ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 12, December 2016

Automatic Region Growing Algorithm based on Combined Filtering Technique to Detect Breast Cancer

M P Gopinath[1], Dr. S Prabu[2], Srinivas Koppu[3], Dr. Madhusudhana Rao[4], J Kamalakannan[5]

Assistant Professor, SCOPE, VIT University, Vellore, Tamil Nadu, India [1]

Associate Professor, SCOPE, VIT University, Vellore, Tamil Nadu, India [2]

Assistant Professor, SITE, VIT University, Vellore, Tamil Nadu, India [3]

Lecturer, Department of IT, Higher college of Technology, Muscat, Oman [4]

Assistant Professor (SG), SITE, VIT University, Vellore, Tamil Nadu, India [5]

ABSTRACT: Day by day the breast cancer becoming the most threating cancer among women especially. But at the same time it has been observed that breast cancer also affecting the males also nowadays. Worldwide the doctors are researching to get a prevention of this disease, as no cancer yet has no medicine to cure. But the most important phase in medical science is to identify the cancer cases in right time. A huge number of people are dying each year in breast cancer only because they were unable to be identified in correct time. As we all know the first stage cancer is curable for most of the time, so the number of cancer survivors can be hiked up if it is possible to identify in the first stage. A lot of existing process is already in market but still there is no such effective one. Most of the segmentation techniques follows an initial manual simulation which is time consuming as well as erroneous in some cases. In segmentation techniques one of the well-known process is the region growing algorithm which is basically nowadays scientists are preferring to use to detect most of the cancer of any organ. In this paper we have forwarded one step more towards the perfection using the automatic region growing technique. In most of the cases we need to choose the seed point manually in region growing algorithm. The newly developed region growing procedure selects the seed point adaptively according to the image intensity values and according that seed point it determines whether the breast is cancer affected or not.

KEYWORDS: Automatic region growing, Fuzzy C means Clustering, K means Clustering, Combined Filtering

I.INTRODUCTION

Breast Cancer is the earliest detected cancer in human history. The oldest recorded case history of cancer was around 1600BC in Egypt [1]. In the last 3000 years still it is the mostly spread one among the humans. The “Cancer Research group UK”[2] represented one survey on the number of deaths per year in Europe and the report was itself a thread to the humanity , as it was the only recorded cases in Europe but still number of uncounted cases are there in Africa and Asia. Breast cancer is the third most basic reason for cancer demise in the UK (2012), representing 7% of all deaths from cancer, notwithstanding the way that it is uncommon in men. It is the second most normal reason for cancer passing among ladies in the UK (2012), representing 15% of female deaths from cancer. Breast cancer was the most widely recognized reason for death from cancer in ladies until 1998; from that point forward there have been more deaths from breast cancer in ladies. In 2012, there were 11,716 deaths from breast cancer in the UK 11,643 (99%) in ladies and 73 (under 1%) in men, giving a female: male proportion of around 160:1.The rough death rate demonstrates that there are around 36 breast cancer deaths for each 100,000 females in the UK and under 1 for each 100,000 guys. The European age-institutionalized death rates (AS rates) don't vary essentially between the constituent nations of the UK for guys or females. In India Breast cancer is the second most cancer in terms of number of deaths per year. According to the “Pink India Statistics” every year number of newly detected cases in India is around 1, 45,000 and number of deaths is about 70,000 [3].

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 12, December 2016

Breast cancer is discovered typically in milk conveying an area i.e. ducts, tubes and the milk producing areas i.e. lobules, glands that produce milk (Fig 1). For ladies this infection is exceptionally regular today yet infrequently it likewise can influence the men. It is likewise recognized sometimes that the development rate of a tumour is corresponding to its temperature. It is constantly exceptionally accommodating for the patients to discover the cancer as prior as could reasonably be expected, on the grounds that the mending chances are more noteworthy for those situation.

