Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 4, Issue. 5, May 2015, pg.127 – 134

RESEARCH ARTICLE

Stage Determination of Cancer in Mammogram Image using SOFT CLUSTERING and ANN

Supriya Kulkarni[1], Shreedhara K S[2]

[1]P G Scholar, Department of Computer Science, UBDTCE (VTU), India

[1]Professor, Department of Computer Science, UBDTCE (VTU), India

supkul47@gmail.com

Abstract: Women around the world are prone to a severe cancer called the breast cancer. This is a disease that arises due to the growth of breast tissue cells that are not normal. The detection of cancer is much important due to the survival of the woman. This research determines to develop a software that can determine the stage of breast cancer based on the size of the cancer. Firstly the sample cancer images are collected and then this images are cropped to find the region of interest. The cancer region is suspected using fuzzy-c means and stage is found out by ANN. (there were different mammogram sample images taken from the mini mammography database of MIAS, the proposed method can detect stage of breast cancer.

Keywords—classification; staging; breast cancer; mammogram; fuzzy-c means clustering and ANN

I. INTRODUCTION

Cancer is a group of diseases that cause cells in the body to change and grow out of control. Most types of cancer cells eventually form a lump or masses called a tumor, and are named after the part of the body where the tumor originates. Breast cancer begins in breast tissue, which is made up of glands for milk production, called lobules, and the ducts that connect lobules to the nipple.Breast is made up of fatty, connective, and lymphatic tissue [1].

Estimated new female breast cancer rates and deaths in 2013-14 survey

Age in years	In situ case	Invasive cases	Deaths
<40	1900	10980	1020
<50	15650	48910	4780
50-64	26770	84210	11970
65+	22220	99220	22870

Non-hispanic white have high incidence rates in breast cancer in an US study 2006-2010 Appoximately 39.620 women are expected to die from breast cancer successful treatment and complete recovery of the patient can be done by detection and diagnosis of breast cancer in its early stage. Screening .mammography technique is an x-ray examination of the breasts in a woman . Mammography detects more than 75% of cancers . Hence mammography technique is useful.

This breast cancer is a silent killer. Many women who are not aware of new diseases ie; people who stay in rural areas cannot identify that it's a breast cancer even if its in final stages. This breast cancer comes silently with no major symptoms. women must undergo a mammography test once in 6 months to determine this cancer in early stages .Like all other cancers this breast cancer also has few different stages. There are nearly 6 different stages.so for every breast cancer we have to first obtain a mammopgraphy image then the preprocessing of the image is done, feature extraction, stage detection. Breast cancer stage is used to describe the condition of cancer, namely its location, its size, where it spreads and the extent of its influence on other organs .
In general, the level of breast cancer stage is stage I , II , III ,IV,V&VI[7]. In fact, determining the level of breast cancer stage is not easy. Many factors in science will influence the stage of the cancer.

This paper aims to propose a method to determine the stage of breast cancer malignancy based on area of mammogram image along with other factors. This work is organized as follows. In Section 2, literature review that related work are presented. In Section 3, we present the proposed method includes the process of segmentation and classification. Next, in Section 4, the results are shown. Finally, Section 5 presents some concluding remarks.

II. LITERATURE REVIEW

Breast cancer detection can be carried out by using a variety of techniques. For successful treatment of the patient the breast cancer has to detected in its early stage and thus the patient can be recovered quickly. For breast cancer detection, the mammogram images will be collected in the first stage, after the image acquisition stage preprocessing will be performed. Next stage will be the image enhancement in which in the resultant image the finer details will be more clearer than the original image since its filtered and noise is removed, the image will be segmented to extract the microcalcification part or cancer detected area.

Various technique used in breast cancer detection is described below:

** Segmentation Technique*
Segmentation is the process of partitioning a digital image into multiple segments. By segmentation technique it is easy to change the representation of an image so it will be easier to analyze and it is easy to locate objects and boundaries in images. In this technique image can be segmented and the set of segments will cover the entire image. Segmentation can be carried out using any of the standard techniques like Local Thresholding, Fuzzy-c means Clustering. bilevel color is used. Bilevel image is a color image which is divided into two colors, 0 (black) and 1 (white). Simplification of color using thresholding is widely used for pattern recognition by eliminating color complexity into simple color so that an observed image has a color pattern which characteristics are easily grouped.

** Region Growing:*
Region Growing is a procedure that classifies the pixels or sub-regions into larger regions based on predefined criteria. The approach basically starts from the beginning of the set of points, then the area is enlarged by adding each neighboring pixel point that has properties similar to those points (for example the range of intensity or color specification).
The selection of similar criteria, in addition to depending on the problem at hand, also depends on the type of image data available, for example descriptor. Examples of descriptors include moment and texture. Region-growing segmentation provides the clear edges of the images. [11]. Region growing segmentation can be implemented to breast cancer detection [12].

