Science World Publication

# Journal of Computational Science and Information Technology

journal homepage: https://www.sciworldpub.com/journal/jcsit


2020, Vol. 1, No. 1, 19-28

**Research Article**

# Revealing Breast Cancer in Mammography via K-Means Clustering Algorithm

Gajendra Sharma

Department of Computer Science and Engineering, Kathmandu University, Dhulikhel, Kavre, Nepal

**Abstract:** Breast cancer is the most common cause of death in women and the second leading cause of cancer deaths worldwide. Primary prevention in the early stages of the disease becomes complex as the causes remain almost unknown. Detection of the masses and their spread in mammogram is important for radiologists. It is also important to detect the shape of the contour or boundary to delineate malignant and benign lesions as malignant lesions have speculated or ill-defined boundary and benign mass have smooth boundary. Automatic detection of boundary helps the doctors in analyzing the lesion in less time and prevents unnecessary biopsies. In this Mammogram image enhancement was done using homomorphic filtering and adaptive histogram equalization. The enhanced mammogram image is segmented using K-means clustering and extracted geometric features from the lesions. The 25 range value of radius (R) between the maximum and minimum value of radius classify malignant and benign are done by distance versus angle of signature.

**Key words:** Breast Cancer Detection, Mammography, K-Means Clustering, Boundary

## 1. INTRODUCTION

The advances in medical imaging over the last three decades have greatly improved the type and quality of medical care that is available to the patient. Medical images are rich in information that can be used for diagonsis and subsequent medical interventions. Information provided by medical image has become an indispensable part of today's patient care. Cancer is the unrepressed development of unusual cells in the body which account for the most dangerous and life threatening diseases in the world. One in 8 deaths in the world is due to cancer. Presently breast cancer is a leading cause of death among women and second main cause of death after lung [4].

The common characteristics of the medical images like unknown noise, poor image contrast in homogeneity, weak boundaries and unrelated parts will affect the content of the medical images. This problem rectified by pre-processing techniques. The preprocessing are fundamental steps in the medical image processing to produce better image quality for segmentation and feature extractions. The preprocessing steps deal with image enhancement, noise and special mark removal. The image segmentation stages several method existed for automatic and semiautomatic medical image segmentation [5].

Early detection and diagnosis of breast cancer are essential for successful treatment. Previously, X-ray (radiograph) which is a non invasive medical test were used by physician for diagnosis of breast cancer but this required exposing a large part of the body to a small dose of ionizing radiation to produce pictures of the inside of the body. Currently, mammography and ultrasound are the basic

**Received:** Feb-09-2020 **Accepted:** Feb-25-2020 **Published:** Mar-12-2020

**Corresponding Author:** Gajendra Sharma, Department of Computer Science and Engineering, Kathmandu University, Dhulikhel, Kavre, Nepal, Email: gajendra.sharma@ku.edu.np



**How to Cite**: Gajendra Sharma, 2020. Revealing Breast Cancer in Mammography via K-Means Clustering Algorithm. *Journal of Computational Science and Information Technology,* 1*:* 19-28.

imaging techniques for the detection and localization of breast tumors. Mammography is specialized medical imaging that uses a low-dose X-ray system to see inside the breasts. A mammography exam, called a mammogram, aids in the early detection and diagnosis of breast diseases in women. Three recent advances in mammography include digital mammography, computer-aided detection and breast tomosynthesis. The segmentation is an image analysis technique. The term segmentation is usually used to describe the process of selecting a specific object or objects from an image. There are three approaches for the segmentation of images: manual segmentation, Automatic segmentation and Semi-automatic segmentation. Manual segmentation involves an expert observer outlining the object of interest in the images on computer screen or paper or film. Manual segmentation is very time consuming, subject to human error and has poor intra-observer reproducibility. In automatic segmentation, image processing tools are used. There should be proper match between references set of images and test images. Semi-automatic segmentation is compromises of two approaches [1]. The human observer usually starts the segmentation and has opportunity to make corrections, but as of work as possible is automated. In this study, semi-automatic segmentation approach is used. Well-defined classification and its summary is very helpful to the radiologist and pathologist and reduced the unnecessary biopsy.

The main objectives of this study are to detect the presence of breast cancer using image segmentation and extraction geometric features from the lesion and its classifications.

Many algorithms are developed for segmentaion of medical images. Segmentation of mass is crucial task in mammograms. Work has been done on segmentation of mass in past to know the spread of spiculation in the breast tissue. Mean shift algorithm and Fuzzy C-means and active contour models are used for the detection of masses [4]. Breast density is calculated by segmenting fibrogladular tissuses [5]. Suspicious focal areas are found for testing morphologic concentric layer (MCL) criteria, to detect mass region in mammogram [2]. Gradient Vector Flow (GVF) snake and multi-scale analysis using Gaussian pyramid have been proposed by Yu *et al.*, to segment masses in mammogram [10]. At first applied Gaussian pyramid to make the image coarse, so that GVF snake is able to converge to the mass contour easily and quickly with less computation. Shape features like elongatedness, eccentricity, Euler number, Max Radius, Min Radius were used to distinguish four different shapes round, oval, lobular, irregular of mass by using C5.0 decision tree algorithm [8]. Gabor filter banks are used for extracting local spatial textural properties of masses at different orientations and scales [3]. Multilevel wavelet decomposition is proposed to extract mean, variance, standard deviation, entropy and mean of absolute deviation from wavelet components [9]. Boundary extraction of this mass is also very important, so that radiologists can judge edge profile acutance benign or cancer. Rangayan *et al.*, proposed a region-based measure shape features like compactness, Fourier descriptors, central polygonal approximation and measured shape features like compactness, Fourier descriptors, central invariant moments and chord- length statics to distinguish between circumscribed and spiculated tumors [7].

