ttp://jgu.garmian.edu.krd https://doi.org/10.24271/garmian.302

Mammography Detection System of Malignant breast mass Cancer Using Hybrid Expert System and Case Based Reasoning

Soran Saeed [1*] Bestan Maaroof [2] and Alla Shally[3]

[1]Sulaimani Polytechnic University, Kurdistan Region, Iraq
Email: soran.saeed@spu.edu.iq
[2]Sulaimani University, College of Informatics, IT Department, Kurdistan Region, Iraq
Email: bestan@mail.com
[3]Sulaimani Polytechnic University, Kurdistan Region, Iraq
Email: alla.shally@spu.edu.iq

[*]Corresponding author

Abstract

This research work, presents a computer-aided mammography detection of mass image for Malignant breast cancer a system has been developed to help radiologists in order to increase diagnostic accuracy and called (ImageCBR). The aim of this work to find or detect similar Malignant image mass of breast cancer from base knowledge by given a target one. similarity Generally, a ImageCBR system consists of four stages: (a) preprocessing of the image (b) segmentation of regions of interest, such as a well-known mass breast features extraction and selection (shape, size, density, margin), and finally (c) image similarity (target and source). The performance evaluation metrics of ImageCBR systems are also reviewed.

Keywords: Case-Based Reasoning, Expert System, Image processing, Image Similarity, Breast Cancer, mammography

1. Introduction

Nowadays medical knowledge is rapidly to the extent that even experts have difficulties in following all the new results, changes and new treatments. The reasoning was rapidly extending breast cancer detection techniques have been a widely researched area.

The aim of this paper is to present analysis, design and implementation of the proposed system within the methods, techniques and analysis to tackle the research problems which, Mammography breast cancer detection (MBCD) by comparing a mass appeared in a mammogram with previous cases retrieved from base knowledge (BK) for diagnosing breast cancer masses using Artificial intelligence techniques Case -Based Reasoning (CBR) [2] and Expert System [3]for the evaluating features, mass detection and similarity purpose. An image processing (4) is used in the pre-processing of the images and preparing the images for the analysis of the features that they reveal. Visual studio2015 and Microsoft access2010 is used for the implementation.

Over 100 mammography images as real case have been collected from two different hospitals Faruq Medical City (FMC), the Breast Disease Treatment Centre (BDTC) at Sulaimani hospital and an additional 38 cases obtained from the Mammographic Image Analysis Society (MIAS) Database [5] Experimental results demonstrated the performance and efficiency of the proposed algorithm and the performance details are given in the evaluation section.

Various methods for shape description have been suggested through the years of research in engineering design and human perception, but none provide a complete and natural solution to the problem. Furthermore, this problem seems to be one of the most challenging problems, and is perhaps equivalent and the negative image of breast masses.

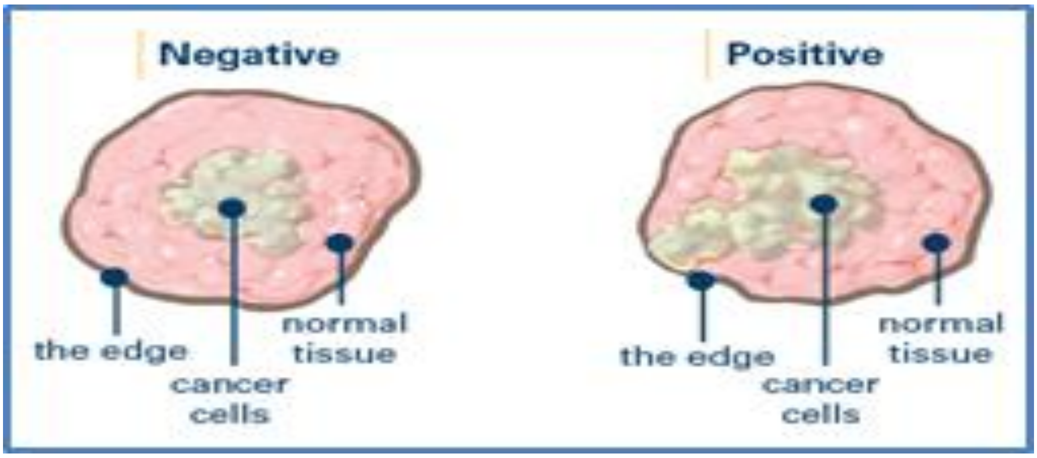


Figure 1.0 The structure of the breast mass.

2. Background

The research is mainly concerned with the similarity process for the **target BCimage** and **source BCmass** retrieval problem. Also, it is concerned with the phases of automating the process of only separate the layers of the BCimage and already well-known BCmass components such as (shape, size, location and density) for an image for this reason needed the graphical representation to allow for the efficient retrieval of similar target BCimage and source BCimage thus reuse of relevant case knowledge.

The background of the problem is a breast cancer disease. It is useful to define what “breast cancer” means. The term “breast cancer” refers to a malignant tumor that has developed from cells in the breast. Usually breast cancer either begins in the cells of the lobules **[6]**, which are the milk-producing glands, or the ducts, the passages that drain milk from the lobules to the nipple.

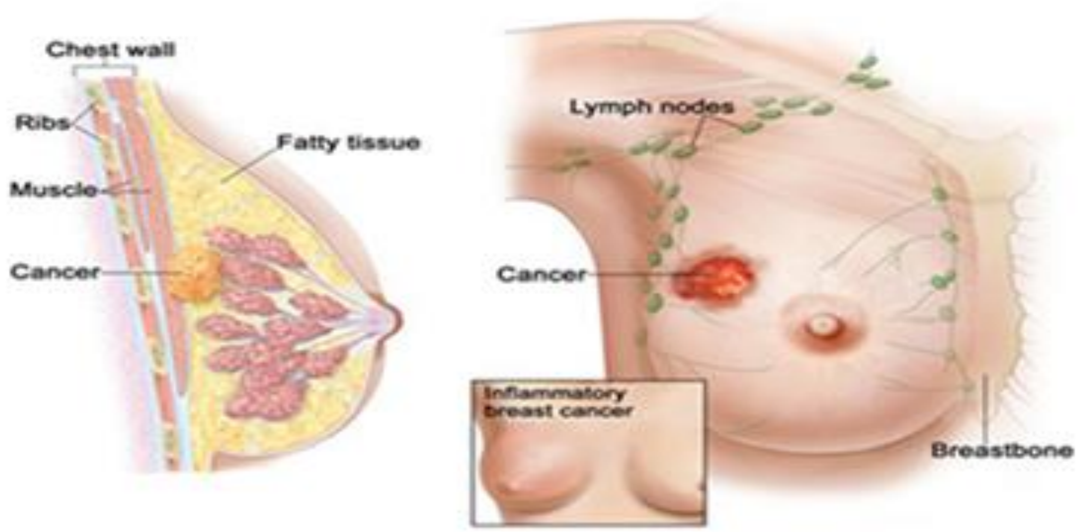


Figure 2.0 The structure of the breast and breast cancer.

