**IJCTA,** 9(10), 2016, pp. 4647-4660
© International Science Press

# Effective Fuzzy C Means Algorithm for the Segmentation of Mammogram images of Identify Breast Cancer

***T.Velmurugan **E.Venkatesan**

***Abstract :*** Fuzzy C-Means (FCM) algorithm is the frequently used algorithm, to examine the various forms of data and images based on its clustering ability. Particularly, FCM and its different methods utilized in the extraction of medical images to find the disease affected regions by means of segmenting the images on the basis of intensity. Mammography is one of the methods to find a tumor in the breast, which is helpful for the doctor or radiologists to detect the cancer. Presently breast cancer detection is very important for women to retain their life a long period.Doctors and radiologist can have a probability of leaving out the abnormality due to inexperience in the area of cancer detection.Clustering oriented segmentation is really a powerful technique, which are helpful for doctors and radiologists to analyze the mammogram images.FCM is applied for the identification of cancer affected regions in Mammogram images in this research work. Also, this work have analyzed mammogram images of normal, benign and malignant using the FCM algorithm. The technique is implemented by focusing a method for finding the tumor affected region in mammogram images in three ways, such as (*i*) Preprocessing the images by selection of region of interest, inverse method and boundary deduction method (*ii*) Segmenting the image by FCM algorithm and (*iii*) finding accuracyusing theclassification algorithms J48, JRIP, Support Vector Machines (SVM), Naive Bayes and CART. The experimental analysis shows that the performance of clustering algorithmsfor the predictionof tumor area and classification algorithms to find the accuracy.

***Keywords :*** ClassificationAlgorithms, Mammogram Images, Fuzzy C-Means Algorithm.

## 1. INTRODUCTION

The breast cancer is one of the most common cancers that found in females. It is a foremost cause of death in the world. A proper screening process can assist an early diagnosis of the cancer so that it is able to cut the death risk.The breast malignant neoplastic disease is mainly a common disease found in females and this disease is increasing every year worldwide.The lack of awareness initiatives, structured viewing, and reasonable treatment facilities continue to result in poor survival [3].Breast cancer is the second most common malignancy in India and also in the world. According to Indian council of medical research statistics, 10,000 breast cancer are being diagnosed every year in India and more than 70% of them are diagnosed at an advanced stage. By 2020, the incidence of breast cancer in India is awaited to be twofold.The cancer registry data indicate that urban women are at almost having doubled the risk of breast cancer than rural women.The aim of mammography is the benign detection of breast cancer. Digital mammograms take an electronic image of the breast and store it directly in a computer [24].

Breast abnormalities that can point out cancer are masses, calcifications, architectural distortions and bilateral irregularity. Segmentation is one of the most vital steps in computer aided detection/diagnosis, especially for the masses, for the reason that breast masses can have uncertain borders and are sometimes covered by glandular tissue in mammograms[25, 34, 37]. During the search for suspicious areas, it is possible that the masses of this case

* Department of Computer Science, D. G. VaishnavCollege, Chennai-600106, India E-Mail:velmurugan_dgvc@yahoo.co.in, venkatelumalai12@yahoo.co.in,

are overlooked by radiologists. The American Cancer Society projected that 1,84,300 women will be diagnosed with breast cancer and those 44,300 women will die from it. Some other survey showed approximately 7,20,000 new cases will be diagnosed worldwide per year.This accounts for approximately 20% of all malignant tumor cases [1].Breast cancer is a kind of cancer originating from breast tissue, most commonly from the inner lining of the milk ducts, milk passages that connect the lobules to the nipple or the lobules, milk-producing glands that supply the ducts with milk. It is the second leading cause of cancer death among women mainly those who are in the 40-55 age group [12].Detection and diagnosis of breast cancer in its early stage increase the successful handling and perfect recovery from disease.The most of the medical examinations not gross.On average, mammography will detect approximately 80-90% of the breast cancers in women without symptoms [2].Doctor or radiologists can miss the abnormality due to inexperience in the area of cancer detection.Segmentation is very valuable for doctor and radiologists to examine the mammogram [4].Analyzing mammogram images is the most challenging job in medical image processing.Fuzzy C- means algorithm analyzing such images for the efficient finding and diagnosis of the disease affected area by segmentation[7].

Data Mining is a convenient method of extracting patterns, which represents knowledge implicitly stored in datasets and focuses on matters linking to their feasibility, usefulness, effectiveness and scalability.Data are preprocessed by data cleaning, data consolidation, and information selection and information translation. Data mining functionalities are Classification, Association, Correlation analysis, Prediction, cluster analysis [5]. The clustering of objects is based on measuring the correspondence between the pair of objects using distance function. Thus, result of clustering is a set of clusters, where object within one cluster is further similar to each other, than to object in another cluster.The Cluster analysis has been broadly applied in numerous applications, including segmentation of medical images, information analysis, and image processing. Clustering is also called data segmentation in some applications because clustering partitions the huge data sets into groups according to their resemblance[4].Classification is a method used to mine models discussing important data classes or to predict the future data. Classification is a two-step process first step, learning or training step where data is analyzed by a classification algorithm. Second is testing step, where data are used for classification and to calculate the accuracy of the classification [6, 38].With these small introduction, the structureof the paper is organized as follows. Section 2 gives the background work of the paper via literature survey and Section 3discusses about Objectivesof research, Section 4 explores the methodologyusedfor this research work. The experimental work of this article is given in the section 5. Finally, section 6 concluded the research work via its findings.

## 2. LITERATURE SURVEY

There is a number of research work carried out by many peoples to find the tumor affected region in mammogram images. The breast cancer tumors are abandoned and abnormal proliferations of cells. Some grow in the breast itself, in such case they are termed primary. Others spread to this spot from somewhere else in the body through metastasis, and are termed as secondary. Primary breast tumors do not circulate to other body sites, and can be malignant or benign. Secondary breast tumors are always malignant. Both cases are potentially disabling and life threatening. Tumors in the breast originally start developing from the breast tissue itself. It is far more common in women than in men [19]. A research work carried byTaraSaikumar et al. in [8].In this, an effort is made to produce an adaptive FCM clustering algorithm for breast image segmentation for the detection of micro calcifications and also a computer based decision system for early detection of breast malignant neoplastic disease.Another researchwork done by GokilaDeepa. G, in which it is proposed that the work utilizes fuzzy C- means clustering for segmentation and PSO for clear identification of clusters. The PSO and FCM is a new approach, using this they have successfully segmented the suspicious breast cancer masses in digital mammogram images [9].

