


Figure 5).Normal type in both sides

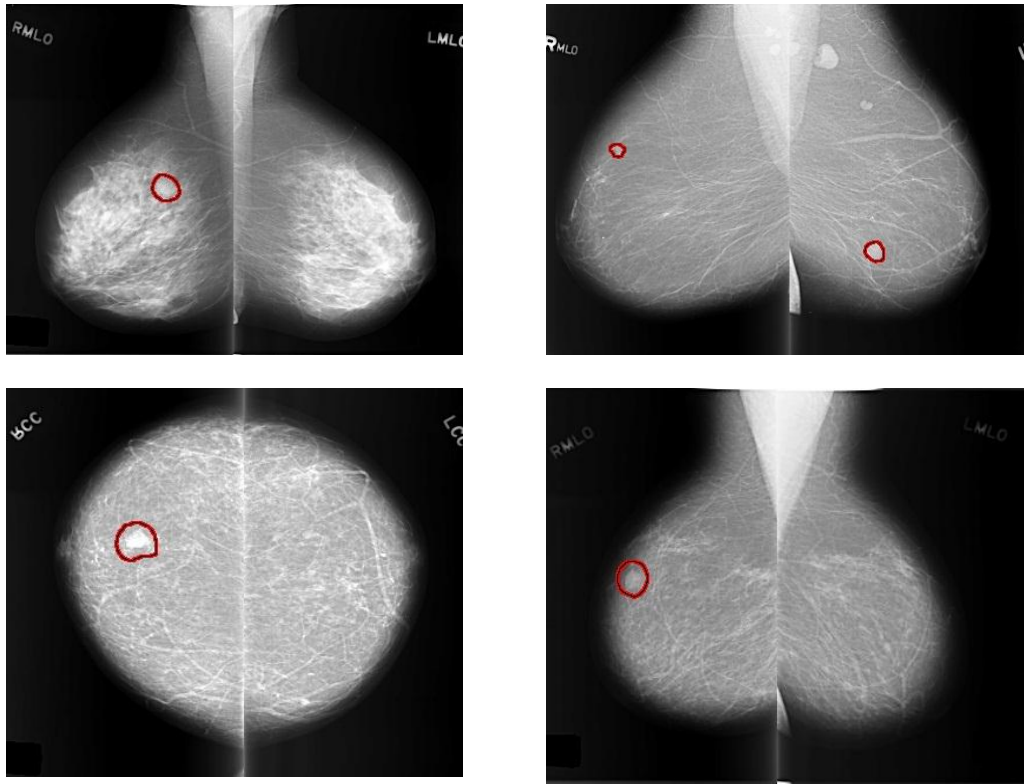


Figure (3). Benign type in both sides

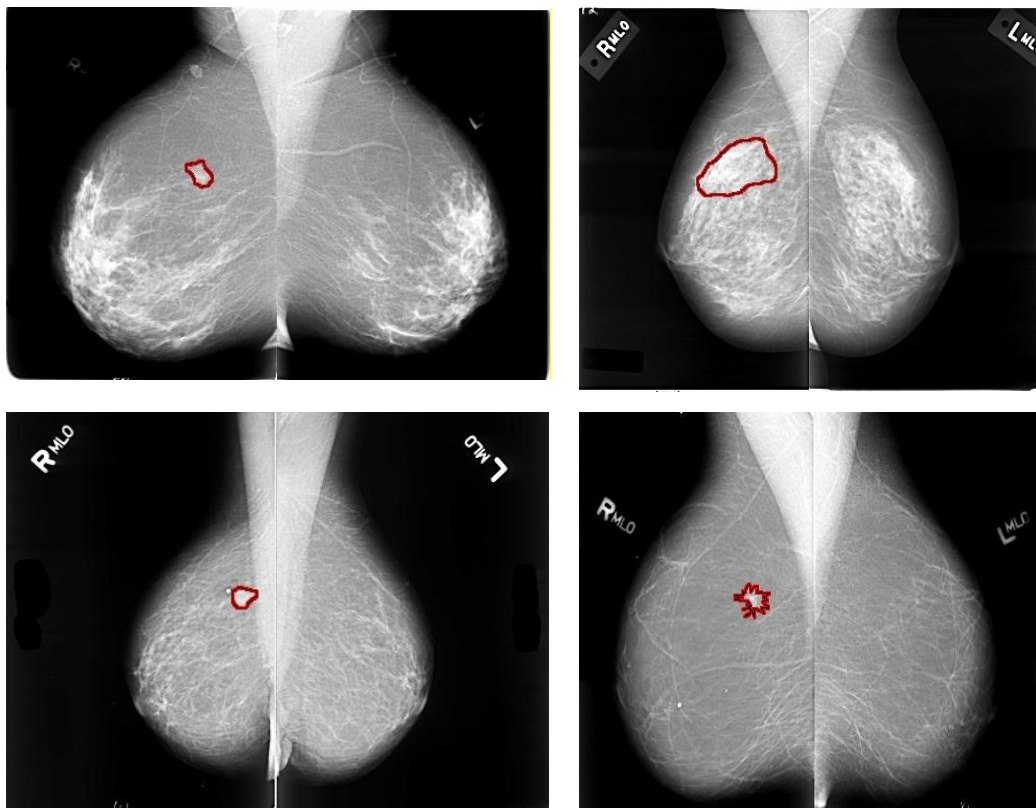


Figure (4) Cancer type in both sides

Table (1) Show the six features for normal cases

name	G0(h)	G1(h)	G2(co)	G2(cr)	G3(co)	G3(cr)
Pic 1	128.931	85.75941	0.63353	0.135645	0.63353	0.135645
Pic2	77.53324	46.53124	0.589002	0.055512	0.589002	0.055512
Pic 3	77.50981	46.41291	0.590589	0.055467	0.590589	0.055467
Pic 4	230.0862	86.23815	0.63306	0.13353	0.63306	0.13353
Pic 5	196.5568	36.38523	0.622907	0.193344	0.622907	0.193344