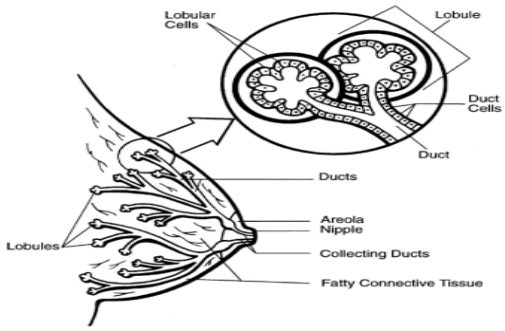


Fig 1 Structural diagram of Breast Tissues

Data got from the breast ultrasound; dedicated breast CT, tomo synthesis and mammography are morphological data. The infrared picture contains data identified with physiological of ordinariness and irregularity of the vascular framework, sensorial and sympathetic sensory system, and inflammatory methods. In Modern days the consolidated demonstrative courses of action are additionally used to get the more prominent probability to recognize the threatening cells. Ionizing radiation is not utilized as a part of Infrared pictures. Additionally the imaging methodology is effortless and there is no compelling reason to contact with the skin surface, so persistent likewise ought not to have any vexation. This strategy is low cost consuming compared to others, for example, mammography, ultrasound and attractive reverberation. This methodology is likewise extremely gainful for youthful patients as they have thick sort bosom tissues and utilizing X-Ray it is exceptionally hard to perceive early stage disease for thick district. The micro calcifications and masses are legitimately unmistakable just in mammograms of ladies who fit in with non-productive ages [4]. But it also has the limitation that it is not a reliable one for those cases where patient is in young age or the cases of dense breast.[5]

II.LITERATURE SURVEY

Dr. G.R. Sinha and Bhagwati Charan Patel in their research work, it is said that the mammogram picture is apportioned into a few non-crossing areas and concentrates the region of intersect and suspicious mass area. Utilizing "K-means clustering", the beginning segmentations are not all that simple to recognize. The collection of information is called as class and the connection of histogram is known as bin. As the state of histogram is reliant on bin, so the essential data may get lost with the addition of measure of bin. So by diminishing the quantity of bins and expanding the quantity of classes in the meantime may issue us a decent result.[6]

Some scientist also got that it is exceptionally powerful to improve the nature of the picture by lessening noise and expanding difference before segmenting the picture. So they used "Adaptive median filter", "Mean Filter", "Markov random field" technique, "Median Filter" are portrayed for filtering and "Region Growing", "Random walk strategy", "Watershed Method" , "Adaptive mean shifting", "Segmentation by Morphological algorithm" are portrays for segmentation.[7].

Scott Doyle,Shannon Agner, ,Shivang Naik, Anant Madabhushi, Michael Feldman and John Tomaszewski proposed that Nuclei Segmentation is mix of essential "Fuzzy C-means clustering" and "Throsholding". The picture is integrated by distinctive scales a)based on the value of pixels b)based on relationship between pixels for object discovery c) in view of connections between histological structures. To produce probability structures of interest a Bayesian classifier

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 12, December 2016

is utilized taking into account picture intensity and textural data and for organ limit discovery and nuclear segmentation, Level set algorithm and templet matching algorithm are utilized individually.[8]

Rajalakshmi Thyagarajan and Dr. S. Murugavalli proposed that the discovery will be all the more simple in the event that we begin work with 3D mammogram picture rather than an organizer picture. In this paper Adaptive Histogram equalization is utilized for pre-processing and afterward it is utilized the "Fuzzy C-means" algorithm. In an another area in the same paper it is said that Expectation Maximization is utilized amplify and upgrading the influenced area.[9]

S. Valarmathy, R. Ramani and N. Suthanthira Vanitha defined that among "K-means clustering", "Adaptive K-means Clustering", "Fuzzy C-means Clustering", "Kernelized Fuzzy C-means Clustering" and "Multiple Kernelized Fuzzy C-means Clustering" ,the barring the first every strategy is exceptionally viable in view of diverse predefined conditions.[10]

Raman et al.[11] used the "region growing algorithm" based on a seed pixel where the next seeds will be chosen when they are greater than or equal to the predefined threshold of their mother seed. Here they followed the 4-connected points to have the next seeds. This region growing is used for mass segmentation.

Shymaa A. Hassan et al.[12] followed a similar process like above. But before choosing the seed point in their pre-processing phase they eliminate the pectoral muscles. Then they choose the seed point that pixel which has the maximum intensity value. The reason is obvious that the malignant cell would have maximum intensity value. We also have followed this approach for our region growing proposed algorithm to detect breast cancer.

