

A NOVEL ALGORITHM FOR EXTRACTING THE FEATURES OF LESIONS FROM MAMMOGRAM USING MODIFIED EXPECTATION AND MAXIMIZATION

Ms.Sushree Prajnabasini Pani¹ Mr. V. Vedanarayanan²

Sathyabama University

Chennai, India

contactprajna@gmail.com Veda_gopal@gmail.com

ABSTRACT

Cancer is uncontrolled growth of cells. Breast Cancer is the uncontrolled growth of cells in the breast region. Breast cancer is the second leading cause of cancer deaths in women today. Early detection of the cancer can reduce mortality rate. Early detection of Breast Cancer can be achieved using Digital Mammography, typically through detection of characteristic masses and/or micro calcifications. A Mammogram is an x-ray of the breast tissue which is designed to identify abnormalities. Therefore, it would be valuable to develop a computer aided method for mass/tumor segmentation based on extracted features from the Region of Interest (ROI) in mammograms. The Modified Expectation-maximization (MEM) algorithm is proposed to segment the breast cancer images. In this work, an efficient approach to search for the global threshold of image using Gaussian mixture model is proposed. Firstly, a gray-level histogram of an image is represented as a function of the frequencies of gray-level. Then to fit the Gaussian mixtures to the histogram of image, the Expectation Maximization (EM) algorithm is developed to estimate the number of Gaussian mixture of such histograms. With this scheme we can detect the lesions more effectively.

Index Terms-Mammogram, Median filtering, Modified expectation maximization, CLAHE, Gray Level Co-occurrence Matrix (GLCM), Support Vector Machine (SVM) Classifier.

1. INTRODUCTION

A mammogram is a radiograph of the breast tissue. It is an effective process for examining the breast, commonly searching for breast cancer. Cancer is not preventable, but early detection leads to a much higher chance of recovery and lowers the mortality rate from this disease. A cancer cell has characteristics that differentiate it from normal tissue cells with respect to: the cell outline, shape, structure of nucleus. When this happens in the breast, it is commonly termed as 'Breast Cancer'. Cancer is confirmed after a biopsy (surgically extracting a tissue sample) and pathological evaluation. The most important risk factor is age.

The incidence of breast cancer increases with age. The majority of women diagnosed with breast cancer are over the age of 50. Women under the age of 25 are least likely to develop breast cancer. However, younger women who do contract breast cancer, suffer from speedy and aggressive growth of cancer. The most common clinical sign of breast cancer is a painless, hard and fixed lump in the

breast. This is one of the reasons that make clinical detection of breast cancer very difficult, as painlessness gives the woman a false sense of security. Based on the morphological and the texture features [1] the masses can be identified. Another technique is based on the DCT features based malignancy and abnormality type detection [2].

Considering the tumor preprocessing segmentation feature extraction and classification [3], a technique known as K-mean algorithm for mass segmentation is done. MEM is also a method that is for dimensionality reduction [4], which is used for tumor detection. Generally all classifiers only classify the images but do not segment the tumor [5]. SVM Classifier can use a technique known as clustering [6]. SVM can also be used for the detection of the microcalcification [7]. The proposed algorithm comprises of expectation maximization method which helps to find the tumor more accurately and gives the more accuracy value. Section II deals with EM, Section III with the proposed algorithm.

2. EXPECTATION MAXIMIZATION

A common task in image processing is the estimation of the parameters of probability distribution function. Expectation maximization is ideally suited to produce the maximum likelihood. This algorithm consists of two major steps, first is Expectation step and second is Maximization step. The expectation step consists of calculating the expected value of the complete data likelihood function. In the Maximization step, first plug in the expected values of the latent variables in the log-likelihood of the given data. Then, maximize this log-likelihood to evaluate the parameters. The main goal of the EM algorithm is to facilitate maximum likelihood parameter estimation by introducing so-called hidden variables which are not observed. Let us denote these hidden variables by some Z , and the unobserved data by a vector z . Note that it does not play a role if this unobserved data is technically unobservable or just not observed due to other reasons. It is just an artificial construct in order to enable an EM scheme. We define the complete data x to consist of the incomplete observed data y and the unobserved data z :

$$\mathbf{x} = (\mathbf{y}^T, \mathbf{z}^T)^T$$

Then, we consider the log-likelihood function of the complete data:

$$\text{Log} L_C(\theta) = \log p(\mathbf{x} | \theta)$$

A. Methodology

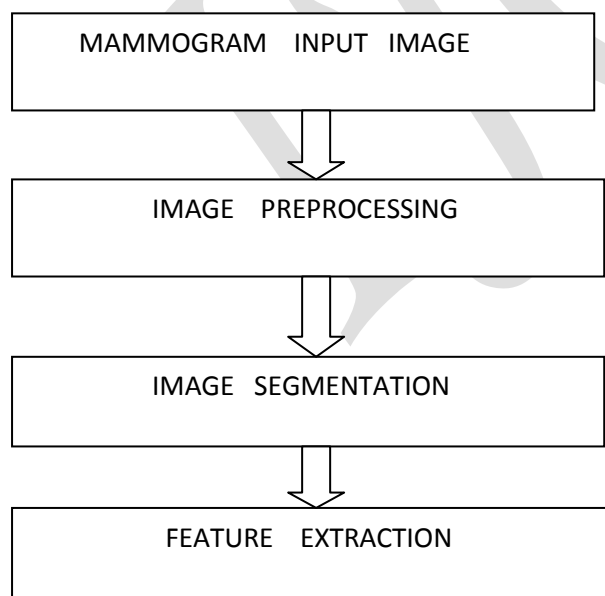


Figure.1 Methodology

3. PROPOSED METHOD

A. Modified Expectation & Maximization (MEM)

MEM is the Modified expectation maximization algorithm is a natural generalization of maximum likelihood estimation. MEM algorithm has been widely applied to the estimation of Gaussian mixture model in cancer image segmentation. However it is deterministic process. In this paper we propose an algorithm for breast cancer segmentation. In this approach the MEM algorithm is performed to estimate the maximum likelihood and to get the more efficient accuracy value. Segmentation of lesions of breast cancer image has been classified into three layers that are Gray matter, White matter and Cerebrospinal fluid.

The objective of our experiments is to classify our image data sets by more than 85% for WM, GM and CSF. Estimating the unknown probability density function (pdf) given a finite set of observation is an important aspect of many image processing problems. MEM algorithm is an iterative procedure that consists of two alternating steps. They are

1. Expectation Step followed by a Maximization Step. Expectation step is used to the unknown cancer tissues using the current estimate of the parameters and conditioned on the observations.
2. The maximization step provides new estimates of the parameters. At each iteration, the estimated parameters provide an increase in the maximum likelihood function.

There are two main applications of MEM algorithm. The first occurs when the data indeed has missing the values due to problems with the limitations of the observation process. The second occurs when calculating the likelihood function is analytically intractable.

B. Algorithm for proposed method

1. Initialize the value for the mixture parameters.
2. Calculate the intermediate function for the given set of initial values.
3. Calculate the new mixture parameters by using the intermediate function and previous mixture values.
4. Continue the calculation until the threshold or validation condition is met.
5. Repeat for each class cancer tissue in the image.

C. Example

Let X be the observed data and suppose a complete data set is $Z = (X, Y)$

We also have a joint density function

$$p(z|\theta) = p(x, y|\theta) = p(y|\theta, x) p(x|\theta)$$

We can define a likelihood function

$L(\theta|Z) = L(\theta|X, Y) = p(X, Y|\theta)$, called the complete data likelihood. The original likelihood $L(\theta|X)$ is referred to as the incomplete data likelihood function. The MEM algorithm first finds the expected value of the complete data log likelihood $\log p(X, Y|\theta)$ with respect to the unknown data Y given the observed data X and the current parameter estimates. That is

$$Q(\theta, \theta^{(i-1)}) = E[\log p(X, Y|\theta) | X, \theta^{(i-1)}]$$

Where $\theta^{(i-1)}$ is the current parameter estimate used to get the expectation value. θ is the new parameter optimized to increase Q . The evaluation of this expectation is called the E-step of the algorithm. The second step of the MEM algorithm is to maximize the expectation calculated in the first step.

$$\theta^{(i)} = \arg \max_{\theta} Q(\theta, \theta^{(i-1)})$$

These two steps are repeated as important. Each iteration is guaranteed to increase the log-likelihood.

B. Methodology of proposed method

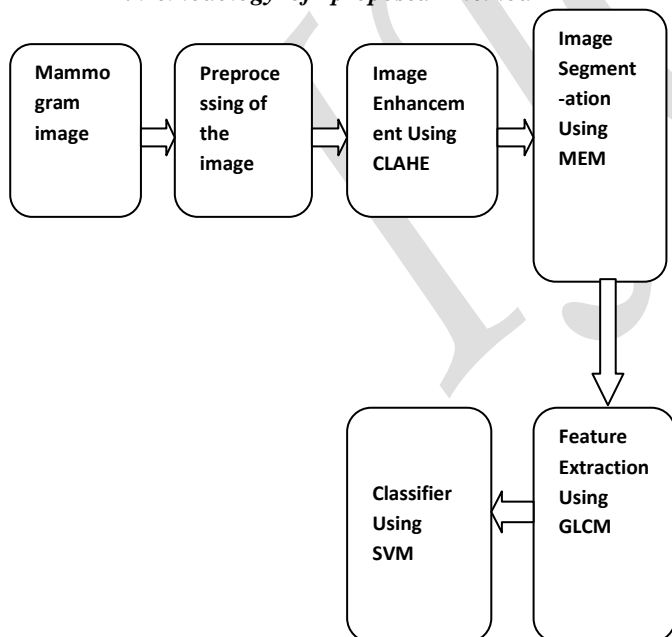


Figure. 2 Methodology of proposed method

4. RESULTS AND DISCUSSION

This section results of the automatic classification on mammogram using GLCM features. The main features are to be calculated as