A research work analyzed byAnamikaAhirwar and R.S. Jadonis using SOM and FCM clustering techniques for tumor detection in digital mammography images. They further evaluate the statistical features of tumor like spot of tumor area, energy, entropy, idm, contrast, mean and standard deviation which helps the radiologist to study the statistical information about breast cancer, so that the doctors can give better treatment to the patients[10]. Aresearch work carried outbyShruthiAnand et al. in [11]. A hybrid of Fuzzy C-means algorithm and Self organizing map algorithm are applied to segment the input image and then categorize tumor affected breast images and normal breast images.Another work by Nalini Singh et al.[13]has presented a novel approach to recognize the presence of breast cancer mass and calcification in mammograms using image processing functions, k-means and Fuzzy C-Means clustering for clear identification of clusters. Combining these have successfully detected the breast cancer area in raw mammogram images.

Another work by N. Golestani, M. EtehadTavakol and E. Y. K. Ng have compared the performance criterion of three image segmentation methods: k-means, fuzzy c-means (FCM) and a region-based level set for segmentation of different breast thermograms.The enhanced level set algorithm (energy functional formulation) is used in this work. Their research have preferable efficiency, accuracy, and much more robust for initialization than the conventional level set methods [14]. Another research done byM. P. Sukassiniand T. Velmuruganhavecompared different methods that proposed by various researchers in segmenting the mammogram imagery. Each category shows its performance vibrantly. From this survey oriented research work, it is identified that most of the hybrid techniques yields better accuracy, sensitivity and specificity in order to segment and classify the mammogram images. They conclude that any hybrid technique performs well in segmenting the mammogram images compared with existing methods [15]. A research work carried out by Sandhya G et al. [16]. They have proposed a segmentation algorithm and compared with k-Means algorithm and FCM algorithm. They discovered that the modified k Means segmentation algorithm outperforms the benchmark k-Means algorithm and FCM algorithm. Further, the detection of abnormalities in human breast like calcification is used by the resultant mammogram.

## 3. OBJECTIVES OF RESEARCH

The FCM algorithm performance is well in many domains, particularly FCM algorithm its superiority in terms of its execution in predicting tumor affected regions in mammogram images. This research is really useful to doctors and radiologist, because to identify the cancer affected area in the breast with the help of fuzzy C-means clustering algorithm.In cancer identification there are so many tests are available. By using FCM algorithm, it is easy to deduct the cancer affected area very easily and accurately.The objective of this research work is to find the cancer affected regions in digital mammogram images by means of segmenting the images based on the intensity of the images. The steps involved in this work have preprocessing, clustering and classification. First step is to apply the preprocessing via selection of a region of interest, inverse method and the boundary deduction methodto mammogram images, the second step is to find breast cancer tumor area using fuzzy C-means clustering algorithm entering the number of clusters into some specific limited count like 4 or 5. The third stepis the use of classification algorithms to find the accuracy of predicted images based its pixel values. The main objective of this process is to find the cancer affected area bymeans of image clustering using FCM algorithms.

## 4. THE METHODOLOGY

Many methods used by various researchers for the analysis and findings of breast cancer in mammogram images. This research uses 50images for the analysis, which includes three types: normal, benign and malignantimages to find the affected and unaffected images of mammograms. The number of normal image is 1, benign images are 10 and malignant images are 39 in the data set. Sinceall the images are extracted by using the clustering and classification to find the tumor area and to find the accuracy respectively. The source code is written in MATLAB. As told, the method used for this research work is preprocessing, clustering and classification. The steps involved inthe proposed method in Figure 1.

**A setp by step procedure utilized in this work is given as follows :**

**Step 1:** Input the images for preprocessing.

**Step 2:** Preprocessing the images using region of interest, inverse method and boundary detection methods to remove the noise and outliers.

**Step 3:** Convert the mammogram image DICOM format into JPG format.

**Step 4:** Apply the FCM algorithm to find the affected region based on the intensity of the images.

**Step 5:** Enter the number of clusters.

**Step 6:** Display the tumor affected area by FCM algorithms via its output images.

**Step 7:** Find the number of pixels in each and every output of the FCM algorithm.

**Step 8:** Input the number of pixel values into the input for classification algorithms.

**Step 9:** Find the accuracy using the classification algorithms J48, JRIP, SVM, Naive Bayes and CART.

**Step 10:** Find the performance of classification algorithms based on its accuracy.

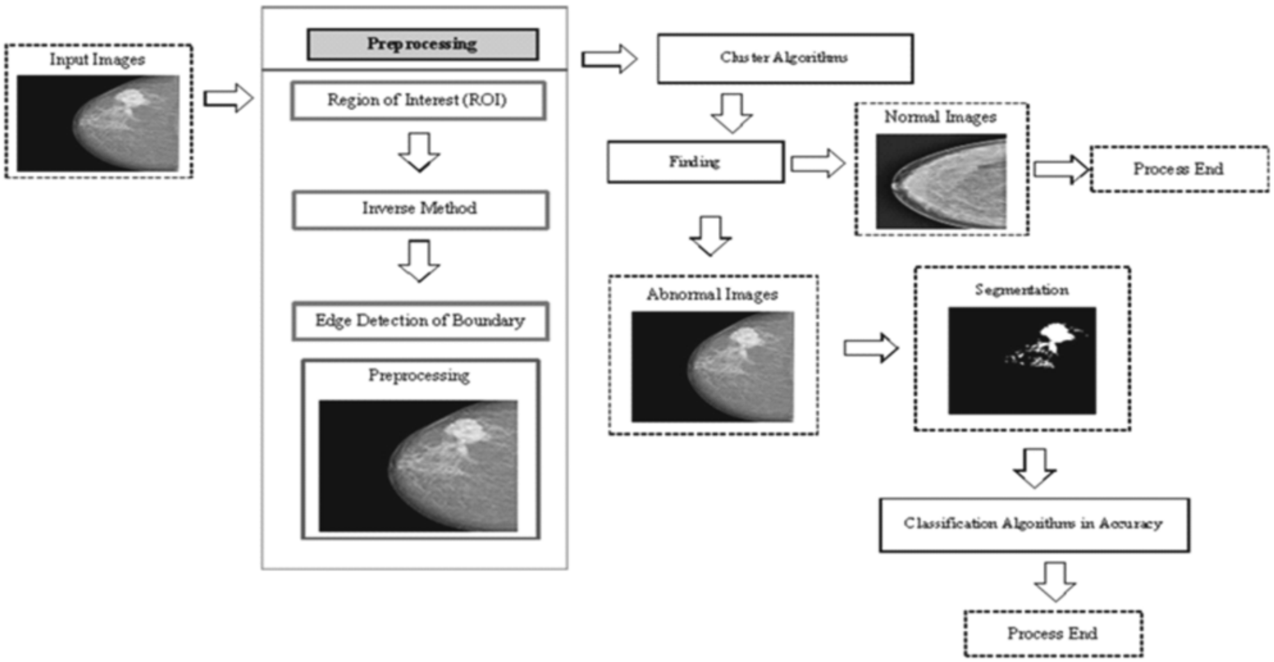


**Fig. 1. Proposed Method.**

Usually, the FCM algorithm utilizes in many domains and its performance is tested. FCM algorithm is a strong algorithm, which haveefficiency in analyzing the medical images very effectively. Also, some of the classification algorithms perform well in some applications and for some other applications these have efficiency [17].