3. Architecture design

Figure2.0 shows the architecture design of the proposed system. The integrated expert system and image processing attempt to increase the accuracy of the case-based reasoning detection of mammography breast cancer obtained from classical Case-based reasoning.

When a new case is arriving the image is pre-processed and analysed by the image processing and the image is re-analysed by our case doctor features are extracted and stored in the base knowledge. The CBR cycle **[7]** starts when the new case (target case) is arrived; in the **RETRIEVE** step all cases in the system are retrieved that are similar to the new case. Then **REUSE** the solution of the most similar case is Reused for the new case, in the **REVISED** step the suggested solution is revised or tested whether it is suitable for the new case or not. Finally the test case is added to the database in the **RETAIN** step as a new learned case in this way the CBR will get more and more clever as a human expert as he sees more cases he gets more clever.

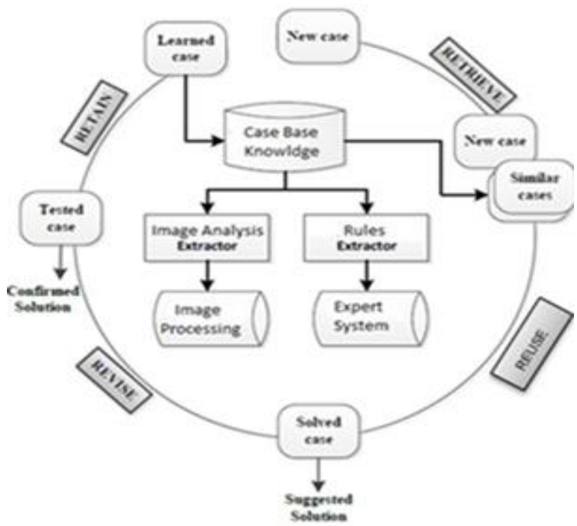


Figure 3.0 Architecture design of the proposed system **(ImageCBR).**

4. Proposed system

The proposed ImageCBR System for breast cancer detection is composed of two main phases: adding archival cases, testing the new case. The proposed approach compares the mammogram image of a new case that contain a mass to old cases in the database (knowledge base) each case is represented as features and weight to represent its importance. The similarity process compares the new case's mass image the mass features with each image and features in the knowledge base. Figure 6.0 shows the CBR System. The system's components are graphical user interface that serves the user, functional level that includes the features, archival case and the testing new case. The functional level interacts with the database that is called knowledge base.

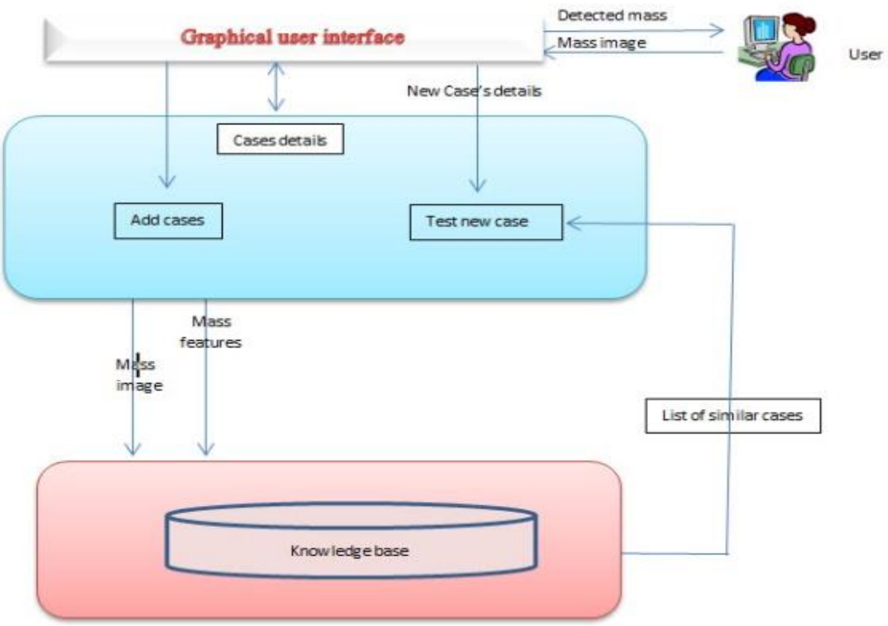


Figure 4.0 Shows the proposed System ImageCBR.

4.1 Phases of the proposed system

The first phase is adding archival cases to form our knowledge base. Each case represents a patient. Each patient had taken a mammogram that contains a mass with the mammogram report. This phase is composed of two main steps. The first step is preprocessing the mammogram image and extracting region of interest (RIO) [8] then putting into our database. Image processing is used for this step. The second step is feature extraction and adding the new and old cases to the database as shows in the figure 7.0 how expert system it works.

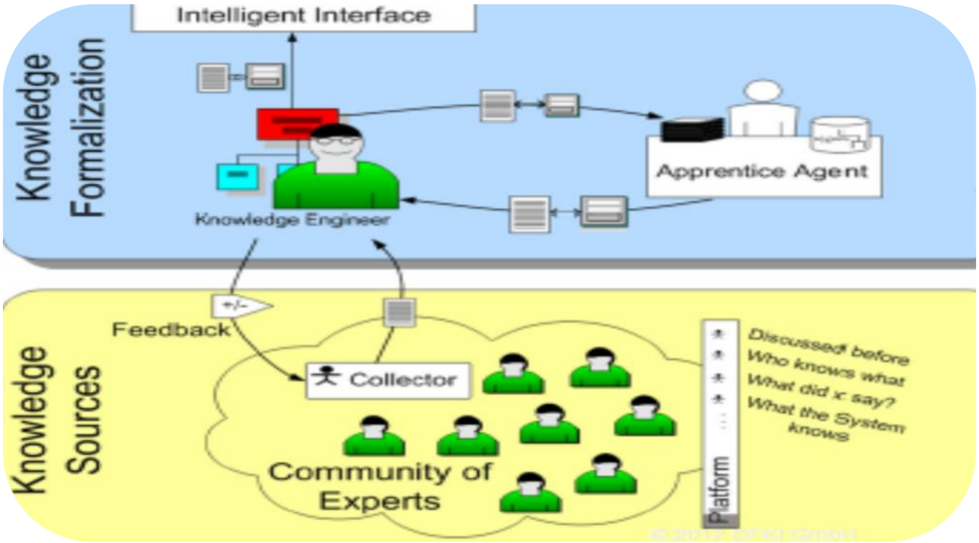


Figure 5.0 Adding archival cases to the system.