**Cancer cell features extraction:** Cell is basic unit of life and all cancer begins in the cell. Figure 1 shows that normal cells have large cytoplasm, single nucleus, single nucleolus, fine chromatin and smooth cell border whereas cancer cells have scanty cytoplasm, multiple nuclei, multiple and large nucleoli, coarse chromatin and irregular cell border [6]. These are key features which are used for recognition of cancer using image-processing techniques.

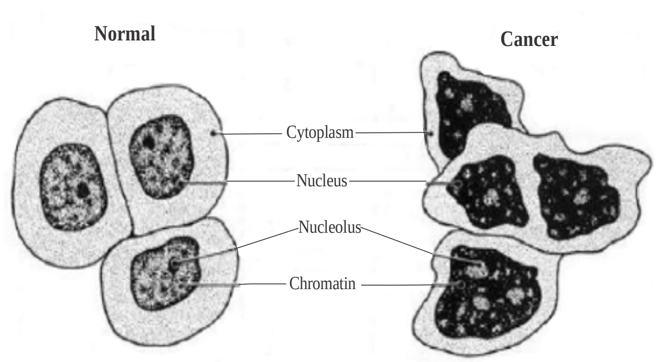


Fig. 1. Normal cell and cancer cell

## 2. METHODOLOGY

Medical image segmentation is an essential step for most subsequent image analysis tasks. The segmentation of an atomic structure in the breast plays a crucial role in breast imaging analysis (Fig. 2). Successful numerical algorithms can help researchers, physicians and neurosurgeons to investigate and diagnose the structure and function of the breast in both health and disease. This has motivated the need for segmentation techniques that are robust in application involving a broad range of anatomic structure, disease and image type. The problems associated with segmentation have been well studied and a large number of approaches have been developed, many specific to a particular image. General approaches to segmentation can be grouped into three classes: Pixel-based methods, Region methods, and Edge based methods.

### 2.1. Image pre-processing

Mammograms breast cancer images have the ability to assist physian in detection disease caused by cell abnormal growth. Developing algorithms and software to analyze these images may also assist physicians in their daily work. Microcalcification are tiny calcium deposits in breast tissues. They appear as small bright spots on mammograms. Since microcalcification are small and subtle abnormalities, they may be overlooked by examing radiologist.

### 2.2. Homomorphic filtering

Homomorphic filtering is used to remove multiplicative noise. Illumination and reflectance are not separable, but their approximate locations in the frequency domain may be located. Illumination and reflectance combine multiplicatively, the components are made additive by taking the logarithm of the image intensity, so that these multiplicative components of the image can be separated linearly in the frequency domain. Illumination variations can be thought of as a multiplicative noise, and can be reduced by filtering in the log domain.

An algorithm based on hybrid approach combination of both frequency domain homomorphic filtering and spatial domain morphology as described by Chaofu *et al*., [1] and adaptive histogram equalization technique to the output of hybrid approach. Homomorphic filtering is applied to the input image to improve the contrast of image and morphological operations are applied to remove the noise and to smooth the edges of the image.

### 2.3. Top hat transform

In mathematical morphology and digital image processing, top-hat transform is an operation that extracts small elements and details from given images. The top-hat

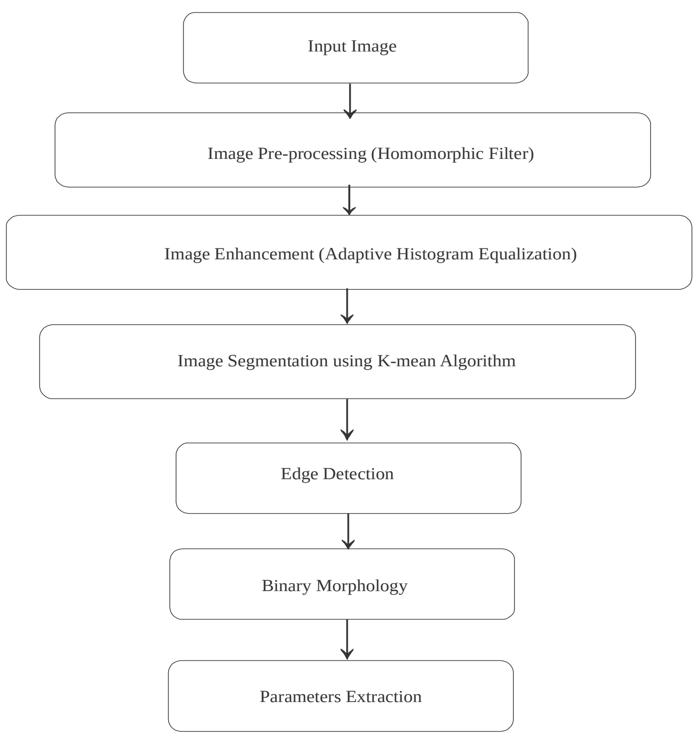


Fig. 2. Segmentation process diagram

transform is defined as the difference between the original image and its opening. The opening of an image is the collection of foreground parts of an image that fit a particular structuring element. An opening is erosion followed by dilation. The top-hat finds the intensity peaks that match the structuring element.