## 4.1. Preprocessing Techniques

The main objective of the preprocessing is to develop the image quality to make it ready for further processing by removing or reducing the not related and surplus parts in the background of the mammogram images.This work uses the region of interest, inverse method and the boundary deduction method to remove the noises [28]. Segmentation is performed to partition the images into homogeneous regions and extract the affected regions [21, 20, 27, 30]. In digital mammograms, it is often useful to highlight regions of interest (ROI).Regions of interest in most cases will isolate those parts of the breast image that are of further interest to the radiologist. These regions of interest can be highlighted using segmentation techniques in image processing. There are several reasons why we want to detect such regions.Regions of interest highlighted using pseudo-color analysis makes the diagnosis easier, these regions can be x-rayed in greater detail for more information,the changes in these regions can be used for monitoring the effects of therapy. Automated analysis of breast cancer can be carried out if details on shape, texture and spectral information for these regions is available, computerized analysis in this manner can be used for training new radiologists on unseen cases.

## 4.2. Regions of interest (ROI)

Region of Interest is the region which covers the area of abnormalities seen on the mammograms [36, 39]. First, a study region ofinterest with n pixels in an image, mean value of pixels is P, and similarly of P in the Region of Interest is

$$P = \frac{1}{n}\sum_{i=1}^{n} f_i \tag{1}$$

Where $f_i$ are the individual pixel values. The variance of the ROI value is

$$X(P) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{Cov}(f_i, f_j)$$

or

$$X(P) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \rho_{ij}\sqrt{X(f_i)\,X(f_i)} \tag{2}$$

where Cov is covariance and $\rho_{i,j}$ is the correlation coefficient between pixel values *i* and *j* within the ROI. X(F) can be determined from independent approximations of the pixel variance values and the inter pixel correlation values. Let $\hat{V}$ be the average pixel variance within the ROI, for a small region, $\hat{V}$ can be approximated by the variance of the central pixel, and large regions, $\hat{V}$ can be estimated by averaging variance estimates from several pixels distributed throughout the region.

$$X(P) = \frac{\hat{V}}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \rho(\mathrm{dis}_{i,j}). \tag{3}$$

The above assessment formula for the variance of a particular pixel, and the correlation coefficient $\rho_{i,j}$ is approximated by a function ρ(dis) that is the mean correlation between pairs of pixels separated by distance dis. inserting into (2) the approximations of $\hat{V}$ for $X(f_i)$ and $\rho(\mathrm{dis}_{i,j})$ for produces [32,33,35].

### 4.3. Inverse Method

Let A beinvertible n x n matrix. Suppose that a sequence of elementary row-operations reduces A to the identity matrix. Then the same sequence of elementary row-operations when applied to the identity matrix yields $A^{-1}$. The intensity of the first pixel is taken. Subtract the maximum intensity value(255) from the intensity value taken from the pixel. Then repeat the first two steps for all the pixels in the image to invert the image [40].

### 4.4. Boundarydetection method

Basically pre-processing is applied to remove unwanted information and distorting as well as a ringing effect in order to get the enhanced and much clear image for the efficient use of the data set. Containing cues in an image are called boundaries that are very significant to high level pictorial tasks such as object recognition and scene understanding. Detecting boundaries have been vital problem since the commencement of computer vision. Some critical roles played in the development of boundary detection methods based on the values of datasets along with their evaluation methods. The behavior of data items is responsible for the progress in the problem of boundary detection. They provide not only an objective quantity to judge the value of each newly created algorithm, but also because of the images used in the approach. The evaluation ethics they set forth have heavily inclined the researchers during the development of a boundary detection algorithm [41].

$$R(X, Y) = A\partial^{1/\Delta f(x/y)} \tag{4}$$

Where, A is an arbitrary constant, $f(x,y) = 255 – f(x, y)$, and $f(x, y)$ is the intensity value of the input image at the coordinate points $(x, y)$. Therefore, R will be very large (resonance condition) at the boundary, where $f(x, y)$ H" 255 and will be small at the points away from the boundary. Hence, by computing the value of R and traversing the co-ordinates $(x, y)$ where $R(x, y)$ gives highest value, the boundary of the breast cancer can be identified and extracted.

$$\text{midx} = \text{image\_width}/2 \tag{5}$$

$$\text{midy} = \text{image\_height}/2 \tag{6}$$

### 4.2. Results of Preprocessing

The main objective of this procedure is to develop the quality of the image, to make it ready for further processing.The detection of the ROI consists in finding a region of the image which appears different from the background with respect to low-level features such as contrast, color, region size and shape, distribution of contours or texture pattern. Different methods have been proposed to detect regions of interest in an image [31]. The pixel having highest intensity value in the digital image is chosen, then that pixel is compared to the neighboring pixels.The comparison goes on till there is a modification in the pixel value.In pre-processing, the mammogram image outer unwanted area is removed.Figure 2is the input images; figure 3 shows the result of preprocessing50 images and table 1 shows the result of preprocessed images [23]. The number of pixels before and after preprocessing the images is given in the columns against the BP and AP respectively in table 1. The difference between BP and AP is given the column against "Difference".

