Mammograms Segmentation and extraction for breast cancer regions based on region growing

تقسيم صور الثدي واستخراج مناطق اصابتها بالسرطان باستخدام طريقة نمو المنطقة

Semaa Ibrahim M.Ali

Dr Nassir H. Salman

كلية العلوم / جامعة بغداد

Abstract:

Medical imaging is an essential part of modern healthcare, where it's technologists take X-rays, mammograms, ultrasounds and computed tomography images to help diagnose patients' injuries and diseases. In this work, the application of region growing technique was explored to the problem of image segmentation, extracting and finding the boundary of different breast tissue regions cancer in mammograms images. Our search focus on two parts, the first is detecting the cancer area. The detection algorithm used on 117 dataset images, the result is detection cancer in 115 image and two images are shifted. Then the second step started with origin these images using segmentation method based on region growing. The goal of the segmentation algorithm here is to see if the region-growing algorithm could separate different intensities for the different breast patterns. This algorithm is applied with selecting a seed point to provide the hard constraint, whereas the seed point are selected based on user-defined. Region growing has been explored on images of various imaging modalities but not on mammograms just yet. Therefore, this article is mainly focused on using region-growing algorithm to perform segmentation to increase the visibility of different breast densities in mammography images. Our proposed methodology for the segmentation of mammograms has been tested on Mini-MIAS database mammogram images. The results show that the proposed algorithms are fast for image segmentation into regions with edges detection, region extraction, and region features calculations such as region area, mean, max, minimum intensity values, and no. of pixels etc.

Keywords*:* mammograms, breast cancer, image segmentation, image processing. Region growing, Mini-MIAS database

الخلاصة:

التصوير الطبي هو جزء أساسي من الرعاية الصحية الحديثة. يقوم أخصائيو التصوير الطبي بأخذ صور الأشعة السينية والتصوير الشعاعي للثدي والتصوير بالموجات فوق الصوتية والصور المقطعية المحوسبة للمساعدة في تشخيص إصابات المرضى وأمراضهم. في هذا العمل ، نستكشف تطبيق تقنيات نمو المناطق على مشكلة العثور على حدود مناطق نسيج الثدي المختلفة في صور الثدي الشعاعية المصابة بالسرطان. لقد تم استخدام خوارزمية تحديد الصور التي بها سرطان وهي 115 صورة من اصل 117 وتم استبعاد صورتين لوجود خلل فيهما اصلا كما ورد في نص البيانات . لتبدا طريقة التقسيم باستخدام خوارزمية نمو المنطقة المختارة .ان الهدف من خوارزمية التقسيم هو معرفة ما إذا كانت خوارزمية النمو في المنطقة يمكن أن تفصل الكثافات المختلفة لأنماط الثدي المختلفة. يتم تطبيق نمو المنطقة مع اختيار نقطة البدايه لتقسيم المنطقه على اساس حد عتبه معين ، في حين يتم اختيار نقطة البدايه على أساس تعريف المستخدم. وقد تم استكشاف تزايد المنطقة على صور لطرائق التصوير المختلفة ولكن ليس على صور الثدي بالأشعة السينية حتى الآن. لذلك ، يركز هذا المشروع بشكل رئيسي على استخدام

خوارزمية النمو في منطقة الثدي المصابة لإجراء تجزئة لزيادة الرؤية من كثافات ثديية مختلفة في صور التصوير الشعاعي للثدي. يعتبر تقسيم صورة الماموجرام إلى كثافات مختلفة للثدي مفيدًا لتقييم المخاطر والتقييم الكمي لتغيرات الكثافة. تم اختبار منهجيتنا المقترحة لتجزئة تصوير الثدي بالأشعة على أساس منطقتهم إلى فئات مختلفة من الكثافة على قاعدة بيانات Mini-MIAS. وكانت الطريقتين المستخدمة سريعة في تقسيم صور الثدي واستخلاص المنطقة المصابة وحساب خواصها لاغراض التحليل.

1. Introduction

Recently, computer-aided detection (CAD) has become a portion of the routine clinical work for detection of breast cancer on mammograms at many hospitals and screening sites in many countries. This seems to appear that CAD is beginning to be applied largely in the detection and differential diagnosis of many various types of abnormalities in medical images obtained in various examinations by use of different imaging modalities.