**2.4. Bottom hat transform**

The bottom-hat transform is defined as the difference between the closing of the original image and the original image. The closing of an image is the collection of background parts of an image that fit a particular Structuring Element (SE). The difference then yields an image with only the removed objects. The top-hat is used for light objects on a dark background and the bottom-hat for dark objects on a light background. An important use of top-hat transformation is in correcting the effects of non-uniform illumination. Bottom- hat morphological operator subtracts input image from result of morphological closing on the input image. Applied to binary image, the filter allows getting all object parts, which were added by closing filter, but were not removed after that due to formed connections/fillings.

**2.5. Enhanced image**

Morphological operators are used to analyze binary image, however these operators can be extended as applications on grayscale images. Top hat and Bottom hat transformations are two examples of morphological sharpening techniques. These transformations are executed using either the opening or closing of the original image. The process of adding the original image to the Top-hat transform and subtract the Bottom-hat transform are useful for enhancing detail in an image if shading is present. Let the output of top hat transform be thf1, the output of bottom hat transform be bhf and G is an input image.

**2.6. Histogram**

Histogram is a graphical display of tabular frequencies. An image histogram is a type of histogram which acts as a graphical representation of the tonal distribution in a digital image. The histogram of a digital image with gray levels in the range [0,L-1] is a discrete function: Where $r_k$ is $k^{th}$ gray level and $n_k$ is the number of pixels in the image having gray level $r_k$.

**2.7. Binary thresholding**

Thresholding is a process of binarization of an image. It is the simplest method of image segmentation. During the threshold process, individual pixels in an image are marked as “object” pixels if their value is greater than some threshold value (assuming an object to be brighter than the background) and as “background” pixels otherwise that, based on the contiguity hypothesis, a test data d is expected to have the same label as the training data located in the local region d. Binary thresholding converts the gray scale image I to a binary image. The output image replaces all pixels in the input image with luminance greater than level with the value 1 (white) and replaces all other pixels with the value 0 (black), specify level in the range [0, 1]. Therefore, a level value of 0.5 is midway between black and white.

**2.8. K-means clustering algorithm**

An algorithm to group objects into a K number of clusters based on features, where K is a positive integer number. The image segmentation of mass region is done by K-means algorithm. We consider the input as image pixels and their features are their grey-level values. The algorithm aims at minimizing sum of any pixel to cluster centroid distances, we have chosen Euclidean distances as distance measure. K-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. K-means is one of the simplest unsupervised learning algorithms that solve the well- known clustering problem. Cluster center, is an indicator of the distance of the n data points from their respective cluster centers (Fig. 3).

The n sample feature vectors $x_1$, $x_2$, ..., $x_n$ all from the same class, and we know that they fall into k compact clusters, k < n. Let $m_i$ be the mean of the vectors in cluster i.

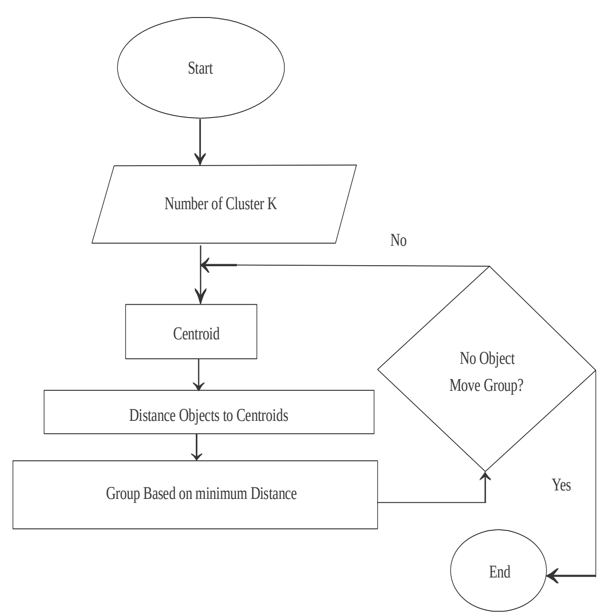


Fig. 3. Flowchart of K-means algorithm

If the clusters are well separated, we can use a minimum-distance classifier to separate them. That is, we can say that x is in cluster i if ‖ x - $m_i$ ‖ is the minimum of all the k distances. This suggests the following procedure for finding the k means:

- Make initial guesses for the means $m_1$, $m_2$, ……., $m_n$
- Until there are no changes in any mean
- Use the estimated means to classify the samples into clusters
- For i from 1 to k
- Replace mi with the mean of all of the samples for cluster i
- end_for
- end_until

**2.9. Edge detection**

Edge detection includes a variety of mathematical methods that aim at identifying points in a digital image at which the image brightness changes sharply or, more formally, has discontinuities. The same problem of finding discontinuities in 1D signals is known as step detection and the problem of finding signal discontinuities over time is known as change detection. Edge detection is a fundamental tool in image processing, machine vision and computer vision, particularly in the areas of feature detection and feature extraction. Edge Detection is by far the most common approach for detecting discontinuities in gray levels.