III.SYTEM ASSUMPTION AND DEFINITION

Still no successful approach to keep the breast cancer is found. Subsequently, early identification is the first critical step. These days X-ray mammography is the most widely recognized approach to recognize a breast cancer. Typically in mammographic image in starting stage it is by all accounts a white spot just. In this manner an accomplished specialist or radiologist additionally can miss that starting stage side effect if he is watching it outwardly. So there are various digital supported systems where the indications are broke down utilizing image segmentation system. Fitting image segmentation procedure to recognize a cancer in introductory state is extremely difficult.[7]

According to the tissue structure and its density type of a breast region, BIRADS(American College of Radiology Breast Imaging Reporting and Data System) classified the structure of the breast into four sub categories:

Level 1 - When the region is made of less than 25% glandular tissue, then the rest of the part is full of fatty tissue.

Level 2 - When the region is made of approximately 25%-50% of glandular tissue and the fibro-glandular tissues are scattered over the region uniformly.

Level 3 - When the fibro-glandular tissue is scattered in an un-uniform manner and the percentage is hiked to 51%-75%.

Level 4 - The extremely densed region where the fibro-glandular tissue is near about more than 75%.[13]

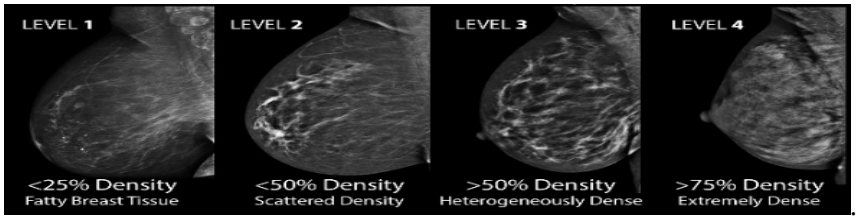


Fig 2 Different type of Density Level

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 12, December 2016

Fatty tissue is appear to be dark and fibro-glandular tissue is appear to be white in mammogram. The difficulty is to detect a small malignant cell on a dense breast region on the mammographic screen .So for the disentanglement of our work we are considering the last two category (Heterogeneously Dense and Extremely Dense) as a one. The principal indication of Cancer is to notice some white spots in mammogram. In those situations where the breasts are made of essentially fatty tissue, it is anything but difficult to distinguish a white spot in light of the fact that the majority of the district in the mammogram is dark. In dense breasts, the foundation is for the most part white, so it is exceptionally hard to identify a cancer cell in preparatory stage. The test of recognizing a little cancer in a thick breast.

IV.PROPOSED SYSTEM MODEL

Segmentation is the system to partition a digital image in such a way, to the point that the fragments can shape the entire image when they are joined together. The segmentation of an image is fundamentally done by pixel esteem. Segmentation algorithm can be grouped into different sort in view of the segmentation methods like-feature thresholding, clustering, region growing. Every strategies has their own points of interest and disadvantages.

Here we have chosen the region growing technique to proceed our research. Before come into our main algorithm we found that one more pre-processing can give us more refined result. We have eliminated the spectral muscle areas before the pre-processing phase, so that the final algorithm could not consider the muscle area as an affected area.In the next phase we have chosen four filtering techniques- Average Filter, Gaussian Filter, Median Filter and Adaptive Filter. Those resulted filtered images are combined into one and we would do further processing one that. The reason behind this type of pre-processing is that we are normalizing the other pixel points, whereas enhancing the pixel with high intensity value.

According to the signal processing while we are filtering an image, it would smooth the image more by normalizing the neighbourhood pixels. This type of normalization may differ the pixel intensity value, but highest intensity valued pixel will be relatively highest compared to the other pixels after filtering also. Now after such type of four filtering the pixel intensity values of them in the same position would be different, but for the all four cases the highest intensity valued pixel will be relatively highest in the respective image. So at the end when we combine then the highest intensity valued pixel will be more enhanced, whereas intensity value of the other pixels will be nearly as similar to the original one.

