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
IMPACT FACTOR: 5.258

IJCSMC, Vol. 5, Issue. 5, May 2016, pg.376 – 381

Breast Cancer Detection Algorithm by Adaptive Thresholding

[1]M. Selvamurugan, [2]G. Kharmega Sundararaj
[1]Department of CSE, PSNCET, India
[2]Associate Professor, Department of CSE, PSNCET, India
becoolandclear@gmail.com

Abstract*— Breast cancer is the most common cancer among the Indian women and it ranges from 25 to 31% of all cancers among Indian women. It is better to treat this dreadful disease at the earliest in order to save invaluable lives. For this sake, we developed a system that automatically detects and classifies the suspicious lesions present in the mammograms. The results are accurate because two levels of segmentations namely coarse and fine segmentation are employed. Coarse segmentation is done with the help of histogram based fuzzy c means technique, which is known for its accuracy, since it takes degree of truth and false into account. After obtaining the local sketch of the suspicious region, fine segmentation is applied in order to improve the rough representation of coarse segmentation and this is achieved by window based adaptive thresholding method. Finally, the outcome of fine segmentation is superimposed over the coarse segmentation to arrive at the perfect result. Then, the features such as area, circularity, correlation of pixel intensity, eccentricity and entropy of intensity distribution are extracted and the image is classified as normal, benign or malignant.

***Keywords*— Mammogram, segmentation, feature extraction**

I. Introduction

Breast cancer is the cancer that is a type of malignant tumour and grows in the cells of breast. Mostly, breast cancer is common among women but in very rare occasions it can be found in men too. The cancer cells of malignant tumour spreads up to the nearby area, if not treated at the right time. Breast cancer is the most common cancer among the Indian women and ranges from 25 to 31% of all cancers in India.

The average age of the occurrence of breast cancer is 30-50 years and previously it was 50-70. The severity or aggressiveness is directly proportional to the age of the person. The survey of Globocan (WHO) in the year of 2012, reports that the death count of breast cancer patients in India is 70,218 and stood first in the death rate. China ranked second with the death count of 47,984 and United States ranked third with 43,909 deaths [1].

The earlier the cancer is detected, the easier it is to cure it. Thus, early detection of cancer is necessary. Mammography is the best reliable technique for the detection of breast cancer, since it can detect 85 to 90% of all kinds of breast cancer. The abnormalities cited in a mammogram are masses and calcifications. Calcifications are the deposits of calcium. The cancer can be categorized as benign or malignant and this is determined by the shape of the mass.

Usually, benign tumours are round or oval in shape and a malignant tumour can be observed with a partially rounded mass with a spike or an irregular outline [7]. Breast cancers can be detected earlier with the help of mammograms and can help physicians to diagnose and deal with it.

The abnormalities present in the mammograms are detected and extracted in order to arrive at the correct diagnostic results. Some grayscale based segmentation methods are found to be effective in extracting the exact edges of the homogenous grayscale regions. The appearance of breast cancer is not stable at their early stages. Hence, the physicians may not be able to locate the abnormalities. In such cases, this automated system helps the physicians to detect the abnormalities easily. A tumor detection algorithm has to identify the lesion and is needed to be accurate with reduced number of false negatives.

Segmentation of medical imagery is a challenging task due to the complexity of the images, as well as the absence of those anatomy models that fully capture the possible deformations in each structure. Particularly, Brain tissue is a complex structure, and its segmentation is an important step for derivation of computerized anatomical images.

Recently there has been lot of research work carried out in mammogram image segmentation based on various soft computing technique [2-12] shows that other than tumor nothing was detected in MR images. Manual segmentations can also be used to detect tumor cell at any specific stage but it is a time consuming process. Soft computing refers to a collection of computational techniques in computer science, artificial intelligence, machine learning, attempting to study, model, and analyze very complex phenomena, like brain tumor cell detection. The objective of this work is to detect the breast cancer cells at early stage and increase the accuracy and also to reduce the computation time of the whole process.

II. LITERATURE SURVEY

Image segmentation has been an area of active research for the past two decades resulting in several image segmentation techniques that have been proposed and described in the image processing research literature. This proliferation is in part due to the fact that there exist several problem domains and applications that need to process and interpret image data in a domain-specific or application specific manner. Moreover, depending on the problem domain or application, there are several types of images that could be processed and analyzed such as, light intensity (grayscale), color, range (depth), thermal (infrared), sonar, X ray (radiographic), nuclear magnetic resonance images (MRI), and so on.