**2.10. Binary morphology**

Opening and closing are two important operators from mathematical morphology. They are both derived from the fundamental operations of erosion and dilation. Like those operators they are normally applied to binary images, although there are also gray level versions. The basic effect of an opening is somewhat like erosion in that it tends to remove some of the foreground (bright) pixels from the edges of regions of foreground pixels.

**2.11. Morphological gradient**

Dilation acts like a local maximum operator. Erosion acts like a local minimum operator. Combination of dilation, erosion, and image subtraction gives morphological gradient. The dilation thickness region in an image and the erosion shrinks them. Subtraction operation tends to remove the constant intensity area and edges are enhanced. The f is an input image.
Where, M(x, y) is the morphological gradient.

**2.12. Data collection**

Mammogram image of breast cancer was obtained from National Cancer Hospital Jawalakhel, Lalitpur Nepal. Total 15 cancer and 10 mammogram images were used to extract the geometric features of the segmented images.

Image processing tools of Matlab 2014 and Spring Tool Suite were used as programming tool. The Spring Tool Suite is an Eclipse-based development environment that was customized for developing Spring applications. Mysql data System was used to access and manipulate data.

## 3. RESULTS AND DISCUSSION

**3.1. Image pre-processing and enhancement**

The proposed algorithm based on hybrid approach combination of both frequency domain homomorphic filtering and spatial domain morphology and adaptive histogram equalization technique to the output of hybrid approach. Homomorphic filtering was applied to the input image to improve the contrast of image and morphological operations were applied to remove the noise and to smooth the edges of the image (Fig. 4).

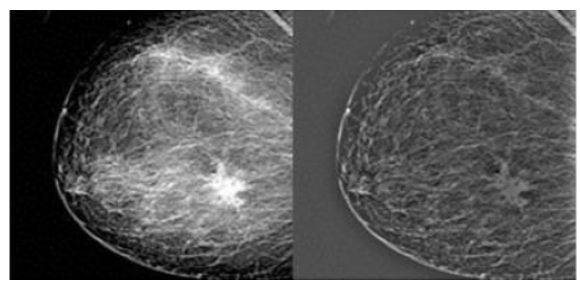

Fig. 4. Input and homomorphic filtered image

**3.2. Image Segmentation**

K-Means clustering algorithm and morphological operators were used to segment mass and extract the border. The procedure of image segmentation consists of :

**Step 1:** K-means Clustering
**Step 2:** Morphological operations
**Step 3:** Morphological gradient

This is an algorithm to group objects into a K number of clusters based on features, where, K is a positive integer number. We segment mass region using K-means algorithm (Fig. 5). We considered the input as image pixels and their features are their grey-level values. The algorithm aims at minimizing sum of any pixel point to cluster centroid distances, we had chosen Euclidean distance as distance measure.

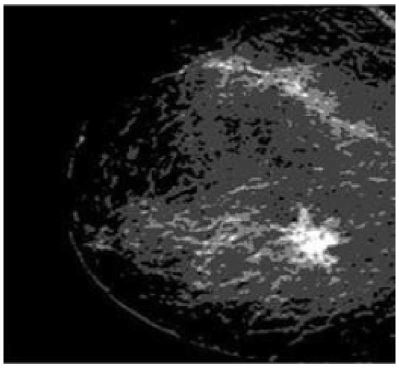

Fig. 5. Image segmentation by K-means clustering algorithm

### 3.3. Geometric Features Extraction

Geometric feature learning is a technique combining machine learning and computer vision to solve visual tasks. Feature extraction techniques are applied to get features that will be useful in classifying and recognition of images. Feature extraction techniques are helpful in various image processing applications e.g., character recognition. As features define the behavior of an image, they show its place in terms of storage taken, efficiency in classification and obviously in time consumption also. Geometric to Features of lesion boundary can be used to characterize malignant or benign lesion (Table 1,2). Seven morphologic features were extracted from each lesion to describe features such as shape, contour and size.

a) **Area:** Number of pixels contained in the lesion. Greater the value of area, it is more likely the lesion is malignant
b) **Perimeter:** The distance around the boundary of the region. Regionprops computes the perimeter by calculating the distance between each adjoining pair of pixels around the border of the region. Perimeter is the circumference of Lesion
c) **PA_ratio:** It is the ratio of perimeter to area of the lesion
d) **L: S Ratio:** It is the length ratio of the major (long) axis to the minor (short) axis of the equivalent ellipse of the lesion. If L: S ratio is more, it is likely the lesion is malignant
e) **ENC (Elliptical Normalized Circumference):** Anfractuosity is common morphological feature for malignant contour. ENC is circumference ratio of the lesion and its equivalent ellipse. Anfractuosity of a lesion contour is characterized by ENC