Pic 6	91.4123	65.96956	0.661849	0.273052	0.661849	0.273052

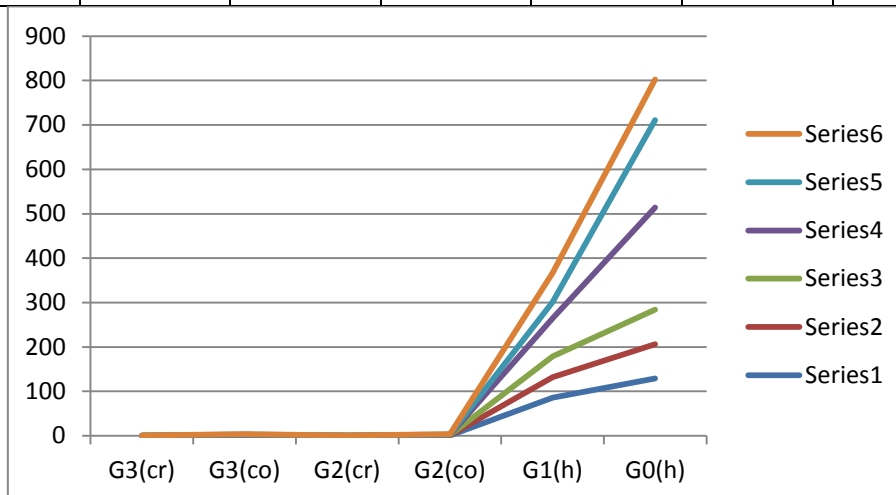


Figure (6) show the 6 features of normal cases

Table (2) Show the six features for abnormal cases

name	G0(h)	G1(h)	G2(co)	G2(cr)	G3(co)	G3(cr)
Pic 1	373.1305	187.1268	0.51897	0.011844	0.599499	0.014362
Pic2	438.3347	251.2601	0.504433	0.011678	0.576356	0.013511
Pic 3	323.143	119.8649	0.574932	0.01955	0.624235	0.021694
Pic4	268.3619	90.26989	0.498716	0.010406	0.546164	0.012156
Pic 5	212.4919	101.8245	0.536959	0.014503	0.600638	0.016301
Pic 6	278.6053	110.8475	0.484992	0.014996	0.541702	0.016085