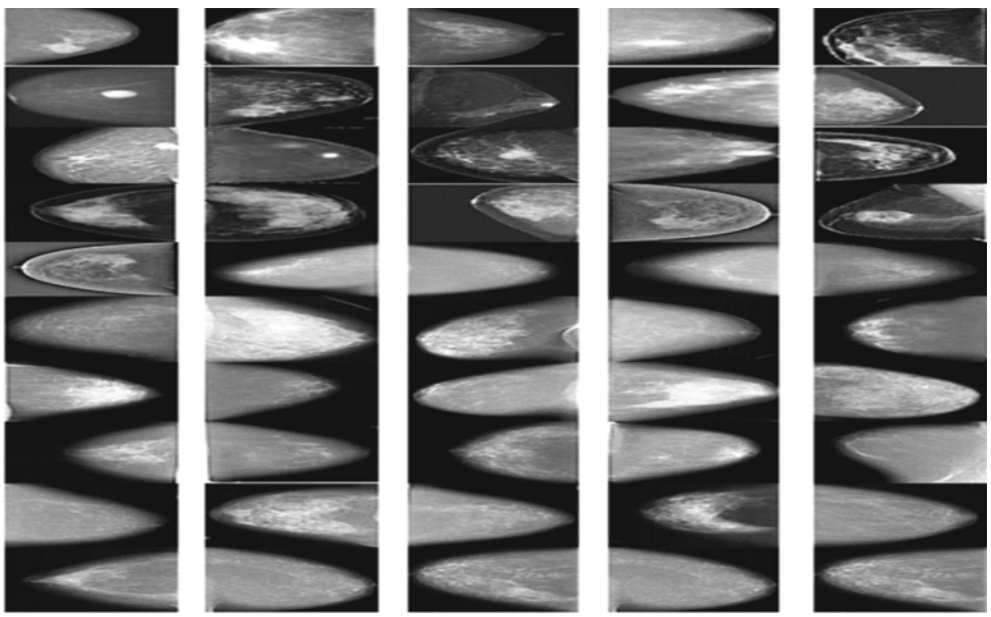

**Fig. 2. Input images.**

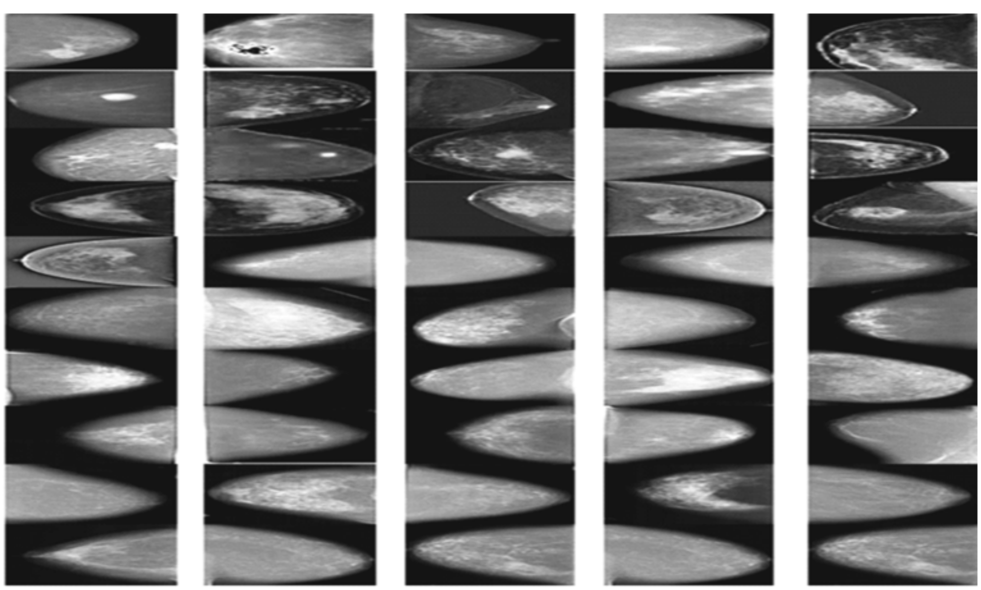

**Fig. 3. Preprocessed images.**

**Table 1. Results of Preprocessing**

| *Image No.* | *Number of Pixels* | | *Difference* | *Image No.* | *Number of Pixels* | | *Difference* |
|---|---|---|---|---|---|---|---|
| | *BP* | *AP* | | | *BP* | *AP* | |
| 1 | 23153 | 23353 | 200 | 26 | 24038 | 24338 | 300 |
| 2 | 27522 | 27822 | 300 | 27 | 23205 | 23320 | 115 |
| 3 | 25345 | 25690 | 345 | 28 | 24157 | 24756 | 601 |
| 4 | 28078 | 28180 | 102 | 29 | 24570 | 25030 | 440 |
| 5 | 27522 | 27630 | 108 | 30 | 27270 | 27370 | 100 |
| 6 | 24720 | 24920 | 200 | 31 | 25515 | 25630 | 115 |
| 7 | 21780 | 22000 | 220 | 32 | 27405 | 27800 | 395 |
| 8 | 23430 | 23730 | 200 | 33 | 26865 | 27005 | 260 |
| 9 | 22220 | 22310 | 110 | 34 | 26325 | 26625 | 300 |
| 10 | 20790 | 21001 | 111 | 35 | 30096 | 30396 | 300 |
| 11 | 23919 | 24019 | 100 | 36 | 30552 | 30752 | 200 |
| 12 | 22512 | 22813 | 301 | 37 | 31516 | 31616 | 100 |
| 13 | 24416 | 24617 | 201 | 38 | 28710 | 28820 | 110 |
| 14 | 22288 | 22388 | 100 | 39 | 27695 | 27795 | 100 |
| 15 | 20496 | 20603 | 293 | 40 | 28542 | 28742 | 200 |
| 16 | 18067 | 18165 | 102 | 41 | 28116 | 28330 | 214 |
| 17 | 17355 | 17456 | 101 | 42 | 30098 | 30315 | 383 |
| 18 | 18774 | 18874 | 100 | 43 | 30098 | 30415 | 483 |
| 19 | 16821 | 17021 | 200 | 44 | 28161 | 28261 | 100 |
| 20 | 16920 | 18020 | 100 | 45 | 33698 | 33899 | 199 |
| 21 | 25578 | 25880 | 302 | 46 | 33532 | 33832 | 300 |
| 22 | 24318 | 24520 | 202 | 47 | 34528 | 34728 | 200 |
| 23 | 27852 | 27953 | 101 | 48 | 32617 | 32800 | 217 |
| 24 | 26268 | 26670 | 302 | 49 | 32617 | 32900 | 317 |
| 25 | 24318 | 24722 | 304 | 50 | 24762 | 24962 | 200 |

## 5. IMAGE CLUSTERING

The main purpose of clustering is to divide a set of objects into significant groups. The clustering of objects is based on measuring of correspondence between the pair of objects using distance function. Thus, result of clustering is a set of clusters, where object within one cluster is further similar to each other, than to object in another cluster.

The Cluster analysis has been broadly applied in numerous applications, including segmentation of medical images, information analysis, and image processing. Clustering is also called segmentation in images for some applications [4]. The process of grouping a set of physical or abstract objects into classes of similar objects is called clustering. By clustering, one can identify dense and sparse regions and therefore, discover overall distribution patterns and interesting correlations between data attributes [22]. Thus, clustering in measurement space may be an pointer of similarity of image regions, and may be used for segmentation purposes[26].