1. Autocorrelation
2. Energy
3. Maximum probability
4. Sum of variance
5. Sum of average
6. Entropy
7. Homogeneity
8. Information measure of correlation
9. Inverse difference normalization
10. Inverse difference moment normalized (IDN)
11. Maximal correlation coefficient
12. Minimal correlation coefficient
13. Mean
14. Variance

By the proposed algorithm the following output cancer images will be detected where the exact segmentation of the masses is possible.

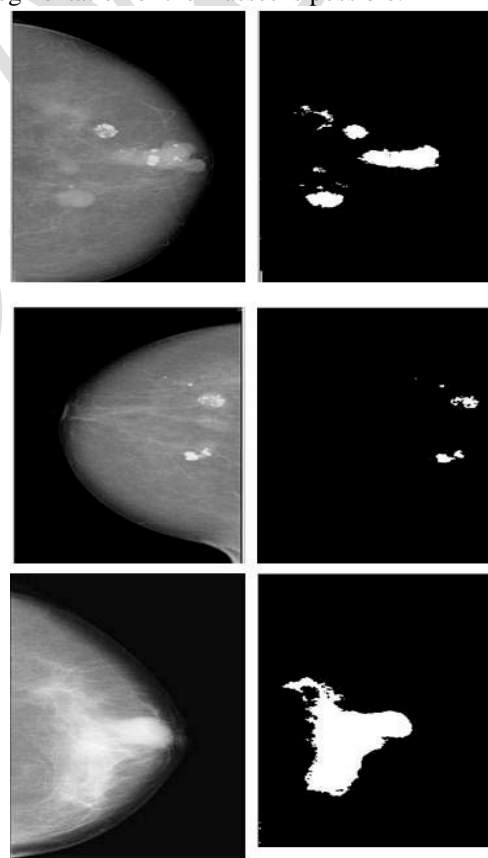


Figure.3.1 Input images 3.2 Output images

5. CONCLUSION

The method employed in this paper has given the better performance. The Modified Expectation Maximization (MEM) is developed for the segmentation of mammogram with the aim of supporting the radiologists in visual diagnosis. The usage of MEM increases the accuracy and sensitivity of the tumor detection.

REFERENCES

- [1]. Faleh H. Mahmood, Remote Sensing Unit, College of Science, Baghdad University, Baghdad, Iraq, Email: faleh_sine@yahoo.com "Early detection of breast cancer mass lesions by mammogram segmentation images based on texture features" *Iraqi Journal of Physics*, 2012, Vol. 10, No.17, PP. 29-40.
- [2]. M. Arfan Jaffar, Nawazish Naveed, Sultan Zia, Bilal Ahmed and Tae-Sun Choi "DCT Features Based Malignancy And Abnormality Type Detection Method For Mammograms" *International Journal of Innovative Computing, Information and Control ICIC International* ©2011 ISSN 1349-4198 volume 7, Number 9, September 2011 pp. 5495-5513.
- [3]. Ms. D. Roja Ramani, Ms. B. Ben, Sujitha, Ms. C. Parameswari, "Digital Mammogram Tumor Preprocessing Segmentation Features Extraction and Classification Using Kmean Algorithm and Distance based Skeletonization" *International journal of computer application*, issue 2 (April, 2012) ISSN-2252-1797.
- [4]. Dr. K. Revathy, Dept. of Computer Science, University of Kerala, India Roshni V. S., ER&DCI Institute of Technology, CDAC-T, India "Applying EM Algorithm for Segmentation of Textured Images" Proceedings of the World Congress on Engineering 2007 Vol I, WCE 2007, July 2 - 4, 2007, London, U.K.
- [5]. Nasseer M. Basheer, Mustafa H. Mohammed Computer Engineering Department, Iraq Technical College/ Mosul, Iraq "Classification of Breast Masses in Digital Mammograms Using Support Vector Machines" *International journal of advanced research in computer science and software engineering*, volume 3, issue 10, October 2013, ISSN 2277 128X.
- [6]. Moi Hoon Yap, Eran A. Edirisinghe, and Helmut Bez, Department of Computer Science, Loughborough University, Loughborough, U.K. M.H.Yap@lboro.ac.uk "A Novel Algorithm For Initial Lesion Detection In Ultrasound Breast Images" *JOURNAL OF APPLIED CLINICAL MEDICAL PHYSICS*, VOLUME 9, NUMBER 4, 2008.