Now we have worked with this resultant image. We are considering the combined image for region growing algorithm. In this part there was two big challenges-one which pixel should be taken as a seed point and other one was how to terminate the region from false growing. For choosing a seed point we started comparing all the intensity values of pixels and the pixel which had the maximum intensity value is considered as the suspected lesion. Considering that pixel as a focus point the region would be spread outwards till the termination condition is not force it to stop. The termination condition has been chosen in heuristic way. After reading a lot of research papers and based on some different parameters we found some heuristic values to terminate the region to grow.Finally one value is chosen after several refinement based on the final outcome images. In this proposed algorithm only this part is based on several heuristic parameters, so we can suggest that this algorithm can be more précised if the termination condition can be taken from a specific mathematical viewpoint.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 12, December 2016

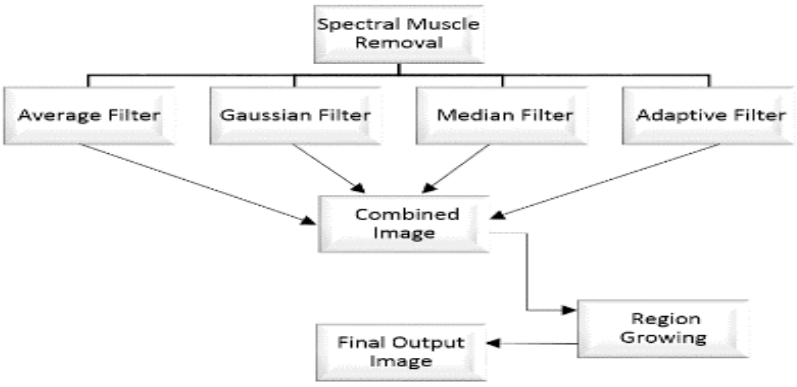
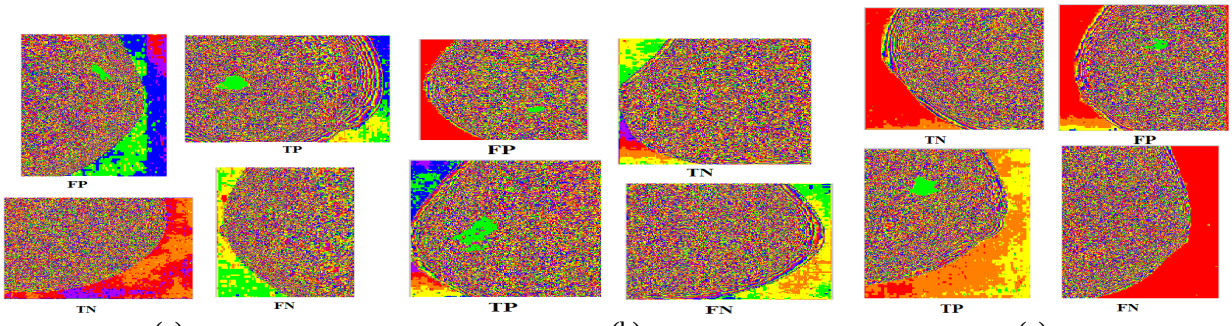


Fig 3 Flow diagram of proposed algorithm

V. RESULT AND DISCUSSION

Here we categorizes the results into four different classes. The categories are-(i) FP (False Positive): Here the case was originally normal but our proposed algorithm identified it as a malignant one. This type of case can lead to a wrong treatment which can affects the health of an individual. So we have tried to avoid this condition as much as possible. (ii)TN (True Negative): Here the case is normal and our algorithm has been identified it successfully. That means it is correctly rejected the case for being that one considered as a malignant one. (iii) FN (False Negative): The image is a case of malignancy, still the proposed algorithm could not identify it as malignant one. This is the thread for medical science as the doctors would not continue further treatment for that patient whereas she needs it. So we also took care about to reduce such cases. (iv)TP (True Positive): Here the image is a case of malignancy and it is identified by the proposed algorithm.

In table 1 we have discussed this part in an understandable and details manner of identification.



(a) (b) (c)

Fig 4 (a) Fatty Tissue (b) Glandular Tissue (c) Dense Tissue

Figure 4 describes the various tissue of breast cancer and now for fatty tissue cases it was very easy to identify the malignancy cases but for normal cases it was also showing positive results sometimes.