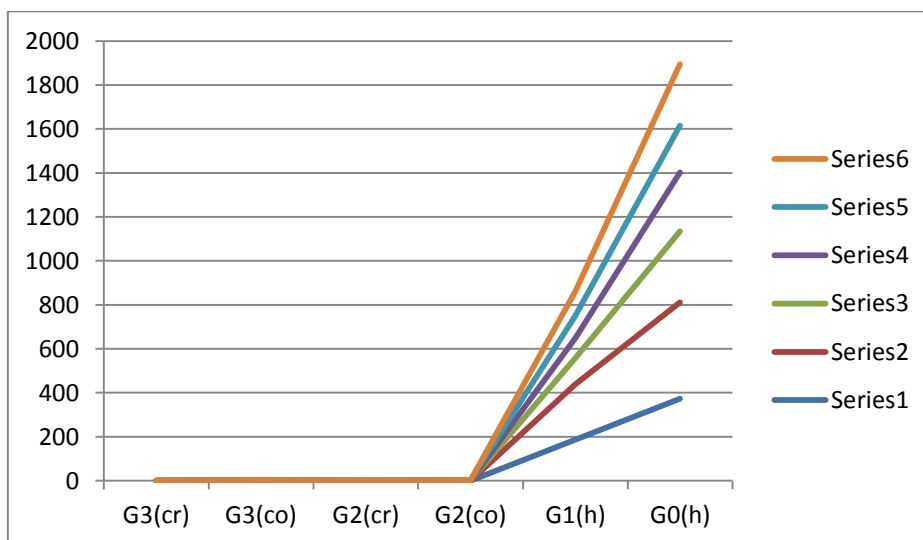


Figure (7) shows the 6 features for the abnormal cases

Table (3) Show the 2 features for benign

name	G0(h)	G2(h)
Pic 1	88.25202	85.24909
Pic 2	78.20492	81.658
Pic 3	77.89426	76.36395
Pic4	119.9024	126.0436
Pic5	57.88611	59.28165
Pic 6	126.6398	50.77

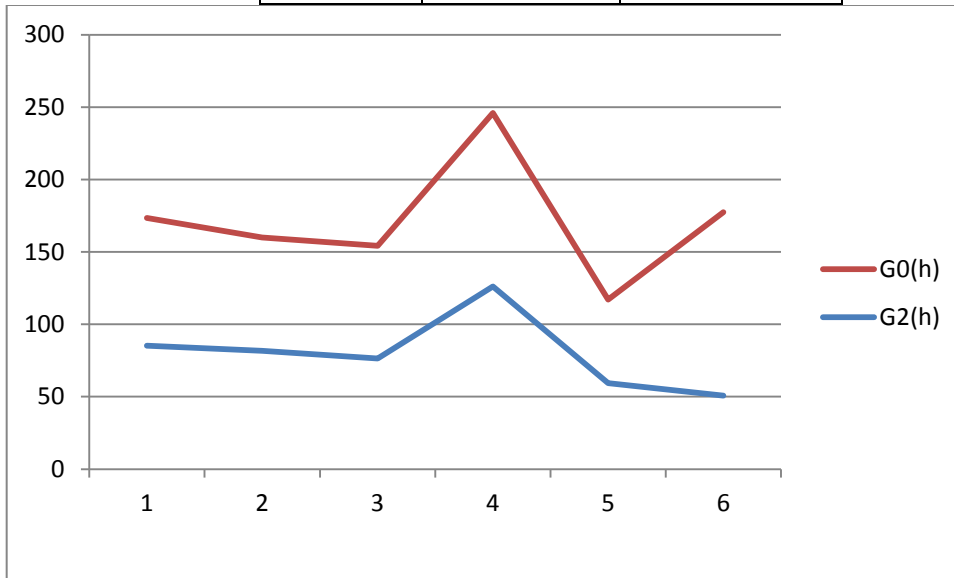


Figure (8) shows the 2 features of the benign cases

Table (4) Show the 2 features for cancer

name	G0(h)	G2(h)
Pic 1	159.0779	162.3309
Pic 1	183.9032	188.8339
Pic 1	354.574	352.9578
Pic 1	177.4063	173.1872
Pic 1	178.1519	170.4073
Pic 1	131.2075	130.7798