The second phase is the classification of the new case. This phase focuses on matching the inputs of a new case with the existing image and features of each case in the knowledge base.

The result of the comparison/similarity is the highest similarity value that supports decision making for the detection of the case. The case with the similarity value is the most suitable case for the new case.

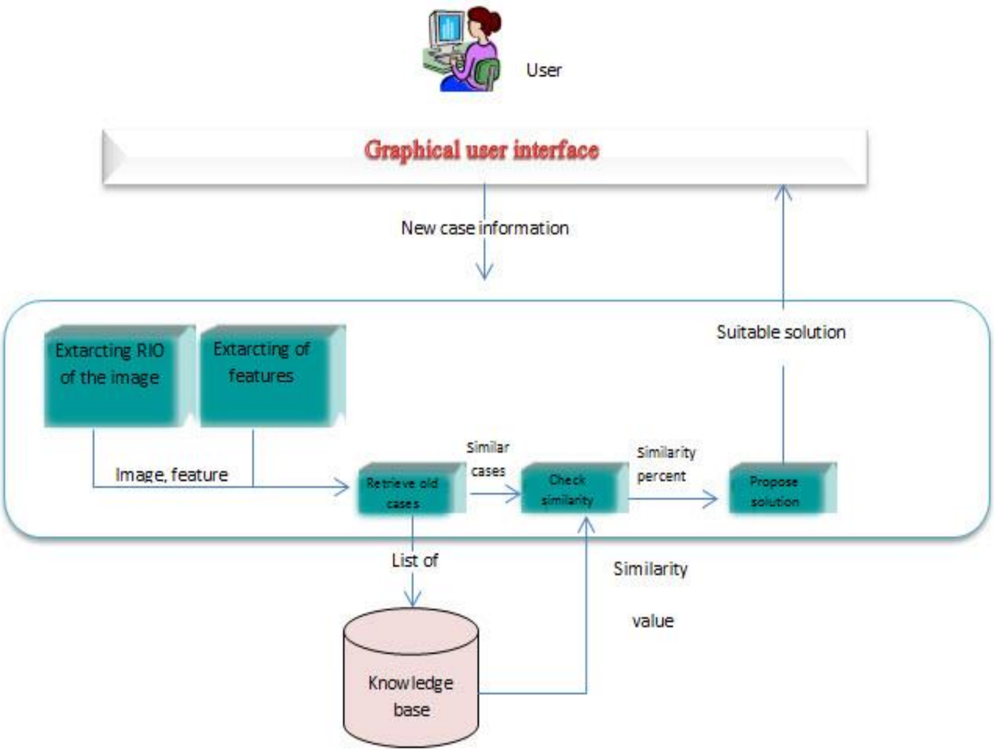


Figure 6.0 Shows the testing process for the system.

5. Data Collection and Investigations

Datasets

Three different data sets are used which they are:

Breast Disease Treatment Centre (BDTC)

25 Cases were obtained from BDTC all of the cases are biopsy based proved cancer cases the radiology report of the cases was diagnosed by two different radiologists of the centre and as the second decisions all cases are re diagnosed by our case expert (Dr. Alla A. Shally). The images are all of a very good quality all are 2394x3062 pixels.

Faruq Medical City (FMC).

Fifty (50) cases have been collected and they are composed of 4 benign and the rest of malignant cases, they are diagnosed by our case expert. Images are of better quality than Breast centre, all are 2800 x 3518 pixels.

Data collection from Mammography Image Analysis Society (Mini MIAS Database). (The mini-MIAS database of mammograms, 2012) [5].

Mini MIAS is a database that provided facilities for the researchers that are interested in breast cancer research area, and widely used it contains 322 images of the breast. They are of three classes (normal, benign, malignant) on each (benign and malignant) image given a description about (shape, margin, density and location) of the abnormality.

36 images are taken for this work which contains mass, 22of them contain benign mass and the other 14 contains malignant mass.

Images are of lower quality than the two other data sets, about 1024 x 1024 pixels in Portable Grey map (PGM) format.

Images of the three data sets are of Digital Imaging and Communications in Medicine (**DICOM**) focus of this they are converted into a readable format by **DICOM** viewer program version 0.9.1 (Build 618) 64 bit, copyright 2006-2016, which can be read by special programs.

6. Quest for information (Expert System)

A medical CBR system in order to detect cancer, the system has to be first fed with information from experts, in our case doctors. A number of interviews are made by the system expert (the case doctor), the cases are diagnosed by the expert then the expert detected the mass and extracted the features and the features are copied into a form that was designed in order to record the description of the mass images that our case doctor diagnose it. After ending up with data collection it's time to convert the data into information and then encoding to our system. As mentioned in the previous section. Then each image diagnosed by two different specialist radiologists experts, figure 9.0 and figure 10 shows the process of how the quest of information works in expert system.

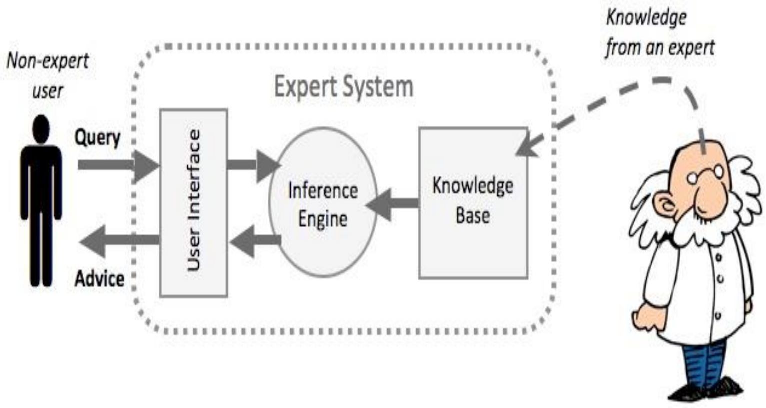


Figure 7.0 the architecture of Expert system.