Seven geometric features are calculated from the 15 cancer and 10 benign mammogram images. Most of cancer mammogram have the greater value in comparison to the benign mammogram in terms of area and ENC. It is seen that greater the value of area and ENC, it is more likely the lesion is malignant. Geometric features are features of objects constructed by a set of geometric elements like

Table 1. Features extraction of cancer

| Image | Area | Perimeter | PA Ratio | Major Axis | Minor Axis | LS Ratio | ENC |
|---|---|---|---|---|---|---|---|
| C1 | 245 | 17.6619 | 0.3474 | 29.122 | 119.801 | 1.4707 | 55.486 |
| C2 | 1428 | 42.6402 | 0.2524 | 256.3406 | 69.6153 | 3.6822 | 133.95 |
| C3 | 348 | 21.0496 | 0.3728 | 34.2217 | 24.8925 | 1.3748 | 66.129 |
| C4 | 1741 | 47.082 | 0.0169 | 168.8417 | 119.563 | 1.4122 | 147.91 |
| C5 | 409 | 22.8201 | 0.3760 | 51.2765 | 36.0112 | 1.4239 | 71.691 |
| C6 | 1280 | 40.3701 | 0.2683 | 296.8708 | 94.355 | 3.1463 | 126.82 |
| C7 | 1022 | 36.0729 | 0.2868 | 152.2419 | 74.3404 | 2.0479 | 113.32 |
| C8 | 336 | 20.6835 | 0.4104 | 51.5467 | 34.4832 | 1.4948 | 64.979 |
| C9 | 136 | 13.1590 | 0.4614 | 30.1860 | 11.8856 | 2.5397 | 41.340 |
| C10 | 316 | 20.0585 | 0.4207 | 65.3548 | 41.1373 | 1.5445 | 63.015 |
| C11 | 1420 | 42.5206 | 0.3140 | 149.9097 | 89.5525 | 1.6740 | 133.58 |
| C12 | 1722 | 46.8243 | 0.3338 | 135.3552 | 81.9427 | 1.6518 | 147.10 |
| C13 | 1054 | 36.6332 | 0.3131 | 116.4685 | 73.6592 | 1.5812 | 115.08 |
| C14 | 2885 | 60.6077 | 0.2765 | 236.2648 | 77.1992 | 3.0605 | 190.40 |
| C15 | 1762 | 47.3651 | 0.3196 | 143.0897 | 98.2015 | 1.4571 | 148.80 |

Table 2. Features extraction of benign

| Image | Area | Perimeter | PA Ratio | Major Axis | Minor Axis | LS Ratio | ENC |
|---|---|---|---|---|---|---|---|
| B1 | 168 | 14.6255 | 0.3857 | 368.5450 | 17.2290 | 21.3909 | 45.9473 |
| B2 | 548 | 26.4147 | 0.3587 | 78.0633 | 76.6882 | 1.0179 | 82.9842 |
| B3 | 319 | 20.1535 | 0.4292 | 58.9635 | 50.9702 | 1.1568 | 63.3141 |
| B4 | 608 | 27.8232 | 0.4027 | 118.0374 | 62.2037 | 1.8976 | 87.4091 |
| B5 | 134 | 13.0619 | 0.4229 | 25.9273 | 15.9948 | 1.6210 | 41.0353 |
| B6 | 52 | 8.1369 | 0.5318 | 11.2679 | 10.0193 | 1.1246 | 25.5627 |
| B7 | 325 | 20.3421 | 0.3967 | 52.1927 | 34.8037 | 1.4996 | 63.9067 |
| B8 | 319 | 20.1535 | 0.4292 | 58.9635 | 50.9702 | 1.1568 | 63.3141 |
| B9 | 579 | 27.1515 | 0.0529 | 267.2050 | 55.5585 | 4.8094 | 85.2911 |
| B10 | 115 | 12.1005 | 0.4449 | 21.9452 | 16.4682 | 1.3326 | 38.0149 |

points, lines, curves or surfaces. These features can be corner features, edge features, blobs, ridges, salient point's image texture (Fig. 6).

### 3.4. Quality Measures of Image Enhancement

Restoration of high-frequency information of an image is a common problem in image processing. High-frequency information is corrupted of either lost during various image corruption and degradation procedures like down sampling or blurring.

It is not possible to completely reconstruct lost high-frequency information, therefore artifacts appears in restored images. Typical artifacts of image enhancement algorithms caused by loss of the high frequency information are blur and ringing effect near sharp edges. Development of image metrics is important for the objective analysis of image resampling, deringing, deblurring, denoising and other image enhancement algorithms.

Image metrics perform comparison of the ground truth image and the restored image. Since the ground truth image is unavailable in most cases, the simulation approach is used. In this approach, artifact free images are corrupted to simulate the effect which is aimed to be suppressed by the being evaluated image enhancement algorithm. Then the corrupted images are restored using the given algorithm and compared to the corresponding reference images using image metrics. There exists large variety of image metrics ranging from simple but fast approaches like MSE, PSNR to more complicated metrics based on modeling the human visual system. Most of image metrics can provide the estimation of perceptual image quality but they cannot be used to develop effective image enhancement algorithms because they do not focus on typical artifacts caused by the corruption of high-frequency information. Two image enhancement algorithms can give the same metrics values but the results can be very different if the first algorithm processes edges well and corrupts non-edge area while the second one corrupts only edges.