Image Segmentation is a process by which raw input image is partitioned into different regions such that each region in the image should satisfy the following properties:
(1) Regions of segmented image should be uniform and homogeneous with respect to some characteristic, such as gray level, color, or texture.
(2) Region interiors should be simple and without many small holes.
(3) Adjacent regions of segmentation should have significantly different values with respect to the characteristic on which they are uniform.
(4) Boundaries of each segment should be smooth, not ragged, and should be spatially accurate.
General techniques for image segmentation are listed below
(1) Clustering-based approaches
(2) Edge-based approaches
(3) Region-based approaches
(4) Based on Thresholding

A. Clustering based Approaches

The general problem in clustering is to partition a set of vectors into groups having similar values. In image analysis, the vectors represent pixels or sometimes small neighborhoods around pixels. The components of these vectors i.e. Features can include:

- Intensity values
- RGB values and color properties derived from them
- Calculated properties
- Texture measurements

1) K-means algorithm

In traditional clustering, there are K clusters C1,C2,...,CK with means m1,m2,..,mK. A least squares error measure can be defined as measures how close the data are to their assigned clusters. Least-squares clustering procedure could consider all possible partitions into K clusters and select the one that minimizes D.

B. Edge based Approaches

Edge detection is the most common approach in detecting the discontinuities in intensity value. Such discontinuities are detected by using first- and second order derivatives.
(i) Edge linking and boundary detection

The edges getting from Gradient operator are discontinuous. So, we use edge linking methods to combine those edges.
(a) Local Processing: 3x3 masks can be applied to all pixels in order to find the pixels are edge pixels or not. All pixels that are similar according to predefined criteria are linked.
Two principle properties are used to find the similarity between the edge pixels. The merits if this approach is less computation time and good for small images. The limitation is that the edge is discontinuous.

C. Region based Approaches

Region Growing Approach

In region growing group the pixels that are similar based on predefined criteria. The basic approach is start with set of seed points and from these grows regions by appending to each seed those neighboring pixels that have similar properties to the seed. The selection of similarity criteria and number of seed points depends upon the type of application.
Algorithm:
(a)Start with an initial seed pixel.
(b)Choose neighboring pixels, based on a connectivity and merge pixels that satisfy the homogeneity condition.
(c)If the region does not grow anymore select another seed and repeat the process Until all pixels are accounted for.
(d)A final tidying operation is often performed to remove very small regions. The merit of this approach is that it has continuous contour.

D. Thresholding Approaches

Thresholding: It separates foreground image from background image based on some threshold value. If the pixel gray level intensity value is less than the assumed threshold value then we group that pixel into foreground image and otherwise, it can be grouped into background object.
1) Global Thresholding
This is applied to whole image, and extract foreground from background based on threshold value T. This approach is well suited for images whose background intensity varies with foreground intensity. This approach does not work well for images with high illumination change. The estimation of threshold is also a vital factor.
2) Basic Adaptive Thresholding
Due to the uneven illumination in the given image the histogram can not be partitioned based on single threshold value like in Global thresholding. So, in order to avoid this sub divide the given image into sub-regions, and then segment or extract foreground image from background image for each sub-region based on local threshold for each sub-region. The approach
i. Works well even the image having different illumination.
ii. Works well for simpler images.

III.PROPOSED WORK

In this work, we propose an algorithm that detects suspicious lesions in mammograms, so as to simplify the work of a physician or a radiologist. The proposed system proves its accuracy which when compared with Fuzzy C means, K-Means, Otsu and Adaptive Thresholding methodologies. The work flow followed in this work is presented below. In the proposed system, wavelet lifting transform is done twice on an image and the sub-images of lower frequency are obtained at various resolutions.

The suspicious lesions are localized from the roughest representation through the histogram based fuzzy c-means technique. Then, the rough representation is improved by the window based adaptive thresholding in order to arrive at the fine segmentation. Then, the first order and run length features are extracted, in order to classify an image into normal, benign or malignant.
The proposed algorithm is presented below.
Step-1: Initially, the image is applied wavelet lifting transform twice and the low frequency sub-images at various resolutions are reaped out of it.
Step-2: I0 is the original MRI image that has the finest resolution and I1, I2 be the sub-images of the low resolution of the original image, where I0,I1 and I2 collectively represent the multiresolution of the original MRI image.
Step-3: Initially, the coarse segmentation is done by employing histogram based fuzzy means C technique, in order to obtain the rough sketch so as to locate the suspicious cells.
Step-4: The so-produced rough sketch of the coarse segmentation is further improved by the fine segmentation, such that this system can arrive at a perfect result.
Step-5: This fine segmentation is done by window based adaptive thresholding method.
Step-6: Then, the outcome of fine segmentation is superimposed over the coarse segmentation results.

Step-7: Extract features such as area, circularity, correlation of pixel intensity, eccentricity and entropy of intensity distribution.
Step-8: Image classification is done by SVM in order to distinguish between normal and the abnormal cells.
The detailed flow of the proposed work is presented in fig 1. The major components are coarse, fine segmentation and classification. This work employs SVM as its classifier for distinguishing between the normal and the abnormal cells.