** Fuzzyc-meansclustering:*
Data clustering is the process of dividing data elements into classes or clusters so that items in the same class are as similar as possible, and items in different classes are as dissimilar as possible. Depending on the nature of the data and the purpose for which clustering is being used, different measures of similarity may be used to place items into classes, where the similarity measure controls how the clusters are formed. Some examples of measures that can be used as in clustering include distance, connectivity, and intensity.

In fuzzy clustering (also referred to as soft clustering), data elements can belong to more than one cluster, and associated with each element is a set of membership levels. These indicate the strength of the association between that data element and a particular cluster. Fuzzy clustering is a process of assigning these membership levels, and then using them to assign data elements to one or more clusters.

One of the most widely used fuzzy clustering algorithms is the Fuzzy C-Means (FCM) Algorithm. The FCM algorithm attempts to partition a finite collection of $n$elements $X = \{\mathbf{x}_1, ..., \mathbf{x}_n\}$ into a collection of c fuzzy clusters with respect to some given criterion. Given a finite set of data, the algorithm returns a list of c cluster centers $C = \{\mathbf{c}_1, ..., \mathbf{c}_c\}$ and a partition matrix $W = w_{i,j} \in [0,1],\ i = 1, ..., n,\ j = 1, ..., c$, where each element w_{ij} tells the degree to which element $\mathbf{x}_i$ belongs to cluster $\mathbf{c}_j$. Like the K-means clustering, the FCM aims to minimize an objective function:

$$\arg\min_{C} \sum_{i=1}^{n} \sum_{j=1}^{c} w_{ij}^{m} \left\| \mathbf{x}_i - \mathbf{c}_j \right\|^2,$$

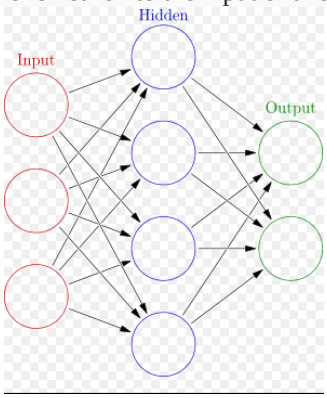
where:

$$w_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{\|\mathbf{x}_i - \mathbf{c}_j\|}{\|\mathbf{x}_i - \mathbf{c}_k\|} \right)^{\frac{2}{m-1}}}.$$

**Artificial Neural Networks:*

They are a family of statistical learning algorithms inspired by biological neural networks are used to estimate or approximate functions that can depend on a large number of inputs and are generally unknown. Artificial neural networks are generally presented as systems of interconnected "neurons" which can compute values from inputs, and are capable of machine learning as well as pattern recognition

The below figure shows an artificial neural network is an interconnected group of nodes. Each circular node represents an artificial neuron and an arrow represents a connection from the output of one neuron to the input of another.



The word *network* in the term 'artificial neural network' refers to the inter–connections between the neurons in the different layers of each system. An example system has three layers. The first layer has input neurons which send data via synapses to the second layer of neurons, and then via more synapses to the third layer of output neurons. More complex systems will have more layers of neurons with some having increased layers of input neurons and output neurons

This has applications in real life applications like

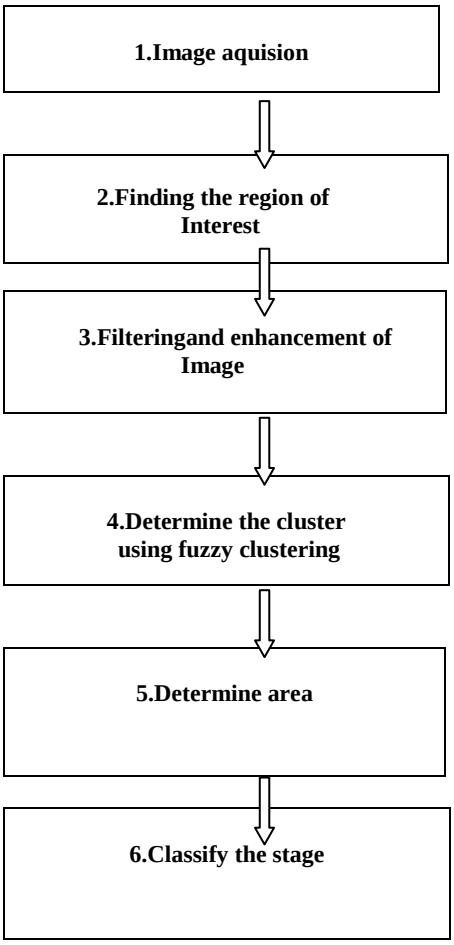
*classification

*data processing

*regression analysis

III. PROPOSED METHOD

The method proposed in this paper is to classify the stages of malignancy of breast cancer based on the mammogram image, through segmentation by sample fuzzy clustering method.
Stages of the proposed method are outlined