Morphological operators are used to analyze binary image, however these operators can be extended as applications on grayscale images. Top hat and Bottom hat transformations are two examples of morphological sharpening techniques. These transformation are executed using either the opening or closing of the original image. The process of adding the original image to the Top-hat transform and subtract the Bottom-hat transform are useful for enhancing detail in an image if shading is present. Let the output of Top hat transform be thf1, the output of bottom hat transform be bhf and G is an input image B.

### 3.5. Entropy

Image entropy is a quantity which is used to describe the `business' of an image, i.e., the amount of information which must be coded for by a compression algorithm. Low entropy images, such as those containing a lot of black sky, have very little contrast and large runs of pixels with the same or similar DN values. An image that is perfectly flat will have entropy of zero. Consequently, they can be compressed to a relatively small size. On the other hand, high entropy images such as an image of heavily cratered areas on the moon have a great deal of contrast from one pixel to the next and consequently cannot be compressed as much as low entropy images.

### 3.6. Standard Deviation

It is a most widely used measure of variability or diversity used in statistics. In terms of image processing it shows how

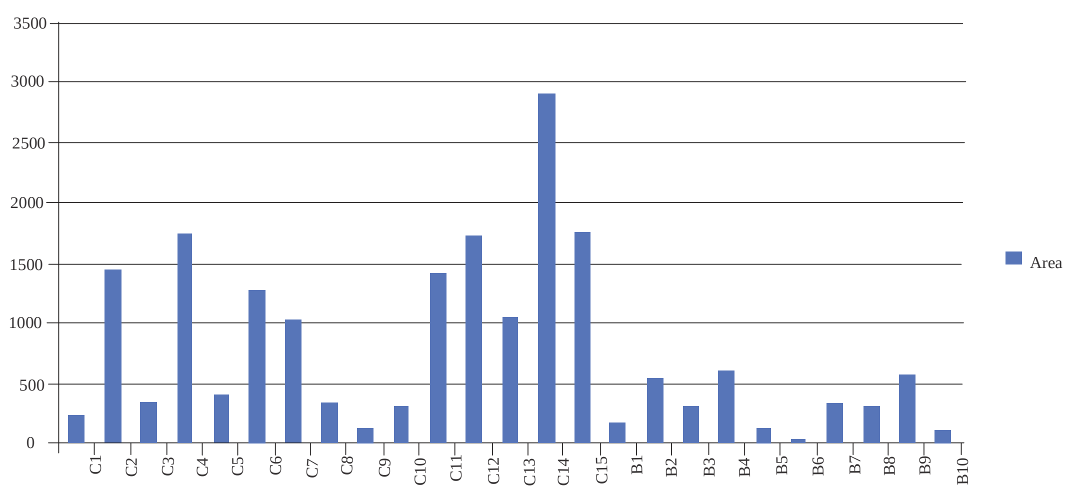


Fig. 6. Comparative analysis of area of cancer and benign

much variation or "dispersion" exists from the average (mean, or expected value).

### 3.7. Edge Based Contrast Measure

Edge detection includes a variety of mathematical methods that aim at identifying points in a digital image at which the image brightness changes sharply or, more formally, has discontinuities. It is expected that an enhanced image should have more edge pixels than the original image and this parameter of EBCM does measure the intensity of edge pixels in small windows of the image.

**E1:** Entropy of Homomorphic Filtered Image.

**E2:** Entropy of Enhanced Image using AHE.

**SD1:** Standard Deviation of Homomorphic Filtered Image.

**SD2:** Standard Deviation of Enhanced Image using AHE

**EBCM1:** Edge Based Contrast Measure of Homomorphic Filtered Image.

**EBCM2:** Edge Based Contrast Measure of Enhanced using AHE.

Homomorphic filtering is used to remove multiplicative noise. Homomorphic filtering is used for correcting non-uniform illumination in image. For image enhancement, the process of adding the original to to-hat transform and subtract the bottom-hat transform are used. Entropy, standard deviation and edge based contrast measure are the three parameters used to measure the quality of image. Table 3 and 4, shows that the value of quality measure parameters of the outputed enhanced image using AHE have the greater