In fact, CAD has become one of the major research subjects in medical imaging and diagnostic radiology. Although early attempts at computerized analysis of medical images were made in the 1960s, serious and systematic investigation on CAD began in the 1980s with a fundamental change in the concept for utilization of the computer output, from automated computer diagnostics to computer-aided diagnosis [4] .

Medical images play an important role in helping diagnosis and treatment, and can be used to teach health care students, and explaining these images will help them learn. Developments in digital imaging technology have greatly increased the number of digital images captured in recent years [5]. Interpreting and analyzing medical images is an important and exciting part of computer vision and pattern recognition. The development of computer assisted diagnostic systems for cancer such as breast cancer has become very important for hospital physicians and has become a top priority for many clinical researchers and centers [8].

This paper concerned with the breast cancer as well as fatty tissue, where breast cancer is a group of abnormal cells is found in the tissues of the breast. In very recent years there are numerous modalities have been developed for the prediction of breast cancer. In biopsy testing[1] ,the biopsy is occupied from the tissues of the breast. The test provides higher accurate result but the procedure to take the biopsy from breast is very painful and pathetic. So, the most of the patients are not intrigued with this test.

A mammogram is the most widely used technique for the detection of breast cancer which provides the 2D projection images of the breast [2]. The proposed method starts with the center pixel of the image as initial seed. The region growing formula uses three homogeneity criteria local,. If the pixel is closer to the growing region as compared to its neighbors then it is included in the growing region, otherwise it is labeled as boundary

pixel. a region growing technique was presented for color image segmentation. In this technique a multi seeded region growing technique for image segmentation was proposed, which defined by user to segment the breast region to assist the radiology to detect the cancer region

2. Region growing method

Planting the area is a bottom-up procedure that starts with a set of pixels. The aim is to develop a unified area connected to each seed. Partitioning by developing an area of sowing point using one criterion is to look at the maximum difference and combine the adjacent areas whose maximum variation is within the tolerance of the seed blocks. The area is frequently grown by comparing all non-assigned adjoining pixels to the region. The difference between the pixel density value and the center area can be used as a measure of similarity. Pixel is allocated with smaller difference measuring this method for the region. This process stops when the density difference between the average region and the new pixel becomes greater than a certain limit. Starting with seed, the density values for each pixel are compared to its neighbors and if it is within the minimum, it will be marked as one. The pixel is added to a region if:

- Does not belong to any other area.
- Be adjacent to that area
- Still unified when creating new area by adding pixels

3. MIAS Dataset

The data set used in this paper is international data collection (MIAS), a British research organization interested in understanding mammograms. The original 322 images (161 pairs) are included with a resolution of 50 microns in PGM format and the real data associated with them. Films filmed in the National Breast Screening Program (NBSP) in the UK. The image is numbered to a 50-micron pixel and represented in 8-bit words per pixel. Minimized to 200 µm pixels and filled, all 1024 x 1024 images (including 202 normal and 120 abnormal images) [3]. See some images of that dataset (chose 117 images) in figure 1:

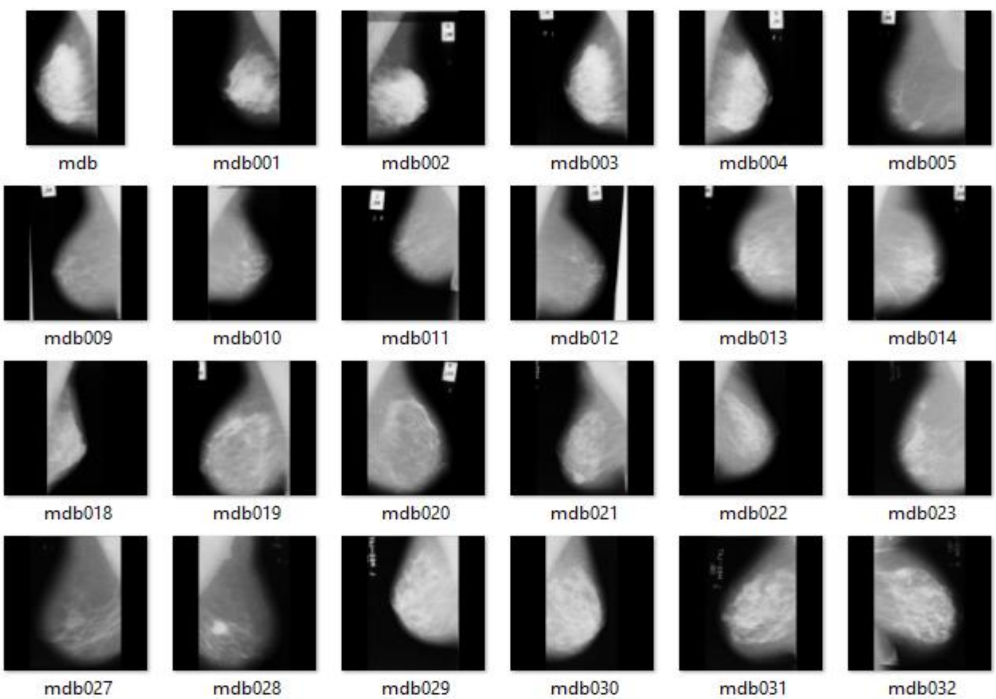


Figure 1. Samples of mamagram images

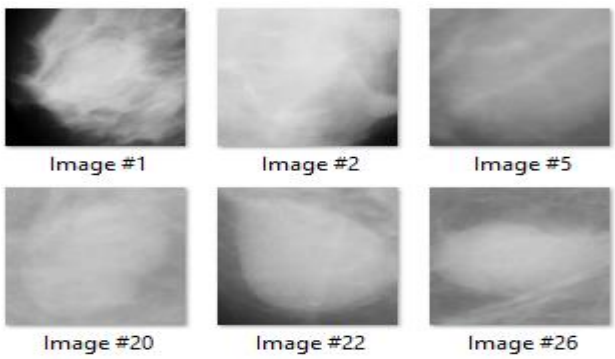
4-Region of Interest (ROI) Extraction process

Original mammogram images have different types of noises, artifacts in their background, pectoral muscles etc., which are unwanted for feature extraction and classification. Hence, a cropping operation has been applied on mammogram image to extract the ROIs, which contains the abnormalities, apart from the unwanted portions of the image. MIAS database gives all the details about each mammogram image, viz., size in pixels, character of background tissue, class of abnormality, Xc and Yc coordinate value of center of abnormality, 'r' radius of circle enclosing the abnormality by the radiologists. ROIs extraction performed by manual cropping operation considering the center of the

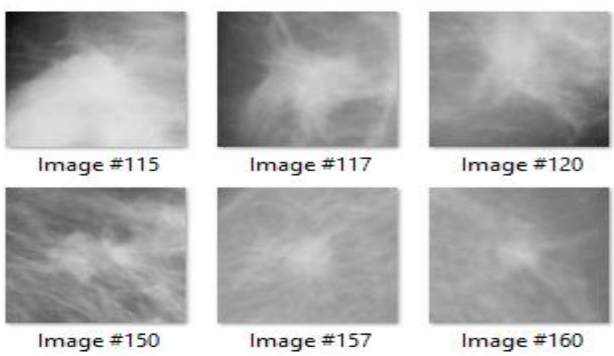
abnormal area as the center of ROI and taking the approximate radius (in pixels) of a circle enclosing the abnormal area as shown in Fig. 2. For the extraction of normal ROI, the same cropping procedure has been performed on normal mammographic images with random selection of location. In this work, all the ROIs are resized in to 256 x 256 for uniformity. In this phase, the rectangular ROIs are extracted and the ROIs are free from the background information and artifacts, which is defined by equation (1) [7].

$$\boldsymbol{I_{ROI}} = \boldsymbol{I}[\boldsymbol{X_c} - \boldsymbol{r}, (\mathbf{1024} - \boldsymbol{Y_c}) - \boldsymbol{r}, \mathbf{2}\boldsymbol{r}, \mathbf{2}\boldsymbol{r}] \qquad \textbf{(1)}$$