For Fatty-glandular or glandular tissue cases it was showing the results in a satisfactory manner for both normal and malignancy cases. For this glandular tissues, the tissues are spread over the region in an uniform or non-uniform manner, so we found that for 98% of cases used to be normal when the region had been grown over a large area of tissues. In that cases we also consider it as a normal case, not malignant case. Specially for this part we have also designed a back door method of understanding. Though it is also true that this method is not an efficient option in such sensitive domains. During the region growing algorithm we have counted the total no of pixels in the region and store it

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 12, December 2016

into a variable called 'count'. Then we have calculated the percentage of pixels in 'count' for that image. Now the percentage is calculated by the following formula:

$$P_{matched} = \frac{count}{number\ of\ valid\ pixels}$$

Here 'P_{mtched}'*means* the percentage of pixels chosen for the region and '*number of valid pixels*' means the pixels relate to the tissue parts, not the background. Initially we already cut out the spectral macules, so remaining parts are only the black background which is represented with zero in grey scale images. So whenever the value of 'P_{mtched}' goes more than one or little bit more than one, we used to consider it as a normal case, because according to the medical science it very rare case to get atumour in such a huge range on a breast surface.

The most troublesome was to identify a malignant cell for the dense tissue case. It shows the revers problem of Fatty tissue case, i.e. it was easy to identify a normal case, but difficult to find out the malignancy case. In the following table 1 we have discussed our details results of identification of malignancy in different situations and its prediction rate has been discussed

Table 1: Prediction Rate for Various Tissue Type

Tissue Type	FP	TN	FN	TP
Fatty Tissue	**5.78%**	**39.33%**	**8.36%**	**47.25%**
Glandular Tissue	**4.93%**	**41.57%**	**4.98%**	**49.09%**
Dense Tissue	**8.17%**	**37.73%**	**15.01%**	**39.67%**

In the next diagram,figure 5 shows the prediction rate based upon the different types of tissue in the form of a graph..

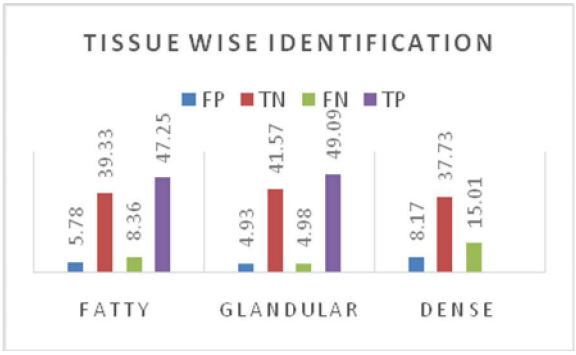


Fig 5 Prediction Rate as a Graph

The diagram represents that for fatty tissue region cases, it is more reliable to identify a malignancy, then a normal case and at the same time it identifies more a normal case as a malignant one than a malignant case as a normal one. That means it is not so much reliable for a normal fatty tissue cases. Now for the glandular tissue region, it identifies a positive case of malignancy than a positive case of normality. At the same time it identifies more normal cases as malignant one falsely, than identifying a malignant one as a normal one. Anyway but for the glandular tissues it is most reliable one for both cases. Lastly for the dense tissue cases, more normal cases than malignant cases in positive way, whereas the malignant cases are considered as normal one falsely more than the cases where the normal cases are considered as malignant one.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 12, December 2016

Now if we consider the rate of correctly identified or rejected cases, we could understand the overall success rate for the proposed method. The rate of correctly identified cases are represented as sensitivity and rate of correctly rejected rate is measured by specificity. In the Table 2 they are given tissue wise. The calculation formula of sensitivity and specificity are as follows:

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

Table 2: Specificity and Sensitivity Rate

	Sensitivity	Specificity
Fatty Tissue	**0.85**	**0.87**
Glandular Tissue	**0.91**	**0.89**
Dense Tissue	**0.72**	**0.81**

Now in figure 6, we represent the above specificity and sensitivity in terms of percentage and represent them as a graph, we could visualize the success rate for this proposed system.