Table 3. Quality measures of enhanced images in cancer

| Patient No. | E1 | E2 | EBCM1 | EBCM2 | SD1 | SD2 |
|---|---|---|---|---|---|---|
| C1(Cancer) | 3.3023 | 4.0535 | 58.4866 | 110.0120 | 0.0215 | 0.0648 |
| C2 | 3.5350 | 4.4489 | 95.6216 | 99.0381 | 0.0132 | 0.0363 |
| C3 | 2.9220 | 3.3419 | 85.7870 | 95.1428 | 0.0130 | 0.0338 |
| C4 | 3.5648 | 4.3746 | 111.5529 | 81.9789 | 0.0120 | 0.1237 |
| C5 | 2.6656 | 3.0532 | 93.0772 | 90.4355 | 0.0118 | 0.0443 |
| C6 | 3.4680 | 4.2274 | 90.3118 | 88.2642 | 0.0121 | 0.0369 |
| C7 | 2.4541 | 2.8501 | 60.1479 | 92.1521 | 0.0148 | 0.0549 |
| C8 | 2.7730 | 3.0913 | 97.9758 | 90.1452 | 0.0074 | 0.0252 |
| C9 | 1.7965 | 2.1380 | 45.2919 | 87.588 | 0.0168 | 0.0570 |
| C10 | 2.7979 | 3.2509 | 120.8388 | 86.9146 | 0.0076 | 0.0269 |
| C11 | 2.6002 | 3.1353 | 66.6428 | 90.4510 | 0.0141 | 0.0554 |
| C12 | 2.0944 | 2.5337 | 55.0490 | 88.4676 | 0.161 | 0.0580 |
| C13 | 2.722 | 23.1991 | 70.5318 | 92.8534 | 0.0139 | 0.0478 |
| C14 | 2.9045 | 3.5197 | 74.0608 | 90.8344 | 0.0175 | 0.0550 |
| C15 | 2.2406 | 2.6668 | 55.8464 | 87.9221 | 0.0158 | 0.0578 |

Table 4. Quality measures of enhanced images in benign

| Patient No. | E1 | E2 | EBCM1 | EBCM2 | SD1 | SD2 |
|---|---|---|---|---|---|---|
| B1(Benign) | 2.4258 | 2.9183 | 44.6844 | 96.3544 | 0.0359 | 0.0842 |
| B2 | 3.2388 | 3.7782 | 79.7023 | 150.5199 | 0.0188 | 0.0307 |
| B3 | 3.0914 | 3.8303 | 91.4749 | 111.0193 | 0.0075 | 0.0215 |
| B4 | 2.5049 | 3.0083 | 98.2106 | 149.3233 | 0.0126 | 0.0469 |
| B5 | 2.3660 | 2.8334 | 92.3206 | 126.5848 | 0.0135 | 0.0350 |
| B6 | 3.2005 | 3.7094 | 87.7859 | 171.2703 | 0.0073 | 0.0131 |
| B7 | 2.8993 | 3.6234 | 97.0811 | 69.0031 | 0.0135 | 0.0459 |
| B8 | 3.0914 | 3.8303 | 91.4749 | 111.0193 | 0.0075 | 0.0215 |
| B9 | 3.9796 | 4.5394 | 80.3545 | 124.8425 | 0.0071 | 0.0191 |
| B10 | 2.4264 | 2.7446 | 96.8110 | 126.5910 | 0.0149 | 0.440 |

Table 5. Value of signature radius (R) in cancer

| Patient's No. | Maximum value of R | Minimum value of R | Range |
|---|---|---|---|
| C1(Cancer) | 50.9218 | 14.9847 | 35.9370 |
| C2 | 115.2748 | 22.2520 | 93.0230 |
| C3 | 66.3737 | 14.3572 | 52.0065 |
| C4 | 119.2221 | 27.0065 | 92.2156 |
| C5 | 45.7967 | 2.8712 | 42.9255 |
| C6 | 100.0809 | 18.8418 | 81.2391 |
| C7 | 83.1588 | 17.0029 | 66.1599 |
| C8 | 48.0704 | 11.8403 | 36.2301 |
| C9 | 24.9662 | 4.5222 | 20.4440 |
| C10 | 50.1638 | 26.2927 | 23.8711 |
| C11 | 125.5841 | 32.6079 | 92.9762 |
| C12 | 132.5617 | 0.5652 | 131.9966 |
| C13 | 93.7184 | 10.333 | 83.3851 |
| C14 | 211.0641 | 3.0218 | 208.0423 |
| C15 | 127.8460 | 27.4866 | 100.3594 |

Table 6. Value of signature radius (R) in benign

| Patient No. | Maximum value of R | Minimum value of R | Range |
|---|---|---|---|
| B1 | 6.1175 | 4.0172 | 2.1003 |
| B2 | 45.8321 | 31.4855 | 14.3466 |
| B3 | 45.6880 | 29.2364 | 16.4516 |
| B4 | 93.7945 | 38.2859 | 55.5087 |
| B5 | 26.6977 | 18.4739 | 8.2238 |
| B6 | 8.2465 | 5.2980 | 2.9485 |
| B7 | 47.2062 | 12.4903 | 34.7159 |
| B8 | 45.6880 | 29.2364 | 16.4516 |
| B9 | 16.5686 | 0.7529 | 15.8157 |
| B10 | 16.3437 | 6.9952 | 9.3504 |