In addition, we used the following equation on the dataset of 117 to detect cancer region, the algorithm detect 115 images of 117 without any errors, where two images are shifted because image as the dataset center. See figure 2



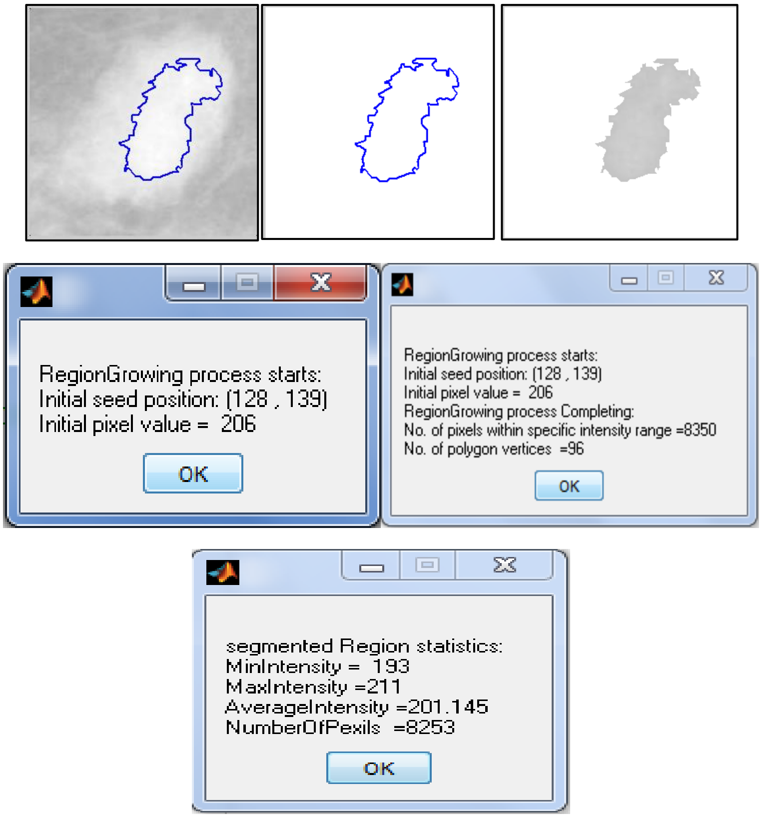
A-Samples of benign images



B-Samples of Malignant images

Figure 2 shows the Samples of dataset used to detect cancer area benign and malignant images

These images then used to extract the cancer area with edge map using region growing method.in this algorithm, many regions can be extracted accurately and fast.

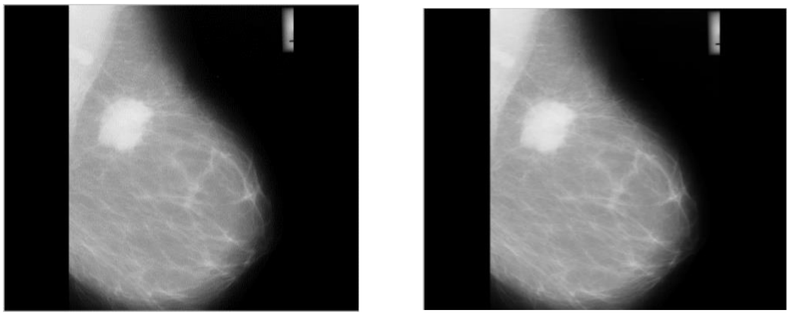


5-Mammogram images Pre-Processing

This section explains shortly some preprocessing steps to improve the results as follows:

3.2.1 Grayscale image

a grayscale digital image is an image in which the value of each pixel carries only the intensity information. Images of this sort, also known as black-and-white, are composed exclusively of shades of gray, varying from black at the weakest intensity to white at the strongest. Usually, each pixel value lies between 0 and 255. Sample images of mammogram images are shown in Figure (1). In a mammogram image each pixel has one color (grayscale), that is, carries only the intensity information. Pixels range from least intense (black) to most intense (white). Pixel values are usually range from 0 to 255. In fact a grayscale color is one in which the red, green and blue components all have equal intensity in RGB space.so using one band of them.