The initial step is the image acquisition to get the data in the form of mammogram digital images that are required in the research. Mammography Image Analysis Society (MIAS) database used in this research . Data is in the form of image format PGM (Portable GrayMap). PGM format is used by many medical image as a PGM is a lossless type image format where at the time of data compression, no parts are removed so that the details of the image will remain intact and not lost. Each mammogram image has a resolution of 1024x1024 pixels and the average file size of 1MB. MIAS database have been grouped into three categories, namely: (1) Dense-Glandular is the mammogram image of breasts that are dense and have many glands by nature (2) Fatty is the mammogram image of breasts that are not dense by nature because they contain a lot of fat, and (3) Fatty -glandular is a mammogram image of breasts that are not dense because they contain a lot of fat and have many glands. These PGM images are converted to jpg images and used.
Each of the three categories is further grouped into three sections, namely: (1) Normal, is a collection of normal mammogram images that are not affected by breast cancer, (2) Benign, is a collection of abnormal mammogram images that have been affected by benign breast cancer on breast tissue, and (3) malignant, is a collection of abnormal mammogram images that have invasive breast cancer.

Stages of the proposed method is explained below

1.Image aquision:
One of the images from MIAS is taken as the input sample and is shown in the figure 1

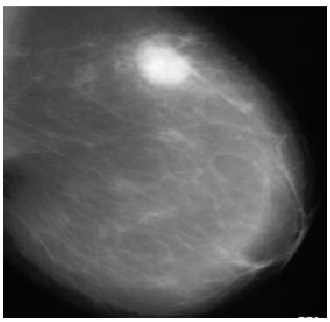


Fig 1 Input image

2.Finding the region of interest:

Later the region of interest is found out Determining ROI is a very important step if there is a certain part of the digital image that is more important than others. The first part in determining ROI is the are cropping process. This step aims to reduce the size of the image to be processed so that the result will be more accurate. Figure 2 is the result of cropping.

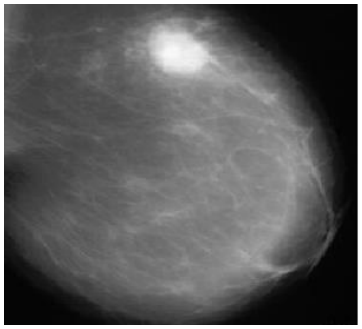


Fig 2 Image after cropping

3.filtering and enhancement of the image:

Later the image is filtered and then the enhancement of the image is done this helps in removal of noise from the image.figure 3 and figure 4 shows it.

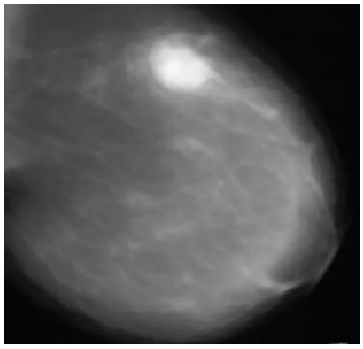


Fig 3Filtered image

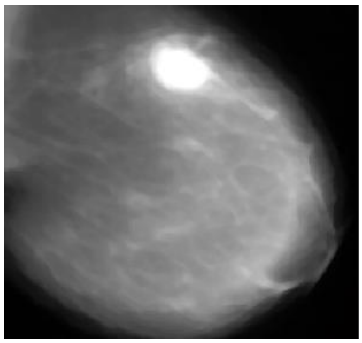


Fig 4 Enhanced image

4.Determining the cluster using the fuzzy c means:

Fuzzy clustering technique is applied and the tumor region is determined.the white color of the image shows the tumor region.figure 5 shows it. separate an area that is estimated to show abnormalities. The estimated area is the one that has a stronger / brighter white color intensity than its surrounding, has a nearly uniform density, has a regular shape with various sizes, and its boundaries are blurred

Figure 5.Tumor region image.

5.determining area:

Now the area of the tumor is determined based on the number of pixels present in tumor region.

All the formalities are completed and now ANN is used to classify the stage of the cancer. whether the image is suspect of breast cancer stage I, II. or III . Stage IV is not used as in the basis of the theory used, patients in stage IV have cancers that vary in size and has spread to several parts of the body so that further examination is needed.

IV. RESULT

The measures are grouped into 3 major groups using fuzzy c means clustering method. The clustering results reflect the 3 group stage, i.e. stage I, II, and III. Each group size has a lower limit and upper limit that are used as a reference to determine the stage of cancer in which the application is made.

Stage 1 has size of the area between 0001to 2800 pixel.

Stage 2 has size of the area between 2800 to 3100pixel.