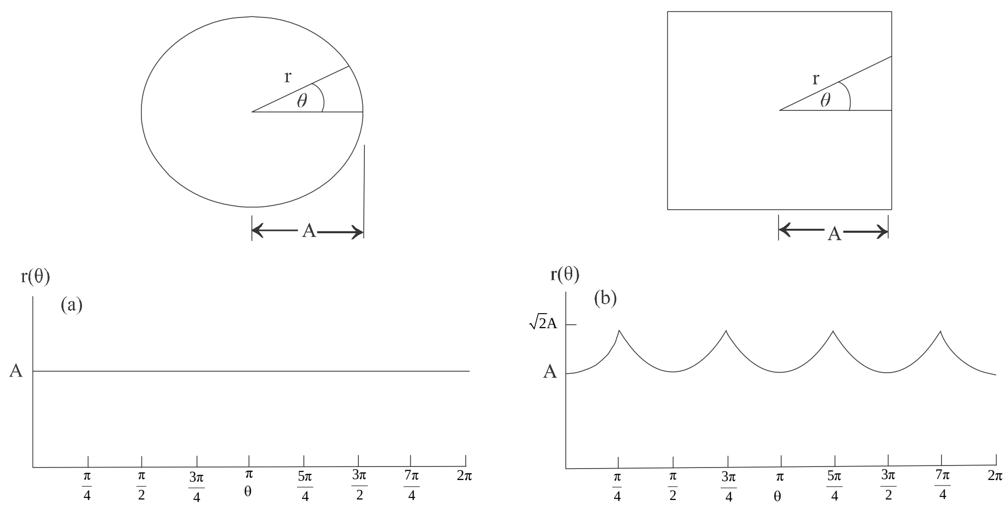


Fig. 7. Distance versus angle signature

value in comparison to the pre-processing outputed image i.e., Homomorphic filtered image.

### 3.8. Signature

A signature is a 1-D functional representation of a boundary. It is a plot of the distance from the centroid to the boundary as a function of angle. A signature is a 1-D representation of a boundary (which is a 2-D thing): it should be easier to describe. For example distance from the centroid versus angle. Signatures are invariant to translation. Signatures are invariance to rotation and depends on the starting point. The starting point could be the one farthest from the centroid. Scaling varies the amplitude of the signature and invariance can be obtained by normalizing between 0 and 1 (Fig. 7).

$r(\theta)$ is constant. The signature consists of repetitions of the pattern. The $r(\theta)=A\sec(\theta)$ for $0\leq\theta\leq\pi/4$ and $r(\theta)=A\csc(\theta)$ for $\pi/2 \leq \theta \leq \pi/4$ (Table 5, 6).

## 4. CONCLUSION

Automatic detection of boundary helps the doctors in analyzing the lesion in less time and prevents unnecessary biopsies. The shape of the contour or boundary to delineate malignant and benign lesions as malignant lesions have speculated or ill-defined boundary and benign mass have smooth boundary. In this study Mammogram image is enhanced using homomorphic filtering and adaptive histogram equalization. The enhanced mammogram image is segmented using K means clustering and extracted geometric features from the lesions. Geometric features of the border are also calculated. Geometric features of lesion boundary are characterized as malignant or benign lesion. Seven morphologic features are extracted from each lesion to describe features such as shape, contour and size. Classifications of malignant and benign are done by distance versus angle of signature. Image enhancement and segmentation methods are implemented to extract the border and distance versus angle of signatures is calculated. Signature value of range in malignant image is higher in comparison to benign image.

## 5. REFERENCES

[1] Chaofu, Z., M. Li and J. Lu, 2012. Mixed Frequency domain and spatial of enhancement algorithm for

infrared image. IEEE international Conference on Fuzzy Systemsand Knowledge Discovery. 29-31, May, 2012 IEEE,

[2] Eltonsy, N.H., G.D. Tourassi and A.S. Elmaghraby, 2007. A concentric morphology model for the detection of masses in mammography. IEEE Transactions on Medical Imaging. 29, May, 2007 IEEE, 880-889.

[3] Hussain, M., S. Khan, G. Muhammad, M. Berbar and G. Bebis, 2012. Mass detection in digital mammograms using mammograms. IET Conference on Image Processing, Springer, 82-91.

[4] Paramkusham, S., K.M.M.Rao and S.N.R. Prabhakar, 2013. Automatic detection of breast lesion contour and analysis using fractals through spectral methods. Advances in Computer Science. AETACS, Elsevier, 233-243.

[5] Paramkusham, S., K.M.M. Rao, B.V.V.S.N. Prabhakar Rao, 2013. Assessment of breast density and lesion feature extraction for early detection of breast cancer. Int. J. Comput. Sci. Commun. Technol., 12: 33-43.

[6] Raghavendra, B.S. and D. Narayana Dutt, 2010. Computing fractal dimension of signals using multiresolution boxcounting method. Int. J. Inf. Math. Sci., 6: 67-78.

[7] Rangayyan, R.M., N.M. El-Faramawy and O.A. Alim, 2007. Measures of acutance and shape for classification of breast tumors. IEEE Trans. Med. Imaging, 21: 23-43.

[8] Vadivel, A. and B. Surendiran, 2013. A fuzzy rule-based approach for characterization of mammogram masses into BI-RADS shape categories. Comput. Biol. Med., 43: 259-267.

[9] Vidhya, M., N. Sangeetha, M.N. Vimalkumar and K. Helenprabha, 2011. Early stage detection of cancer in mammogram using statistical feature extraction. Recent Advancements in Electrical, Electronics and Control Engineering, 15-17 Dec, 2011, IEEE,

[10] Yu, H., L. Li, W. Xu, W. Liu, J. Zhang and G. Shao, 2009. Gaussian pyramid based Multi- Scale GVF snake for mass segmentation in digitized mammograms. IEEE International Conference on Bioinformatics and Biomedical Engineering, 11-13, June, 2009, IEEE,