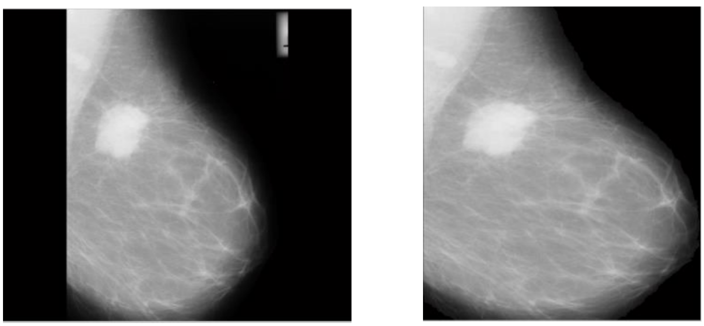


(a) **(b)**

Figure 3- mdb184 (a) original color image (b) convert into gray level.

6-Remove noise and select Region of Interest

We applied to the images two techniques: an image enhancement and a cropping operation one see figure 4. The first one was applied median filter to remove noise then employed in order to cut the black parts of the image as well as the existing artefacts such as written labels etc. For most of the images in our dataset, almost 50% of the whole image comprised of a black background with significant noise. Cropping removed the unwanted parts of the image usually peripheral to the area of interest.



(a) **(b)**

Figure.4 (a) remove noise by median filter (b) select ROI (cropped)

6.6.1 Median filter

An image can filtered either in the spatial frequency or in the frequency domain. The filter used in spatial domain, median filter 3 × 3, The noise in the mammogram is reduced using median filtering. Median filtering is as same as an averaging filter, in that each output pixel is set to an average of the pixel values in the neighborhood of the corresponding input pixel. Here, the value of an output pixel is determined by the median of the neighborhood pixels, rather than the mean. The median filter is better than the mean filtering. Median filtering is less sensitive therefore it is used to remove these outliers without reducing the 'sharpness of the image[4].

6.6.2 Enhancement by Background Removal

The initial process of this research begins from reading the mammography image that has mass with format .jpg in Matlab program. The result of the image reading was then segmented using thresholding method to separate the object and the background. After the image of the segmentation results in the form of binary images obtained, the morphological operation to improve the segmentation result and eliminate noise then conducted. The next stage is to determine the location (coordinates) of the midpoint of the mass.The area of mass of the morphological process results was then calculated using the following equation:

$$\boldsymbol{Area} = \sum_{i}^{N} \sum_{i}^{M} f(i,j) \qquad \textbf{(2)}$$

Whenever always the breast area is biggest than label, choose the biggest one and delete all the other.

7. Clustering algorithms

7.1 K-Means Clustering

Simply speaking it is an algorithm to classify or to group your objects based on attributes/features into K number of group. K is positive integer number. The grouping is done by minimizing the sum of squares of distances between data and the corresponding cluster centroid. Thus, the purpose of K-mean clustering is to classify the data.

The basic step of k-means clustering is simple. In the beginning, we determine number of cluster K and we assume the centroid or center of these clusters. We can take any random objects as the initial centroids or the first K objects can also serve as the initial centroids.

Then the K means algorithm will do the three steps below until convergence

Iterate until *stable* (= no object move group):see figure 5

1. Determine the centroid coordinate
2. Determine the distance of each object to the centroids
3. Group the object based on minimum distance (find the closest centroid)

In this paper our image clustered into different clusters one of them our interest region

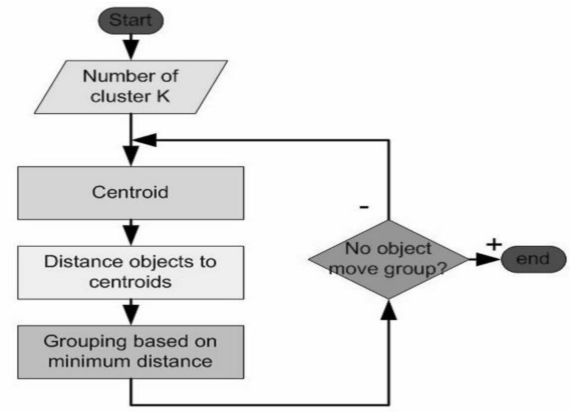


Figure.5 k-means algorithm steps

8. Proposed method

The proposed method includes the following steps:

- Read mammogram image file
- Convert to gray scale image
- Remove noise by median filter
- Select region of interest
- Then determine the number of regions (n) to segment, where the image is used as the input image of the region-growing algorithm.
- Select the initial position within the selected area in gray scale or clustered images , see figure(6), to obtain the pixel position associated with the current axis.