7. Case representation

Each case is about the mass image, represented in the knowledge base as the following features of the Breast Imaging-Reporting and Data System (BI-RADS):

Shape: (round, oval, irregular)

Margin: (well defined, ill-defined, speculated)

Density: (fat equivalent, hypodense)

Size: the approximate size of the mass is calculated as follows:

The approximate size of the mass is calculated as follows:

Let (x1, y1) be a point on the edge of the mass, and (x2, y2) be a point on the center of the mass

Area= $\left(\frac{\sqrt{(x1-x2)^2+(y1-y2)^2)}\times 2.54}{r}\right)^2 \times \pi$ [9].

Where (r) is the resolution of the image.

The experts experience is the most important feature in mass detection is the shape of the mass, then the margin after that the density and then other features.

8. Image processing

In order to do some processing on an image, it should be made ready. To do so pre-processing is a necessary task.

8.1 Image pre-processing

Mammogram images are converted to a process format by DICOM viewer program. After that, the images undergo some image pre-processing algorithm to identify the Region Of Interest [10] (RIO). The very first step is cropping the part of the image which contains the mass.

Then a small program is written with Visual Studio C# 2015 to resize the images to 128 x128 pixels. Figure 11.0 shows the interface of the resize program and figure 11.0 shows the process of the resizing.

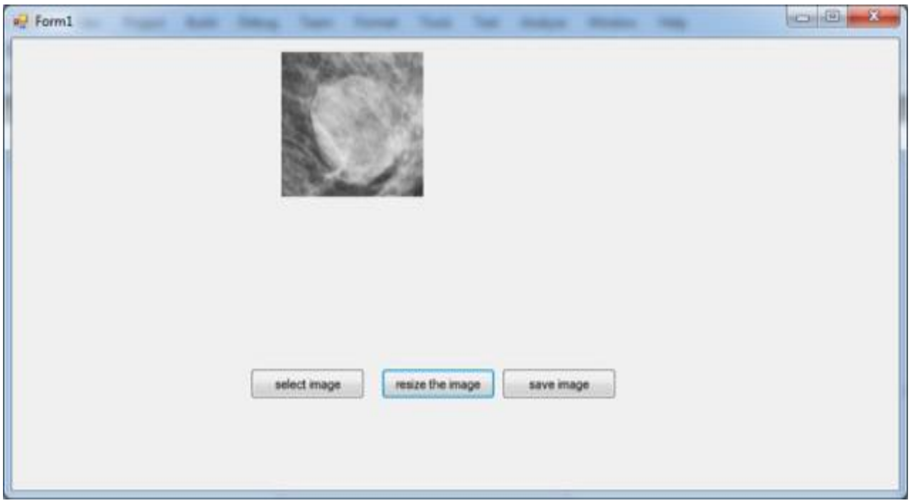


Figure 8.0 Shows the Interface of the resize program of Mammo image.

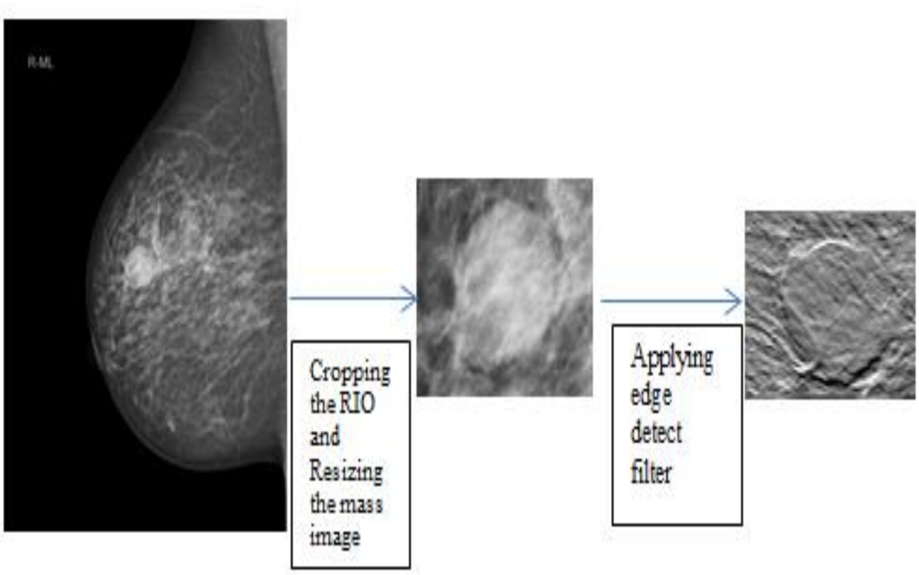


Figure 9.0 A block diagram shows the steps image processing.

8.2 Edge detection filter

After resizing all of the images finished, then quick edge detect filter is designed with convolution filter matrix [13] [14]. Convolution is the treatment of a matrix by another one which is called "kernel", the image is multiplied by a filter matrix to enhance the edges of the mass in the image then used for the similarity. The filter studies successively every pixel of the image. For each of them, which we will call the "initial pixel", it multiplies the value of this pixel and values of the 8 surrounding pixels by the kernel corresponding value. Then it adds the results, and the initial pixel is set to this final result value.

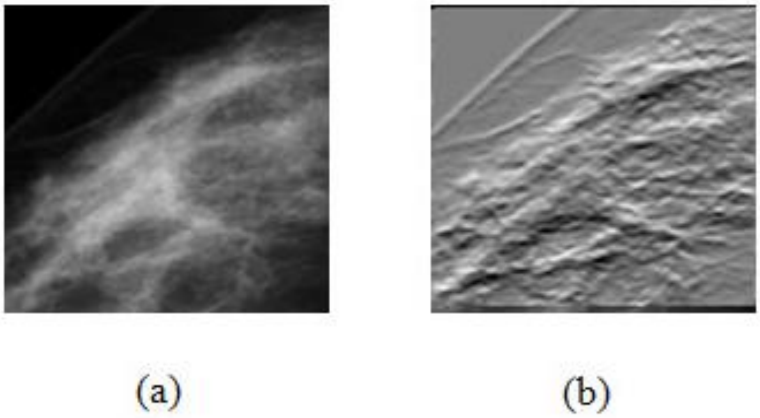


Figure 10.0 Edge detection filter (a) is the original image (b) is the image after adding the filter.