Stage 3 has size of the area between >3100 pixel.

Total of different stage cancerous images were tested and the results is as below

Stage	Test image	Correctly classified image	Incorrectly classified image	accuracy
I	12	10	2	83.3%
II	12	9	3	75%
III	5	4	1	80%

The image with different known stages were tested with fuzzy c means and ANN technique where energy, entropy and most important that is area is used the image after running the code is

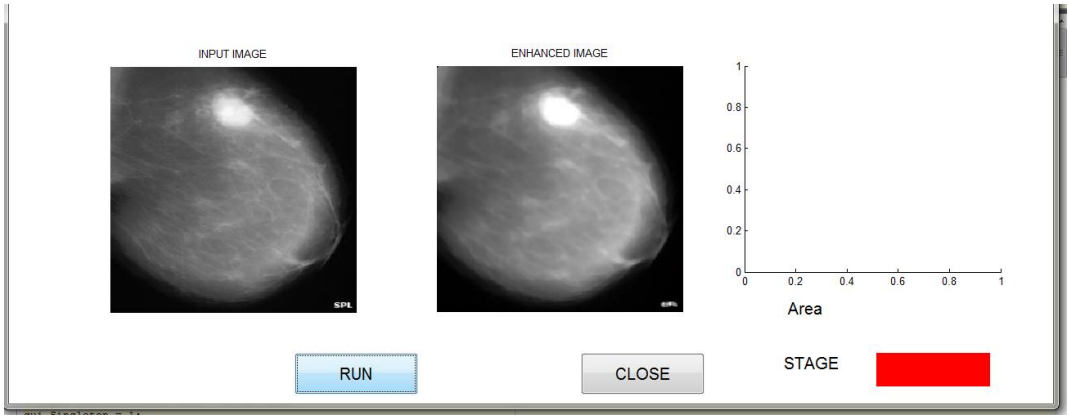


Image after the enhanced image is obtained

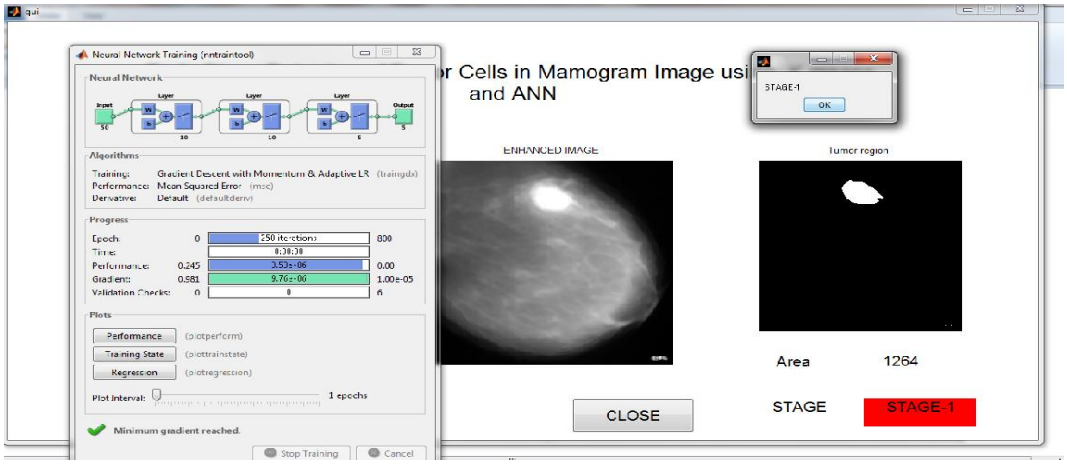


Image where the stage is detected along with the area of the pixels

V. CONCLUSION

This paper determines the different stages of cancer in mammogram images using soft clustering and ANN. Based on the tests the different stages of the cancer from the images of MIAS are determined. This is tested with different samples and the results based on the area, energy and entropy were obtained. The future work can be done with the threshold and pattern of the tumor obtained.

REFERENCES

[1] American Cancer Society, "Breast Cancer Facts & Fig. s, 2013-2014", American Cancer Society, Inc, 2013.
[2] Priya, D.S., and Sarojini, B., "Breast Cancer Detection In Mammogram Images Using Region-Growing And Contour Based Segmentation Techniques", International Journal
[3] Kamdi, S.," Image Segmentation and Region Growing Algorithm", International Journal of Computer Technology and Electronics Engineering (IJCTEE), Vol 2 Issue no.1, October 2011.
[4] Pradeep, N., Girisha, H., Sreepathi, B., and Karibasappa, K., "Feature Extraction of Mammograms", International Journal of Bioinformatics Research, Volume 4, Issue 1, 2012, pp.-241-244.
[5] American Joint Committee on Cancer, "Breast Cancer Staging", AJCC 7th Edition Staging Posters, 2009.