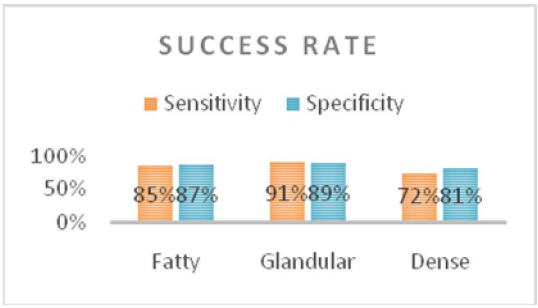


Fig 6 Sensitivity and Specificity Rate Graph for Various Tissue Type

VI.CONCLUSION

In this paper the combined filtering technique helps us to find out the most probabilistic suspected pixel of malignancy and at the same time our automatic region growing algorithm grows with that suspected pixel, so that the exact region of malignancy can be identified efficiently than the other existing techniques. In the vast majority of the cases we were getting the outcomes what we really needed to discover. In a genuine circumstance it was likewise difficult to distinguish if there was any malignancy of not. Considering those cases, we have the capacity to recognize if there is any kind of danger or not. However, according to medicinal motivation behind treatment, it ought to additionally extremely accommodating to a specialist that the amount of have those tumour tissues been spread.

REFERENCES

[1] Yasmin, M., Sharif, M., & Mohsin, S. (2013). Survey paper on diagnosis of breast cancer using image processing techniques. *Research Journal of Recent Sciences ISSN*, *2277*, 2502.

[2] Borchartt, T. B., Conci, A., Lima, R. C., Resmini, R., & Sanchez, A. (2013). Breast thermography from an image processing viewpoint: A survey. *Signal Processing*, *93*(10), 2785-2803.

[3] Jain, A., Jain, A., Jain, S., & Jain, L. (Eds.). (2000). *Artificial intelligence techniques in breast cancer diagnosis and prognosis* (Vol. 39). World Scientific.

[4] Patel, B. C., & Sinha, G. R. (2010). An adaptive k-means clustering algorithm for breast image segmentation. *International Journal of Computer Applications*, *10*(4), 35-38.

[5] Ramani, R., Suthanthiravanitha, D. S., & Valarmathy, S. (2012). A survey of current image segmentation techniques for detection of breast cancer. *International Journal of Engineering Research and Applications (IJERA)*, *2*(5), 1124-1129.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 12, December 2016

[6] Naik, S., Doyle, S., Agner, S., Madabhushi, A., Feldman, M., & Tomaszewski, J. (2008, May). Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology. In *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro* (pp. 284-287). IEEE.

[7] Thyagarajan, R., & Murugavalli, S. (2012, December). Segmentation of Digital Breast Tomograms using clustering techniques. In *2012 Annual IEEE India Conference (INDICON)* (pp. 1090-1094). IEEE.

[8] Ramani, R., Valarmathy, S., & Vanitha, N. S. (2013). Breast Cancer Detection in Mammograms based on Clustering Techniques-A Survey. *International Journal of Computer Applications*, *62*(11).

[9] Raman, V., Sumari, P., & Then, P. (2011). Matab Implementation and Results of Region Growing Segmentation Using Haralic Texture Features on Mammogram Mass Segmentation. In *Advances in Wireless, Mobile Networks and Applications* (pp. 293-303). Springer Berlin Heidelberg..

[10] Hassan, S. A., Sayed, M. S., & Farag, F. (2014, December). Segmentation of breast cancer lesion in digitized mammogram images. In *Biomedical Engineering Conference (CIBEC), 2014 Cairo International* (pp. 103-106). IEEE.

[11] Basha, S. S., & Prasad, K. S. (2009). AUTOMATIC DETECTION OF BREAST CANCER MASS IN MAMMOGRAMS USING MORPHOLOGICAL OPERATORS AND FUZZY C--MEANS CLUSTERING. *Journal of Theoretical & Applied Information Technology*, *5*(6).

[12] Yerushalmi, R., Woods, R., Ravdin, P. M., Hayes, M. M., & Gelmon, K. A. (2010). Ki67 in breast cancer: prognostic and predictive potential. *The lancet oncology*, *11*(2), 174-183

[13] Amini, A. A., Weymouth, T. E., & Jain, R. C. (1990). Using dynamic programming for solving variational problems in vision. *IEEE Transactions on pattern analysis and machine intelligence*, *12*(9), 855-867.

[14] Tambasco, M., Eliasziw, M., & Magliocco, A. M. (2010). Morphologic complexity of epithelial architecture for predicting invasive breast cancer survival. *Journal of translational medicine*, *8*(1), 1.

[15] Lee, K. M., & Street, W. N. (2000). Automatic image segmentation and classification using on-line shape learning. In *Applications of Computer Vision, 2000, Fifth IEEE Workshop on*. (pp. 64-70). IEEE.