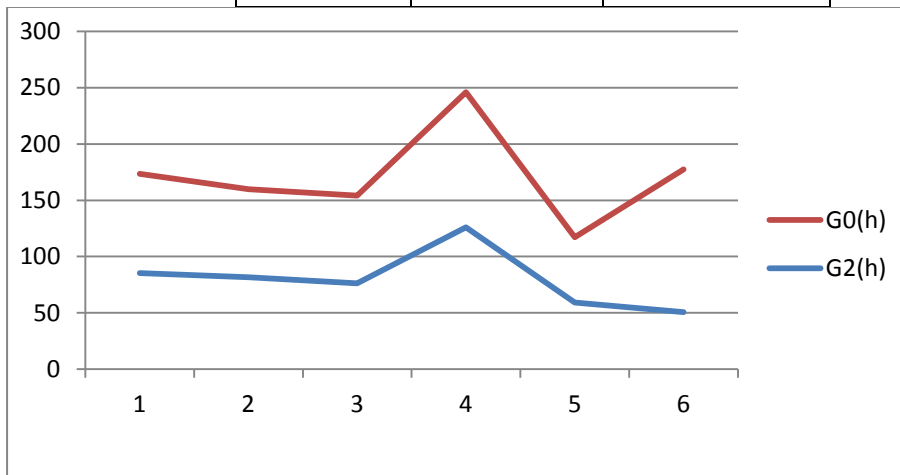


Figure (9) shows the 2 features of the cancer cases

6. Conclusion

We can easy to conclusion the following:

1. GLCM is the most important for textures analysis and easy to get the features which are can be the input for the neural network or Fuzzy logic.
2. Not all the feature from GLCM considered for classification features only 6 feature to classify the normal and abnormal and 2 features for abnormal to classify if it is cancer or no benign.
3. The six feature for classification normal and abnormal are (in 0 degree the homogenate only can consider for classification , in 45 degree only the homogenate, in 90 and 135 degree only contrast and correlation).
4. The two features for classification cancer and benign are (in 0 degree the homogenate only and in 90 degree the homogenate only)
5. Our algorithm can easy used and has benefit for medical research.

References

- [1].S. Aruna, L. V. Nandakishore, S. P. Rajagopalan “Cloud based Decision Support System for Diagnosis of Breast Cancer using Digital Mammograms” published by International Journal of Computer Applications, Vol. EGOV, Issue. 1, pp. 1-3, 2012.
- [2].J. Dheeba & S. Tamil Selvi “A Swarm Optimized Neural Network System for Classification of Micro calcification in Mammograms” Journal of Medical Systems , Vol.36, No. 5, pp. 3051-3061, 2012.
- [3].R. Nithya and B. Santhi “Breast Cancer Diagnosis in Digital Mammogram using Statistical Features and Neural Network” Research Journal of Applied Sciences, Engineering and Technology. Vol. 4, No.24, pp. 5480-5483, 2012.
- [4].Jian Yang Jing feng Guo “Image Texture Feature Extraction Method Based on Regional Average Binary Gray Level Difference Co-occurrence Matrix” International Conference on Virtual Reality and Visualization, pp.239-242, 2011.
- [5].Dhanashree Gadkari “Image Quality Analysis Using GLCM” thesis University of Central Florida Orlando, Florida,2004.
- [6].Sunita P. Aware “Image Retrieval Using Co-Occurrence Matrix & Texton Co-Occurrence Matrix For High Performance” International Journal of Advances in Engineering & Technology, Vol. 5, Issue 2, pp. 280-291. 2013.
- [7].Tuan Anh Pham “Optimization of Texture Feature Extraction Algorithm” M.Sc thesis, Delft university, Netherlands,2010.
- [8].Shaker Ali, Zou Beiji, and Abbas Ali “An Algorithm for Diagnosis of the Three Kinds of Constitutional Jaundice” The International Arab Journal of Information Technology, Vol. 7, No. 4,pp.441-448, 2010.