Similarity Process

The similarity method is the most essential method that the result totally depends on the proposed method occurred between the test image with the images in the system. Similarity between images are found by the PCC (equation (2)).

$$r_1 = |\frac{\sum_i (x_i - x_m)(y_i - y_m)}{\sqrt{\sum_i (x_i - x_m)^2}\sqrt{\sum_i (y_i - y_m)^2}} \qquad (2)$$

PCC equation (2). [10] [11] [12].

when xi is the intensity of the ith *pixel* in test image, yi is the intensity of the i t h pixel in images in the system, xm is the mean intensity of test image, and ym is the mean intensity of the image in the system. The function treats pixel by pixel of the comparing images to find the relation between images. If the result between two images are positive then there is a relation between the two images we can say they are similar we can accept it as one of the similar cases and then, the case if the coefficient is smaller or equal to (0) then the images are not similar, the system reject the approval for dissimilarity by using PCC equation The cases of the images are retrieved from the BK ordered from largest coefficient to the lowest the case with the maximum coefficient is selected and its class (benign, malignant) is suggested for the new case and advice is suggested for the new case.

9. Implementation of the proposed System

For the proposed system to tackle the research problem three algorithms have been designed, the first one to filter the image and resized to have better quality to be ready and hand out to experts for analysis, for this technic image processing be used as has been explained in section similarity using PCC equation.

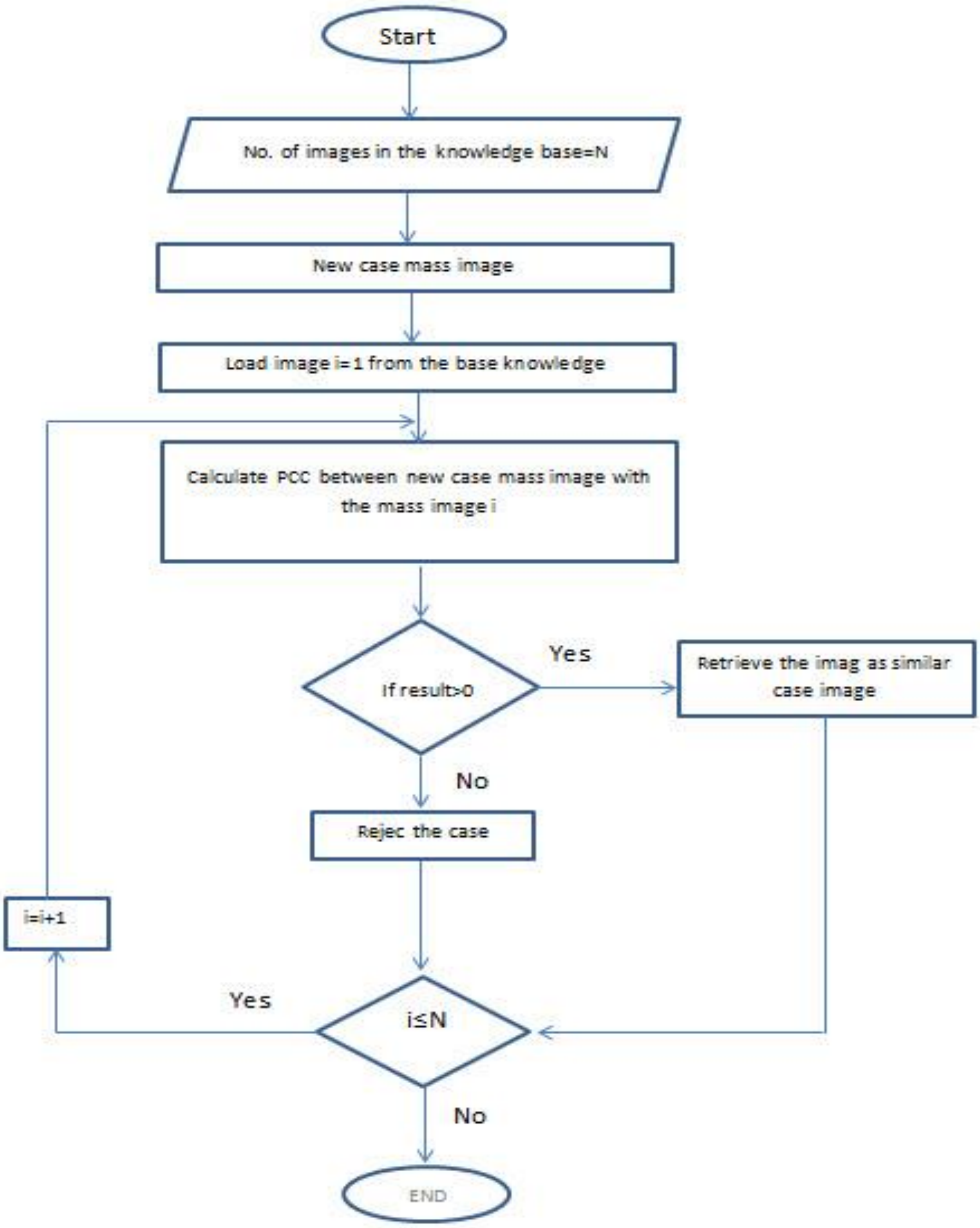


Figure 11.0 Flowchart of the similarity.

9.1 Experimentation and Test.

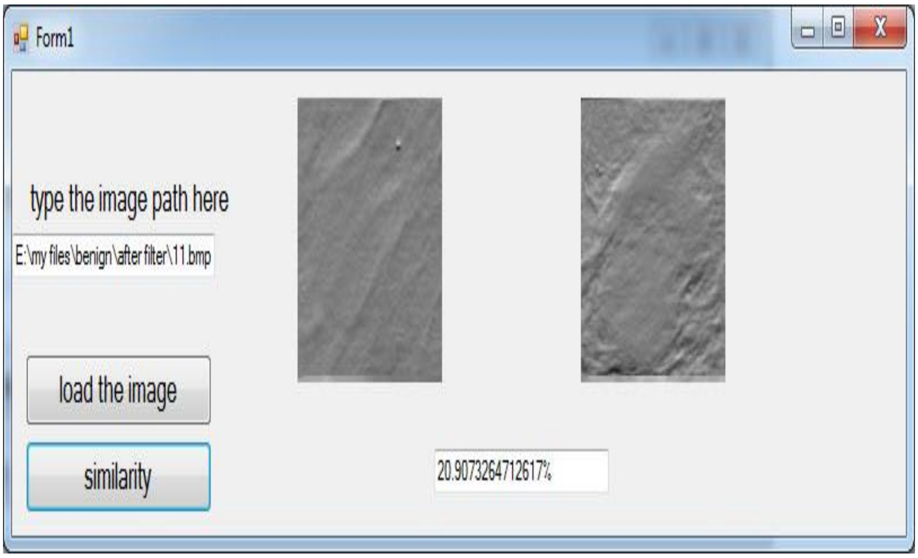


Figure 12.0 Similarity program interface.