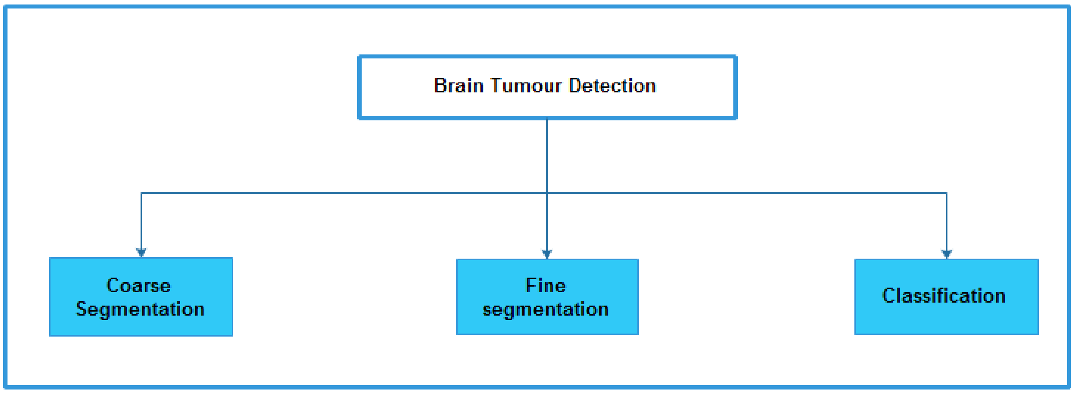

Fig 1: Overall flow of the proposed work

The stepwise screenshots are presented in figure 2.

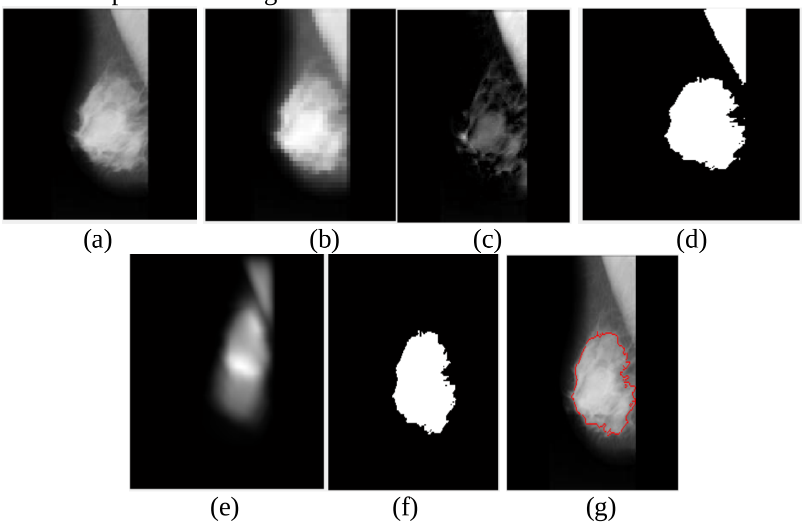

Fig 2: (a) Original Image, (b) Wavelet lifting transform (c) Coarsely segmented image (d) Convoluted image (e) Finely segmented image (f) Convoluted image (g) Cancer cell detection

Thus, all the steps involved in the proposed algorithm is diagrammatically presented above. This work is composed of three modules and they are as follows

- Pre-processing
- Coarse segmentation
- Fine segmentation
- Feature extraction
- Classification

Initially, the images are pre-processed to make them suitable for the forthcoming processes. Lifting Wavelet Transform is applied in the pre-processing step for effective processing. Coarse segmentation is done by the Histogram based Fuzzy C Means algorithm. This level of segmentation is to locate the suspicious cells. This is followed by the application of fine segmentation, which is done by window based adaptive thresholding method. Then, the outcome of fine segmentation is superimposed over the coarse segmentation results. Then, the outcome of fine segmentation is superimposed over the coarse segmentation results.

The process of feature extraction is achieved by extracting features such as area, circularity, correlation of pixel intensity, eccentricity and entropy of intensity distribution. Image classification is done by SVM in order to distinguish between normal and the abnormal cells.

IV. Performance Analysis

The performance of the proposed work is evaluated with respect to different classifiers. The experimental results show that the SVM works better than k-NN classifier. The accuracy of the system is measured by the following equation.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Where TP – True Positive, TN – True Negative, FP – False Positive, FN – False Negative and these are given by the following equations.

$$TP = \frac{Number\ of\ Correctly\ classifed\ images}{Total\ number\ of\ images} \times 100 \quad (2)$$

$$TN = \frac{Number\ of\ correctly\ rejected\ images}{Total\ number\ of\ images} \times 100 \quad (3)$$

$$FP = \frac{Number\ of\ wrongly\ classifed\ images}{Total\ number\ of\ images} \times 100 \quad (4)$$

$$FN = \frac{Number\ of\ wrongly\ rejected\ images}{Total\ number\ of\ images} \times 100 \quad (5)$$

The graphical results are presented below in figure 2.