### 5.1. The Fuzzy C-Means Algorithm

Fuzzy C-Mean (FCM) is an unsupervised clustering algorithm that has been applied to a wide range of problems calling for feature analysis; clustering and classifier design.FCM has a wide area of applications such as agricultural engineering, astronomy, chemistry, geology, image analysis, medical diagnosis, condition analysis, and target identification.With the ripening of the fuzzy theory, the fuzzy c-means clustering algorithm derived from Ruspini fuzzy clustering theory was proposed in 1980s[18].This algorithm is examined to consider based on the distance between the various input data points. The clusters are formed according to the distance between data points and cluster centers are formed for each cluster. The technique FCM is a method of clustering which allows one piece of data to belong to more than two clusters. This method is frequently applied in pattern recognition. It is based on minimization of the following objective function:

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^m \| x_i - c_j \|^2, 1 \le m < \infty \tag{7}$$

Where $m$ is any real number greater than 1, $u_{ij}$ is the degree of association of $x_i$ in the clusterj, $x_i$ is the ith of $d$-dimensional measured data, $c_j$ is the $d$-dimension center of the cluster, and ||*|| is any norm expressing the similarity between any measured data and the center. Fuzzy partitioning is carried away through an iterative optimization of the objective function shown above, with the update of membership $u_{ij}$ and the cluster centers $c_j$ by:

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left[ \frac{\| x_i - c_j \|}{\| x_i - c_k \|} \right]^{\frac{2}{m-1}}}, cj = \frac{\sum_{i=1}^{N} u_{ij}^m . x_i}{\sum_{i=1}^{N} u_{ij}^m} \tag{8}$$

This iteration will terminate when $\max y\{| u_{ij}^{(k+1)} - u_{ij}^{(k)} |\} < \xi$

Were î is a termination criterion between 0 and 1, whereas $k$ is the iteration steps. This procedure converges to a local minimum or a saddle point of Jm. Steps of the algorithm is follows:

**Step 1:** Initialize U = [$u_{ij}$] matrix, U (0)

**Step 2 :** At $k$-step: calculate the centers vectorsC ($k$) = [$cj$] with U ($k$)

**Step 3 :** Update U ($k$),U ($k$ +1)

**Step 4 :** If || U ($k$ + 1) – U ($k$) || < î then STOP;

Otherwise, return to step 2. In this algorithm, data are bound to each cluster by means of a membership function, which represents the fuzzy behavior of the algorithm [18].

### 5.2. Results of FCM

The sensing and identification of breast cancer diseases through the FCM algorithm and the segmentation of affected region are carried out in this research work.This work takes the mammogram breast image as input data.The results of FCM were carried out in the platform of IBM Windows 7 operating system.This source code of research work is written by using Matrix Laboratory MATLAB. Initially, the fuzzy C-Means algorithm is applied to the mammogram images and clustered based on the intensity. The number of clusters in every image is given as 5. Figure 4 shows theresults of a single abnormalimage. It is easy to identify that there is no specific difference in the images up to $4^{th}$ cluster. But, in the $5^{th}$ cluster, there is a change. The 5th cluster has just about disputes in the intensity and which is the most unnatural part of the original image.As per the suggestions of medical practitioners, the affected part is found in 5th cluster.Figure 5is the output of the normal mammogram image. Note that there are no white color pixels in the $5^{th}$ cluster of the image. Therefore the image is treated as normal image. Figure 6 is the output of all the 50 mammogram images. The number of white color pixels and black color pixels of abnormal image for all the five clusters are listed in table 2. Like, for normal image, it is given in the table 3. The number of pixels before and after the preprocessing for each category is given in the respective table for both abnormal and normal images. The output of the preprocessing the 50 images are given as input for the FCM algorithm. After clustering all the images by FCM algorithm, the final cluster results are shown in figure 6.

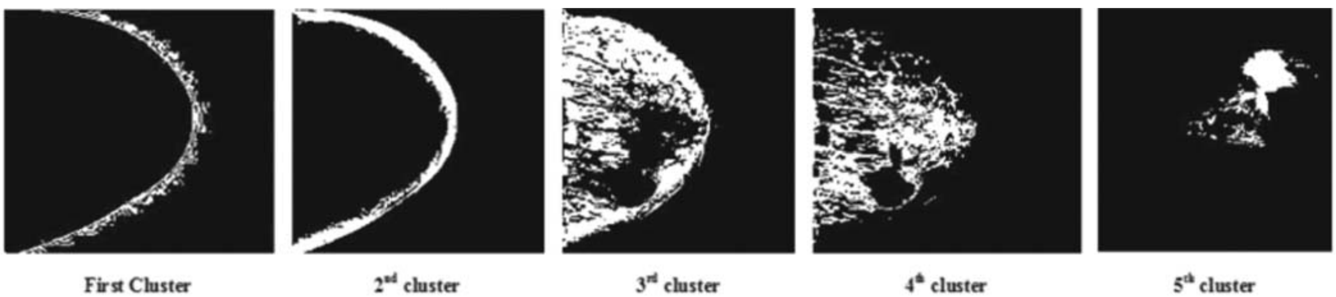


**Fig. 4. Results of Abnormal image.**

**Table 2: Results of Abnormal image**

| ***No of Cluster*** | ***C1*** | | ***C2*** | | ***C3*** | | ***C4*** | | ***C5*** | |
|---|---|---|---|---|---|---|---|---|---|---|
| Area | B | W | B | W | B | W | B | W | B | W |
| BF | 24243 | 3579 | 20649 | 2495 | 18787 | 9035 | 21346 | 6476 | 25163 | 2659 |
| AF | 15818 | 16799 | 30122 | 2495 | 25354 | 7263 | 27794 | 4823 | 31379 | 1238 |

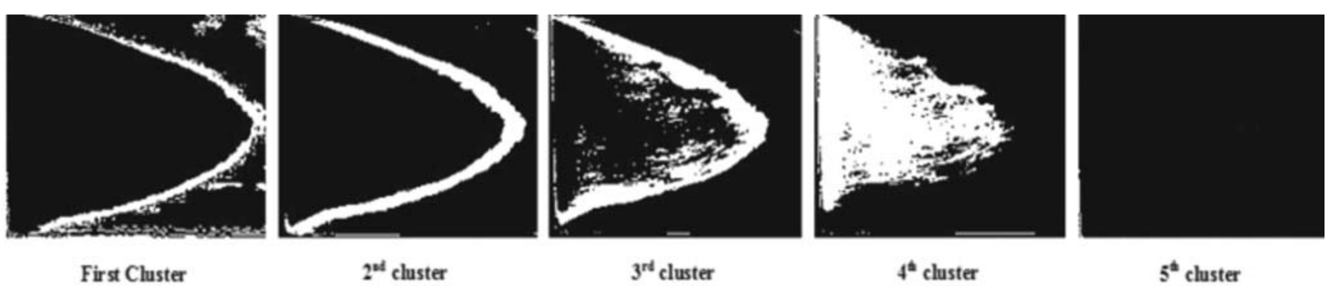


**Fig. 5. Result of normal image.**

**Table 3. Results of normal image.**

| ***No of Cluster*** | ***C1*** | | ***C2*** | | ***C3*** | | ***C4*** | | ***C5*** | |
|---|---|---|---|---|---|---|---|---|---|---|
| Area | B | W | B | W | B | W | B | W | B | W |
| BF | 19983 | 11633 | 28298 | 3318 | 26084 | 5532 | 10642 | 20974 | 31125 | 491 |
| AF | 19983 | 11633 | 28298 | 3318 | 26084 | 5532 | 10642 | 20974 | 31125 | 491 |