Figure 1 Case addition form interface.

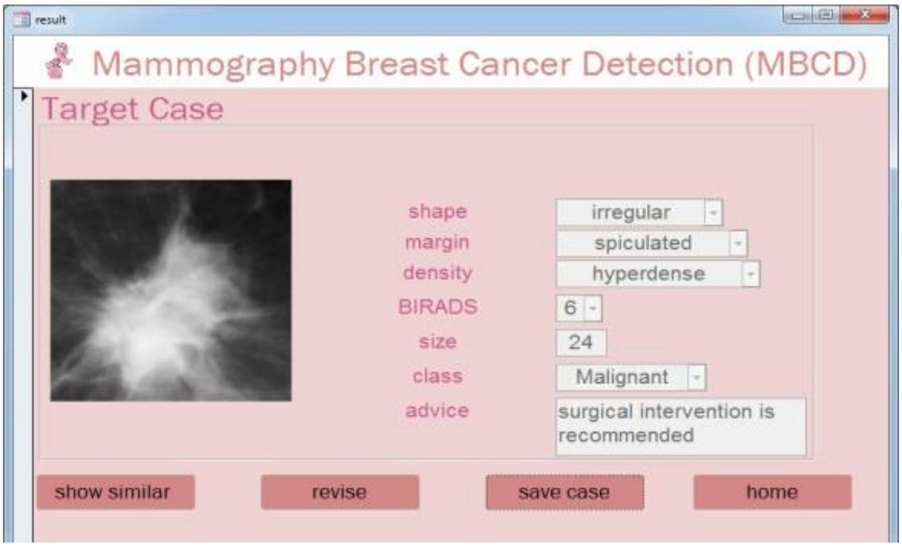


Figure 14.0 Diagnosing interface.

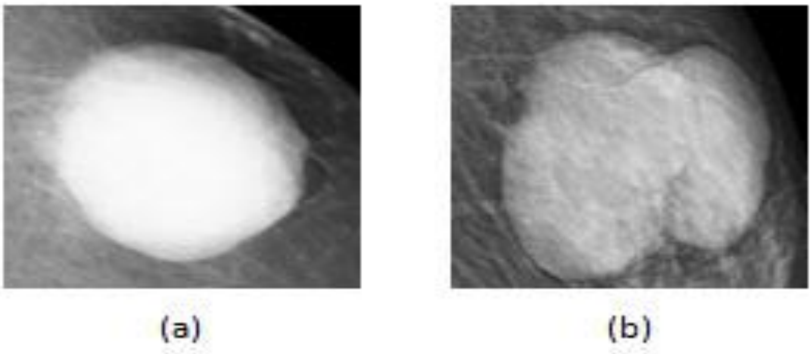


Figure 15.0

FMC that is a malignant mass, the proposed system retrieved the similar cases from the Base Knowledge and selected case (29) as the most similar case to the target case.

Target case diagnosed by:	Shape	Margin	Density	Size	BIRADS	Class
Radiologist	Round	Lobulated	Hyperdense	9	3	Benign
Proposed System	Round	Spiculated	Hyperdense	12.5	4	Malignant

Table 1.0 Radiologist VS proposed system.

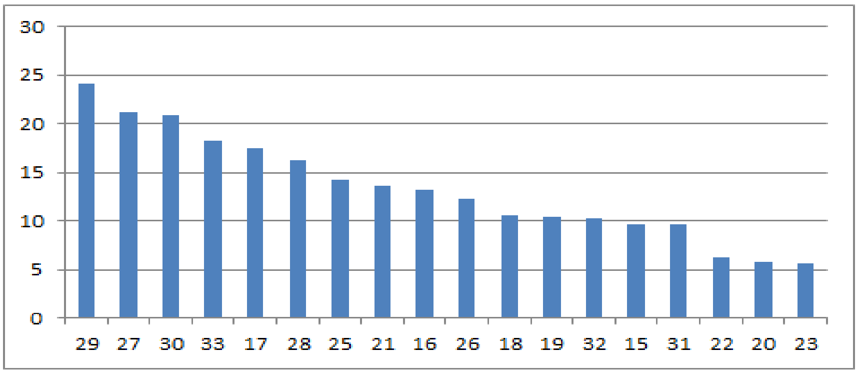
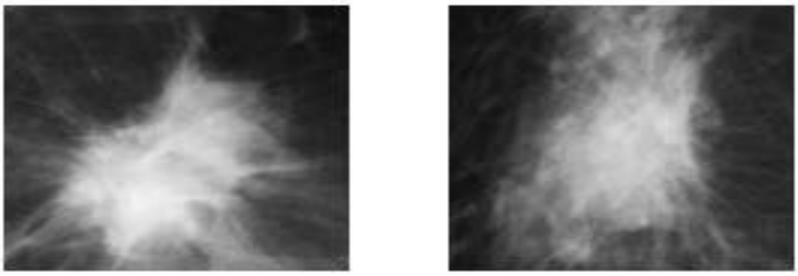


Figure 2 Similar cases to the target case (a)



(a) (b)

Figure 17.0

Target case diagnosed by	Shape	Margin	Density	Size	BIRADS	Class
Radiologist	Irregular	Spiculated	Hyperdense		5	Malignant

Table 2.0 Radiologist proposed system.