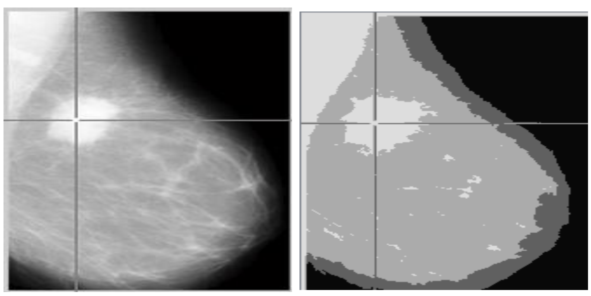


Figure .6 select the initial position in gray and clustered image (4 regions)

- Set the positive threshold to 10% of the value (maximum) in the input image.
- If have many regions, obtain the initial pixel value and add the initial pixels to the queue.
- Use the above parameters to start the growth algorithm. See figure 7

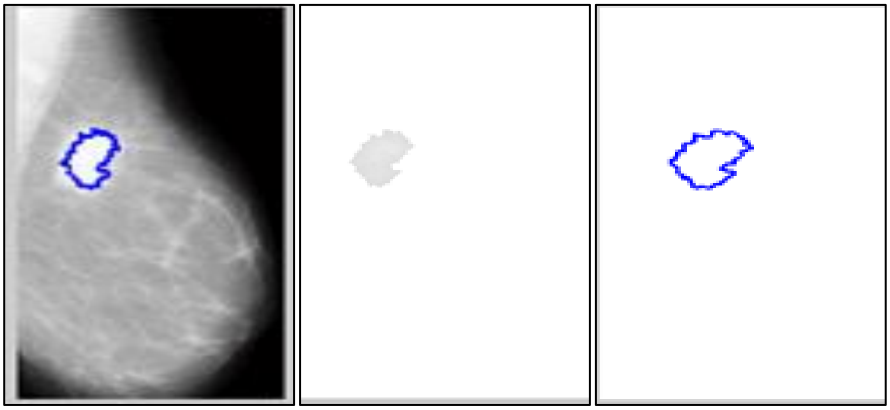


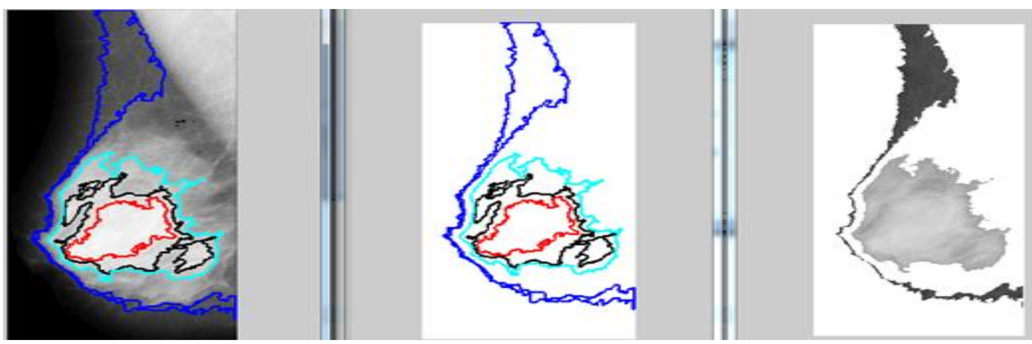
Figure.7 extracted one region by using seeded region growing

the charactristic of the region as follows:see table.1

Table.1 shows region growing method with some features extractions

Process	Initial position	Initial pixel values	No. of pixel in the intensity range	No.of polygons	Max Intensity	Min intensity	Average intensity	Total no. of pixels
Region growing	402 151	230	11020	123	217	235	225.4858	10896

For multi regions firstly select the seeds point then save it to the queue then the region are segment and extract from one to other as the fig .8. Segmenting the region depending on threshold that use positive threshold value to 10% of (max-min) value in the input image.