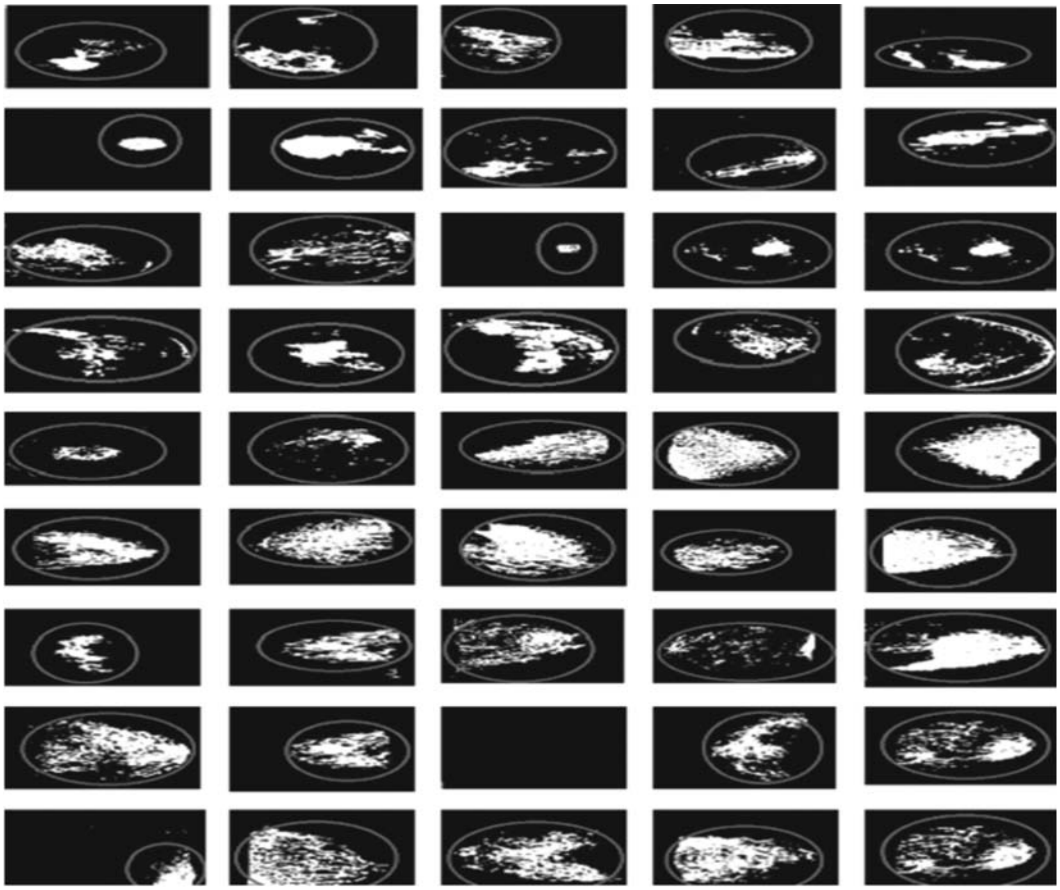

**Fig. 6. Results of Mammogram images.**

## 5.3. Discussions

The segmentation of images by the fuzzy C-means clustering algorithm is carried out in this work to find the tumor affected area. The final results of 5$^{th}$ cluster in each and every mammogram image is listed in table 4. White color pixels and black color pixels of all 50 images are given in the same table. The execution time (in milliseconds) for clustering the images are also calculated. For the verification purpose of the results, this work uses classification algorithms CART, J48, JRIP, SVM, and Naive Bayes. These algorithms are applied for finding the efficiency in terms accuracy, sensitivity and specificity. Theresults of accuracy are given in table 5. The performance of all the five algorithms is shown in figure 7based on its accuracy.The outcomes indicate that the highest accuracy 96.45% is found in JRIP classifier, second accuracy 95.75% is found in CART algorithm, third 94.61% is found in the Naive Bayes algorithm, fourth 93.40% is found in SVM algorithm and finally accuracy 91.85% is found in J48 algorithm. Among the choice of classification algorithms, the performance of JRIP is well for the selected data.

**Table 4. Results of FCM in 5$^{th}$ cluster**

| *Input Images No.* | *Cluster 5* | | *Time in milli seconds* | *Input Images No.* | *Cluster 5* | | *Time in milli seconds* |
|---|---|---|---|---|---|---|---|
| | *B* | *W* | | | *B* | *W* | |
| 1 | 31379 | 1238 | 3320 | 26 | 20842 | 3496 | 2789 |
| 2 | 25163 | 2659 | 3219 | 27 | 18114 | 5206 | 2691 |
| 3 | 23481 | 2209 | 3452 | 28 | 17094 | 7668 | 2849 |
| 4 | 23705 | 4475 | 3818 | 29 | 18625 | 6385 | 2780 |
| 5 | 25364 | 2266 | 3790 | 30 | 26078 | 1292 | 3529 |

| *Input Images No.* | *Cluster 5* | | *Time in milli seconds* | *Input Images No.* | *Cluster 5* | | *Time in milli seconds* |
|---|---|---|---|---|---|---|---|
| | *B* | *W* | | | *B* | *W* | |
| 6 | 24248 | 672 | 3538 | 31 | 22744 | 2886 | 3299 |
| 7 | 19703 | 2297 | 3189 | 32 | 23341 | 4459 | 3576 |
| 8 | 23140 | 1490 | 3366 | 33 | 24756 | 2366 | 2293 |
| 9 | 21131 | 1199 | 3153 | 34 | 20621 | 6004 | 3768 |
| 10 | 18180 | 2721 | 2505 | 35 | 23923 | 6473 | 1935 |
| 11 | 21978 | 2041 | 2577 | 36 | 28168 | 2584 | 2063 |
| 12 | 20660 | 2153 | 3236 | 37 | 31125 | 491 | 2629 |
| 13 | 24278 | 339 | 3657 | 38 | 26007 | 2813 | 3335 |
| 14 | 21542 | 846 | 3222 | 39 | 24087 | 3708 | 3252 |
| 15 | 19497 | 1292 | 3340 | 40 | 25197 | 3545 | 3023 |
| 16 | 16912 | 1257 | 3058 | 41 | 21113 | 7217 | 2692 |
| 17 | 16165 | 1291 | 3649 | 42 | 25795 | 4686 | 2737 |
| 18 | 15932 | 2947 | 2511 | 43 | 21896 | 8302 | 2773 |
| 19 | 15583 | 1438 | 2930 | 44 | 23828 | 4532 | 2944 |
| 20 | 15124 | 1896 | 2492 | 45 | 31621 | 2377 | 2809 |
| 21 | 22869 | 3011 | 3094 | 46 | 29027 | 4705 | 2917 |
| 22 | 21938 | 2582 | 3049 | 47 | 26774 | 5843 | 3276 |
| 23 | 22742 | 5211 | 3490 | 48 | 29735 | 2882 | 3179 |
| 24 | 18064 | 5061 | 2686 | 49 | 31906 | 711 | 1954 |
| 25 | 18135 | 6487 | 3360 | 50 | 30358 | 2259 | 3567 |