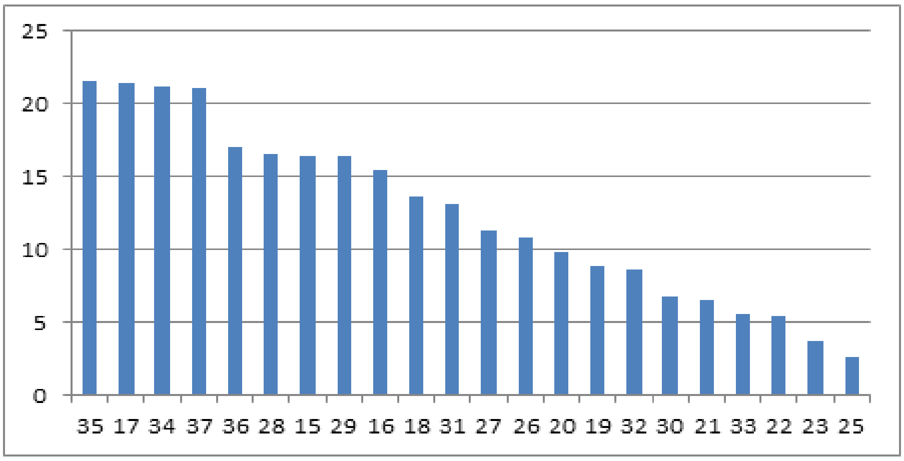
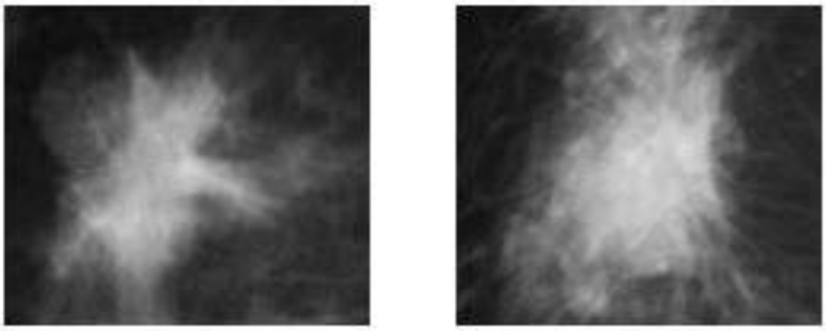


Figure 17.0 Similar cases to the target case (a).



(a) (b)

Figure 18.0

Target case (a) is a case from BDTC that is proved by biopsy to be malignant. The mass is diagnosed by Dr. Kalthum, and has written the radiology report. Table (3) shows the mass description made by the radiologist doctor.

The image (b) is a mass image of a caseID (35) from our Base Knowledge, it is a case of BDTC that is proved by biopsy that it is cancer, the mass is diagnosed by Dr. Shad Muhammad. And re diagnosed by Dr Alla A. Shally the radiology report has been written by Dr Shad Muhammad. The proposed system retrieved the similar cases from the Base Knowledge and selected case (35) as the most similar case to the target case.

Target case diagnosed by	Shape	Margin	Density	Size	BIRADS	Class
Radiologist	Irregular	Ill-defined	Hyperdense	6	5	Malignant
Proposed System	Irregular	Spiculated	Hyperdense	24	6	Malignant

Table 2 Radiologist VS proposed system.

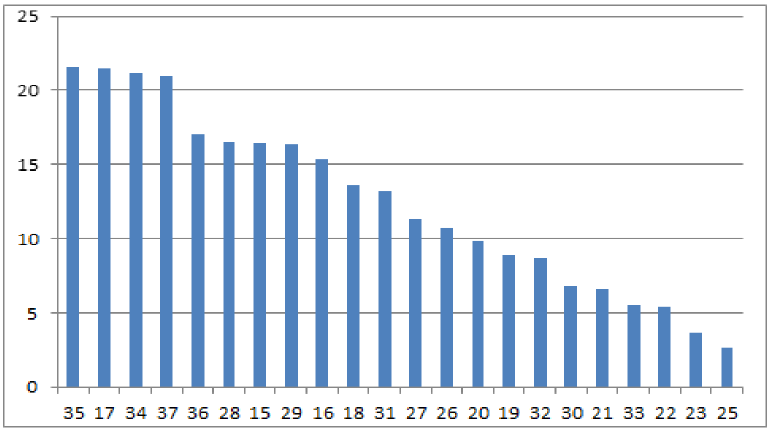


Figure19.0 Similar cases to the target case (a)

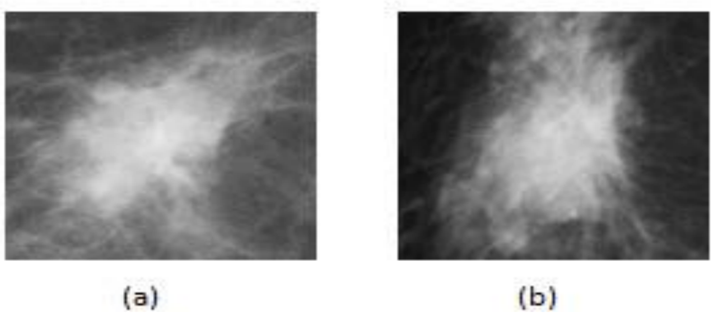


Figure 20.0

Target case (a) is a case from BDTC that is proved by biopsy to be malignant. The mass is diagnosed by Dr. Kalthum, and has written the radiology report table (4) shows the mass description made by the radiologist doctor.
The image (b) is a mass image of a caseID (35) from our Base Knowledge.it is a case of BDTC that is proved by biopsy that it is cancer, the mass is diagnosed by Dr. Shad Muhammad. And rediagnosed by Dr. Alla A. Shally the radiology report has been written by Dr. Shad Muhammad.
The proposed system retrieved the similar cases from the Base Knowledge and selected case (35) as the most similar case to the target case.

Target case diagnosed by	Shape	Margin	Density	Size	BIRADS	Class
Radiologist	Irregular	Ill-defined	Hyperdense	6	5	Malignant
Proposed System	Irregular	Spiculated	Hyperdense	24	6	Malignant

Table 3.0 Radiologist VS proposed system.