(a) (b) (c)

figure.8 segmenting the breast image into four regions(a) with their edge maps(b) and extracted regions (c).

The characteristics of these regions as in table 2:

Table 2 shows regions characteristics of figure.8 above

Region color	Initial position (x,y)	Initial pixel values	No. of pixel in the intensity range	No.of polygons	Max intensity	Min intensity	Average intensity	Total no. of pixels
Blue	230 248	73	36063	665	96	56	73.41	35361
Red	592 250	210	20696	252	222	195	207.10	20478
Black	490 188	192	44513	456	222	173	198.19	44069
light blue	422 170	170	74142	414	222	140	187.195	73702

The importance we gained in this article is that we can divide and extract any number of regions in the original image, we can select any point in the original image as our reference point, then we use the growing algorithm area to test all the neighbors. The seed point is equal to the intensity value (maximum - minimum) * 10% of the original image. If the pixel density is within this value, it will be associated with one region; otherwise, it belongs to the region boundaries. Another technical paper was used to cultivate the area that used the seed point in the center of the original image as primary seed, and then identified the following seeds from the connected pixels in the region according to the growth formula. Then use the force-based similarity and Otsu adaptive threshold to calculate the standard stop growth formula. Other researchers use color images and merge Methods to obtain the final decimal function of the entire image,

8. Conclusion

The suggested method shows that it is easy to determine the cancer images, the number of regions, the number of pixels in a split area, and the edge map, and we can extract the divided areas. On the other hand, the boundaries of fragmented areas are accurate. This is very useful, for example, to highlight small tumors very close to dense glandular tissue in the medical image and calculate the number of pixels (area Measure the effect of the drug before and after taking the medicine. In addition, it is useful to extract any area of

the image separately to diagnose and analyze images. On the other hand, if we first use assembly techniques such as watersheds or k means to include gray images, then using our method, we can provide very accurate edge maps, area extraction and image segmentation results. Finally, this method has been tested on many medical images. The

results of experiments on the computer show that this method can quickly, extract areas of interest in gray biomedical images, and show that the proposed method is effective and possible.
We recommend our method of automatically diagnosing cancer and tumors in medical images. Therefore, we suggest comparing and developing our methods with a team of expert doctors based on output. Also It is recommended to use clustering algorithms (eg, watersheds and K methods) to assemble images in their standard areas for very accurate results, then use our method of separating different applications from one another.[6]

References

[1] A. M. Ahmad, S. A. Mahmud, A. M. Ahmad, G. Muhammad, S. A. Mahmud, and J. F. Miller, "Breast Cancer Detection Using Cartesian Genetic Programming evolved Artificial Neural Networks Breast Cancer Detection Using Cartesian Genetic," no. January 2014, 2013.

[2] C. Science and T. Khulna-, "Prediction of Breast Cancer Using Support Vector Machine and K-Nearest Neighbors," pp. 21–23, 2017.

[3] http://peipa.essex.ac.uk/info/mias.htm

[4] K. Doi, "Computer-aided diagnosis in medical imaging : Historical review , current status and future potential," vol. 31, pp. 198–211, 2007.

[5] P. Smitha, "A Review of Medical Image Classification Techniques," no. Icvci, pp. 34–38, 2011.

[6] P. K. Jain, "An Adaptive Single Seed Based Region Growing Algorithm for Color Image Segmentation," no. December 2013, 2014.

[7] R. N. Panda, M. A. Baig, B. K. Panigrahi, and M. R. Patro, "Efficient CAD system based on GLCM & derived feature for diagnosing Breast Cancer," vol. 6, no. 4, pp. 3323–3327, 2015.

[8] S. Dehghani and M. A. Dezfooli, "Breast Cancer Diagnosis System Based on Contourlet Analysis and Support Vector Machine," vol. 13, no. 5, pp. 1067–1076, 2011.