**Table 5. Accuracy of classification algorithm**

| *Predictive Model* | *Accuracy* | *Sensitivity* | *Specificity* |
|---|---|---|---|
| JRIP | 96.45 | 0.984 | 0.007 |
| Naïve Bayes | 94.61 | 0.928 | 0.362 |
| SVM | 93.40 | 0.93 | 0.404 |
| J48 | 91.85 | 0.928 | 0.094 |
| CART | 95.75 | 0.940 | 0.106 |

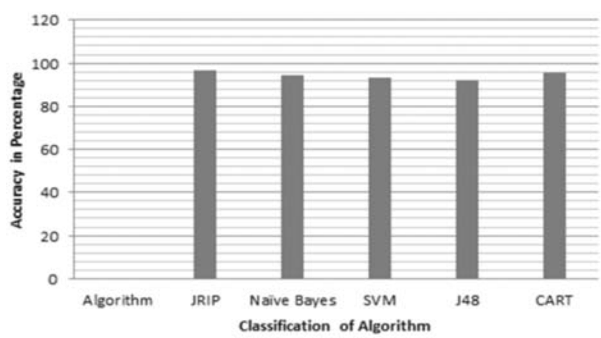


**Fig. 7. Performance of Algorithms.**

## 6. CONCLUSIONS

This research work is carried out to determine the breast cancer segmentation in mammogram images via finding its affected region. The mammogram images of normal, benign and malignant are analyzed by fuzzy C-Means algorithm which helps to get the breast cancer affected area by detecting masses in the images.Totally, 50 mammogram images are chosen for the analysis and segementedby FCM algorithm based on its intensity by taking C value as 5. Before applying FCM algorithm, the images are preprocessed by the preprocessing methods ROI, inverse method and boundary detection method. By fuzzy C-Means algorithm, the cancer region was identified in the fifth cluster by the clustering process in a perfect manner . This work also evaluates the performances of classifying accuracy by using J48, JRIP, SVM, Naive Bayes and CART algorithms by its various accuracy measures. During the implementation process, it is considered only the intensity of the final images in the prediction of breast cancer. Based on the classification results of all these algorithms, the performance of JRIP is better than the SVM, CART, Naïve Bayes and J48. Hence, identifying the affected regions of mammogram images after some preprocessing techniques, the FCM algorithm performs well and FCM is suited to detect the cancer regions effectively.

## 7. REFERENCES


1. Leena Jasmine J.S and A. Safana,"Poossibilistic Fuzzy C- means Algorithm For Mass Classification In Digital Mammogram", International Journal of Engineering Research and Applications, Vol. 4, Issue.12, pp.84-89, 2014.
2. Akila, K and P. Sumathy., "Early Breast Cancer Tumor Detection on Mammogram Images", International Journal Computer Science, Engineering and Technology, Vol. 5, Issue. 9, pp. 334-336, 2015.
3. Shiv Shakti Shrivastava, AnjaliSant and Ramesh Prasad Aharwal., "An Overview on Data Mining Approach on Breast Cancer data",International Journal of Advanced Computer Research, Vol. 3, Issue.13, pp.256-262, 2013.
4. R. Ramani, S. Valarmathy and N. SuthanthiraVanitha, "Breast Cancer Detection in Mammograms based on Clustering TechniquesA Survey", International Journal of Computer Applications, Vol. 62, No.11, pp. 17-21, 2013.
5. G. Sujatha and K. Usha Rani, "Evaluation of Decision Tree Classifiers on Tumor Datasets", International Journal of Emerging Trends & Technology in Computer Science,Vol. 2, Issue. 4,pp. 418-423, 2013.
6. Sujata Joshi and S. R. PriyankaShetty,"Performance Analysis of Different Classification Methods in Data Mining for Diabetes Dataset Using WEKA Tool",International Journal on Recent and Innovation Trends in Computing and Communication, Vol. 3, Issue. 3, pp. 1168 – 1173, 2015.
7. B. Senthilkumar and G. Umamaheswari, "Combination of Novel Enhancement Technique and Fuzzy C Means Clustering Technique in Breast Cancer Detection", Biomedical Research, Vol. 24, Issue. 2, pp. 252-256, 2013.
8. Tara Saikumar, P. Yugander, P. SreenivasaMurthy and B. Smitha, "An Adaptive Fuzzy c-Means Clustering Algorithm forBreast Image Segmentation", International Journal of Image Processing and Applications, Vol.2, Issue.2, pp. 79-83, 2011.
9. GokilaDeepa. G, "Mammogram Image Segmentation Using Fuzzy Hybrid with Particle Swarm Optimization (PSO)", International Journal of Engineering and Innovative Technology, Vol. 2, Issue. 6, pp. 167-171, 2012.
10. AnamikaAhirwar and R.S. Jadon, "Characterization of the tumor region using SOM andNeuro Fuzzy techniques in Digital Mammography", International Journal of Computer Science & Information Technology, Vol. 3, No 1, pp. 199-211, 2011.
11. ShruthiAnand, Vinod andAnandRampure, "Application of Fuzzy c-means and Neural networks to categorize tumor affected breast MR Images", International Journal of Applied Engineering Research,Vol. 10 No.64, pp. 275-281,2015.
12. S. Shanthi and V. Muralibhaskaran, "Automatic Detection and Classification ofMicrocalcification, Mass, Architectural Distortion and Bilateral Asymmetry in Digital Mammogram", International Journal of Medicine, Health, Biomedical, Bioengineering and Pharmaceutical Engineering, Vol. 8, No 11, pp. 814-819, 2014.
13. NaliniSingh, Ambarish G Mohapatra and GurukalyanKanungo., "Breast Cancer Mass Detection in Mammograms using K-means and Fuzzy C-means Clustering",International Journal of Computer Applications, Vol. 22, No.2, pp.16-21, 2011,
14. Golestani.N, M. EtehadTavakol and E. Y. K. Ng, "Level set Method for Segmentation of Infrared Breast Thermo-grams", EXCLI Journal, Vol.13, pp. 241-251,2014.

15. M. P. Sukassini and T. Velmurugan.," A Survey on the Analysis of Segmentation Techniques in Mammogram Images", Indian Journal of Science and Technology, Vol. 8, Issue 22, pp. 1-8, 2015.