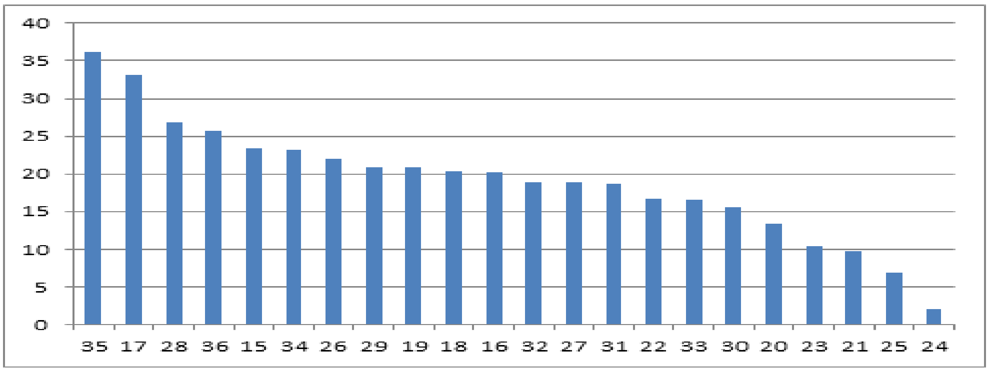


Figure 21.0 Similar cases to the target case (a)

Figure 22.0

Target case (a) is a case from BDTC that is proved by biopsy to be malignant. The mass is diagnosed by Dr. Kalthum, and has written the radiology report. Table (4) shows the mass description made by the radiologist doctor.

The image (b) is a mass image of a caseID (35) from our Base Knowledge, it is a case of BDTC that is proved by biopsy that it is cancer, the mass is diagnosed by Dr. Shad Muhammad and radiologist by Dr. Alla A Shally. The radiology report has been written by Dr. Shad Muhammad.

The proposed system retrieved the similar cases from the Base Knowledge and selected case (35) as the most similar case to the target case.

Target case diagnosed by	Shape	Margin	Density	Size	BIRADS	Class
Radiologist	Irregular	Ill-defined	Hyperdense	6	5	Malignant
Proposed System	Irregular	Ill-defined	Hyperdense	24	6	Malignant

Table 4.0 Radiologist VS proposed system

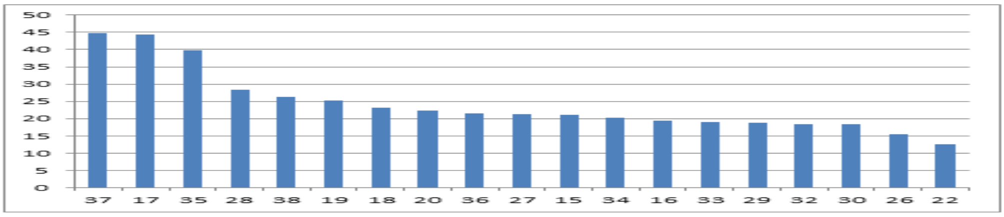


Figure 23.03 Similar cases to the target case (a)

10. Conclusions

This paper proposes a novel method to analyse (MI) objects into separate parts (1234, based on: Three algorithms have been designed to tackle the research problems.

References

[1] what_is_bc: breast cancer. (2015, October 23). Retrieved July 21, 2016, from breast cancer Web site : http://www.breastcancer.org

[2] G. Kamp, S. Lange, C. Globig,. (1998). Case-based Reasoning Technology:related area. Case-based Reasoning Technology:from Foundations to Application,. Berlin: Springer.

[3] Jimmy Singla, Dinesh Grover, Abhinav Bhandari. (May 2014) “Medical Expert Systems for Diagnosis of Various Diseases”International Journal of Computer Applications (0975 – 8887)Volume 93 – No.7.

[2] Mohamed Hachama, Agn`es Desolneux and Fr´ed´eric J.P. Richard” A Bayesian Technique for Image Classifying Registration” (2012). IEEE TRANSACTIONS ON IMAGE PROCESSING, TIP-08569-2012.R1, MAI.

[5] (Online) (2014). “The mini-MIAS database of mammograms. (2012, December 11). Retrieved September 1, 2014, from Mammographic Image Analysis Society” :http://www.mammoimage.org

[6] Kovalerchuck, B, Triantaphyllou, E., Ruiz, J. F., & Clayton, J. (1997). Fuzzy logic in computer-aided breast-cancer diagnosis: Analysis of lobulation. Artificial Intelligence in Medicine, 75-85.

[7] Watson I. (1999). “Case-based reasoning is a methodology not a technology” Knowledge-Based Systems, 303–308.

[8] S.Singh, A. M. (2000). Indentification of regions of intersest in Digital mammograms. Journal of Intelligent Systems.

[9] Jalalian A1, M. S. (2013). Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound Clin Imaging. Clin Imaging., 430-6.

[10]Y. K. Eugene and R.G. Johnston. (1996). The Ineffectiveness of the Correlation Coefficient for Image Comparisons”. Los Alamos: Technical Report LA.

[11] Rodgers J. L. and Nicewander W. A. (1988). “Thirteen Ways to Look at the Correlation Coefficient”. The American Statistician, 42.

[12] Eugene K . Jen, Roger G.Johnston. (n.d.). Nov.2016) “The Ineffectiveness of Correlation Coefficient for Image Comparisons”. New Mexico: Research Paper prepared by Vulnerability Assessment. Edition. 59-66;

[13] Singh R. and Shaw D. (2016). Experimental Analysis of Impact of Noise on Various Edge Detection Techniques. World Congress on Engineering. London: WCE.

[14]Salem Saleh Al-amri, Dr. N.V. Kalyankar and Dr. Khamitkar S.D . (2010). IMAGE SEGMENTATION BY USING EDGE DETECTION. International Journal on Computer Science and Engineering, 804-807.

نظام الكشف كتلة الثدي الخبيثة السرطان باستخدام نظام الخبراء الهجين والحالة المستندة إلى المنطق

الخلاصة

هذا العمل البحثي، ويعرض الكشف عن الثدي بمساعدة الحاسوب من الصورة الجماعية لسرطان الثدي الخبيث وقد تم تطوير نظام لمساعدة علماء الأشعة من أجل زيادة دقة التشخيص ويدعى Image CBR

والهدف من هذا العمل للعثور على أو اكتشاف مماثلة صورة خبيثة كتلة من سرطان الثدي من قاعدة المعرفة من خلال إعطاء هدف واحد. التشابه بشكل عام، يتكون نظام (Image CBR) من أربع مراحل: (أ) المعالجة المسبقة للصورة (ب) تجزئة المناطق ذات الأهمية، مثل خصائص الثدي المعروفة جيدا الاستخراج والاختيار (الشكل والحجم والكثافة والهامش)، و وأخيرا (ج) تشابه الصورة (الهدف والمصدر). كما تتم مراجعة مقاييس تقييم الأداء لنظم (Image CBR).

الكلمات ألمفتاحية :االمنطق القائم على حالة، نظام الخبراء، معالجة الصور، تشابه الصورة، سرطان الثدي، التصوير الشعاعي للثدي**.**