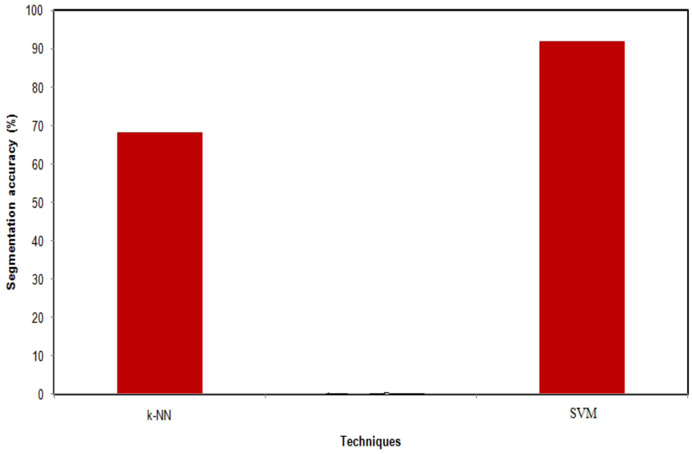

Fig 2: Segmentation accuracy

From fig 2, it is obvious that the performance of SVM is better than k-NN. k-NN classifier shows 70% segmentation accuracy and the SVM outperforms it with 91.5% accuracy rate.

V. Conclusions

In this work, a system that automatically detects suspicious cancer cells is presented. This work utilizes the lifting wavelet transform so as to process the image effectively. This work exploits two levels of segmentation namely coarse and fine segmentation. The coarse segmentation is accomplished by the histogram based fuzzy c segmentation and the fine segmentation is achieved by the window based adaptive thresholding method. Finally, the fine segmented outcome is superimposed over the outcome of coarse segmentation. Then, the features such as area, circularity, correlation of pixel intensity, eccentricity and entropy of intensity distribution are extracted and finally, the image is classified with SVM as normal, or abnormal.

References

[1] Gonzales R.C. and Woods R.E (2002), 'Digital Image Processing', Prentice Hall, Englewood Cliff, NJ.
[2] Linda G. Shapiro and George C. Stockman (2001), 'Computer Vision', New Jersey, Prentice-Hall, ISBN 0-13-030796-3, pp. 279-325.
[3] Clark M.C., Hall L.O., Goldgof D.B., Velthuizen R., Murtagh F.R. and Silbiger M.S. (1998), 'Automatic tumor-segmentation using knowledge-based techniques' IEEE Transactions on Medical Imaging Vol. 117, pp.187 - 201.
[4] Vannier M.W., Speidel C.M. and Rickman D.L. (1988), 'Magnetic resonance imaging multispectral tissue classification', News Physiol. Sci., Vol. 3, pp. 148 - 154.
[5] Liang Z. (1993), 'Tissue classification and segmentation of MR images', IEEE Eng. Med. Biol., Vol. 12, pp. 81-85.
[6] Zhang W. and Pan F.Z. (2002), 'A genetic algorithm based fuzzy clustering', Journal of Hubei University (Natural Science), Vol. 24, No. 2, pp. 101- 104.
[7] Sahoo P.K., Soltani S., Wong A.K.C. and Chen Y.C. (1988), 'A survey of thresholding techniques', Computer Vision, Graphics Image Process, Vol. 41, pp. 233 - 260.
[8] Bezdek J., Hall L. and Clarke L. (1993), 'Review of MR image segmentation using pattern recognition', Med Phys; Vol. 20, pp. 1033 -1048.

[9] Peli T. and Malah D. (1982), 'A study of edge detection algorithms', Computer Graphics and Image Processing, Vol. 20, pp.1-21.
[10] Bouman C.A. (1994), 'A multiscale random field model for Bayesian image Segmentation', IEEE Transactions on Image Processing, Vol. 3, pp. 162-176.
[11] Bhandarkar S.M. and Nammalwar P. (2001), 'Segmentation of Multispectral MR images Using a Hierarchical Self-Organizing Map', Computer-Based medical system CBMS 2001, Proceedings, 14th IEEE Symposium on 26, Vol. 27, pp. 294 - 299.
[12] Bezdek J., Hall L. and Clarke L. (1993), 'Review of MR image segmentation using pattern recognition', Med Phys; Vol. 20, pp. 1033 -1048.