16. Sandhya G, D Vasumathi and G T Raju, "Mammogram Image Segmentation Quality Enhancement Using Clustering Techniques", American Journal of Engineering Research, Vol. 4, Issue. 4, pp-153-158, 2015.

17. E. Venkatesan, T. Velmurugan, "Performance Analysis of Decision Tree Algorithms for Breast Cancer Classification" International Journal of Data Mining Techniques and Applications, Vol. 4 Issue. 2, pp. 29-44, 2015.

18. Velmurugan, T., and T.Santhanam., "Implementation Of Fuzzy C-Means Clustering Algorithm For Arbitrary Data Points", International Conference on Systemic, Cybernetics and Informatics, pp. 68-71, 2011.

19. Natarajan, P, DebsmitaGhosh,Kenkre Natasha Sandeep and SabihaJilani., "Detection of Tumor in Mammogram Images using Extended Local Minima Threshold", International Journal of Engineering & Technology, Vol.5, No 3, pp. 2310-2315, 2013.

20. CholavendhanSelvaraj, Siva Kumar R andKarnanM.,"A Research on Various Filtering Techniques in Enhancing Mammogram Image Segmentation", International Journal of Engineering Trends and Technology, Vol. 9, No 9, pp. 451-453, 2014.

21. D.NarainPonraj,M.Evangelin Jenifer, P. Poongodi and J. SamuelManoharan.,"A Survey on the Preprocessing Techniques of Mammogram for the Detection of Breast Cancer", Journal of Emerging Trends in Computing and Information Sciences, Vol. 2, NO. 12, pp. 656-664, 2011.

22. AmiyaHalder., SoumajitPramanik. And ArindamKar., " Dynamic Image Segmentation using Fuzzy C-Means based Genetic Algorithm", International Journal of Computer Applications, Vol.28,No.6, pp. 15-20,2011.

23. IndraKantaMaitra., Sanjay Nag and Samir Kumar Bandyopadhyay.,"Identification of Abnormal Masses in Digital Mammography Images", International Journal of Computer Graphics, Vol. 2, No. 1, pp. 17-30, 2011.

24. Monica Jenefer and V. Cyrilraj.,"An Efficient Image Processing Methods for Mammogram, Breast Cancer Detection", Journal of Theoretical and Applied Information Technology, Vol.69, No. 1, pp. 32-39, 2014.

25. K. Vaidehi and T. S. Subashini.,"A global approach for detecting mass in Digital Mammograms", International Journal of Advancements in Research & Technology, Vol.1, Issue. 6, pp. 1-4, 2012.

26. Ms. Siva Priya and E. Ashok Ku., "HMPFIM-B: Hybrid Markov Penalized FCM in Mammograms for Breast Cancer", International Journal on Recent and Innovation Trends in Computing and Communication, Vol. 2, Issue.10, pp. 3033-3037, 2014.

27. D.A.Kulkarni, Bhagyashree.S.M and G.R.Udupi., "Texture Analysis of Mammographic images", International Journal of Computer Applications,Vol. 5, No.6, 2010.

28. ShenghuaGu, Yao Ji,Yunjie Chen, Jin Wang and Jeong-Uk Kim., "Study on Breast Mass Segmentation in Mammograms",3rd International Conference on Computer, Information and Application, pp.22-25, 2015.

29. Benjamin joseph A,Baskaran Ramachandran and PriyadharshiniMuthukrishnan., "Intelligent Detection and Classification of micro-calcification in Compressed Mammogram Image",Journal Abbreviation image anal stereol,Vol.34, No.3.pp.183-198, 2015.

30. Kashyap, KanchanLata, Manish Kumar Bajpai, and Pritee Khanna., "Breast cancer detection in digital mammograms", Imaging Systems and Techniques, IEEE International Conference on. IEEE, 2015.

31. Carona.Y., P. Makrisa and N. Vincent., "Use of power law models in detecting region of interest",Jounal of Pattern Recognition Society Elsevier, Vol.40, No.9, pp.2521 – 2529,2007.

32. Craddock, R. Cameron, G. Andrew James, Paul E. Holtzheimer, Xiaoping P. Hu, and Helen S. Mayberg, "A whole brain FMRI atlas generated via spatially constrained spectral clustering", Human brain mapping, Vol. 33, Issue. 8, pp. 1914-1928,2012.

33. Carson, R. E., Y. Yan, M. E. Daube-Witherspoon, N. Freedman, S. L. Bacharach, and P. Herscovitch., "An approximation formula for the variance of PET region-of-interest values", Medical Imaging, IEEE Transactions, Vol.12, No. 2,pp.240-250, 1993.

34. Amygdalos, Iakovos. "Detection and classification of gastrointestinal cancer and other pathologies through quantitative analysis of optical coherence tomography data and goniophotometry." PhD diss., Department of Medicine, Imperial College London, 2014.

35. Young, Carolyn Mei..,"High-speed volumetric in vivo medical imaging for morphometric analysis of the human optic nerve head.", M,SC., diss., Applied Science: School of Engineering Science andsimonfraser university, 2011.

36. Tong, Yunxia, Qiang Chen, Thomas E. Nichols, Roberta Rasetti, Joseph H. Callicott, Karen F. Berman, Daniel R. Weinberger, and Venkata S. Mattay., “Seeking Optimal Region-Of-Interest (ROI) Single-Value Summary Measures for fMRI Studies in Imaging Genetics”, PloSone,Vol. 11, Issue. 3, pp.1-20,2016.

37. Salti, Samuele, AliosciaPetrelli, Federico Tombari, Nicola Fioraio, and Luigi Di Stefano., “Traffic sign detection via interest region extraction”, Elsevier Pattern Recognition, Vol. 48, Issue. 4, 2015, pp. 1039-1049.

38. Kambo, Rubi, and Amit Yerpude.,. “Classification of basmati rice grain variety using image processing and principal component analysis”,International Journal of Computer Trends and Technology, Vol. 11. Issue. 2, pp. 80-85,2014.

39. Deepa S. and SubbiahBharathi V., “ Efficient ROI Segmentation of Digital Mammogram Images using Otsu’s N thres holding method”, International Journal of Engineering Research & Technology, Vol. 2, Issue 1, pp.1-6, 2013.

40. Y.Ireaneus Anna Rejani and S.ThamaraiSelvi., “Breast Cancer Detection Using Multilevel Thresholding”,International Journal of Computer Science and Information Security, Vol.6, No.1, pp.111-115, 2009.

41. Naveen. A and T. Velmurugan.,”Identification of Calcification in MRI Brain Images by k-Means Algorithm”,Indian Journal of Science and Technology, Vol. 8, No.29, pp. 1-7, 2015.