Detection Of A Breast Tumor In Mammogram Images Employing Various Image Processing Techniques

Shaik Khasim Saheb, Dr.B.Narayan, Dr.T.Venkat Narayana Rao, Sai Manoj

Abstract: Breast cancer is the serious issue facing by the many of the women around the world. Image processing (IP) techniques are used to improve the usage of data. There are different techniques like image analysis, edge detection, filters and noise removal methods, which helps in understanding the data present in the image as well as to study the image properties. This paper is discussed about the breast cancer detection and classification from a mammographic image. It is also focused on the two properties, one is to classify the type of tumor and another is to recognize the tumor from low contrast background area. The proposed system follows these steps 1) enhancement of input image 2) segmentation of tumor 3) extraction of features from the image 4) classification of tumor based on the features extracted. The input image is to be a mammographic image which contains the different noises, these noises are removed and image is enhanced by the speckle noise removal method, the enhanced image is used for the segmentation of tumor area by iterative modified watershed algorithm. The features or attributes are taken out from the image that is segment or partitioned by gray level co-occurrence matrix methods and are used for the tumor classification by SVM classifier. This process gives accuracy of 98%.

Index Terms: Mammogram, Breast Tumor, Enhancement, EM method, Segmentation, Classification, SVM, Feature Extraction, GLCM

1. INTRODUCTION

Breast cancers are one of the most dangerous tumors, multiply their cells unanimously and uncontrollably. These are the formation of lumps in the breast. These cells are formed in the breast and in that some of them doesn’t move and some of them move throughout the body parts by the process called metastasis. The cells which don’t move are termed as essential breast tumors (EBT) and the cells which move are termed as optional breast tumors (OBT). OBTS are more dangerous than EBT’s. These cancer cells can be detected or identified in the mammographic images. The lump’s in the breast causes irregularities which in turn are dangerous and more over it is destructive. Mammograms show the clear differences between normal tissues and affected tissues. It is one of the general technique utilized to medicinal imaging. Therapeutic image investigation has great importance in field of pharmaceutical, mainly in clinical study. In this paper it is motivated that classification and detecting the breast cancer accurately and it is comparing with already existing systems. We have used few methods which all together give the output with great accuracy, those methods are removal of speckle noise method, Expectation-Maximization (EM) algorithm[10], repetitively modified segmentation using watershed, Gray-Level Co-occurrence Matrix (GLCM) method, and Support Vector Machine (SVM) classification[7]. These methods that we use were individually proved as good in their previous studies.

2 RELATED WORK

In this section we survey about different techniques and methods which are applied for the image processing, tumor detection and classification. Author Fatima Eddaoudi Et al used a system by taking the Thresholding as main factor and proposed mass and uncommon tissues. These methods are tested on database of MIAS to give greater performance in detecting tumors detection threshold and classified using the SVM classifier. From co-occurrence matrix Haralick features are calculated to exhibit the common when the compared the methods with the methods are proposed. Tingting Mu, by using 22 features of the segmented portion of tumor, proposed strict two-surface proximal(S2SP) classifier. By using the concept of fractal derivatives BhagwatiCharan Patel proposed a method to enhance the image using fractal technique. Fractal technique is always connected with fractional dimensionality and self-similarity criteria. This method was tested on the several images to cancer diagnosis & research using BSR APPOLO images. Osfati. E et al discussed about the Eigen factors and the outcomes are differentiated with GA based results. In one paper a swarm intelligence method based SVM classifier (PSO_SVM) is proposed. In the PSO_SVM, two models are simultaneously solved in PSO optimization (swarm) framework, those two models are model and feature selection. SVM accuracy rates average (ACC), no. of support vectors with the selected attributes concurrently and weighted function which designs PSO function. In the extension of this system to manage the global & local search in algorithm of PSO two features are considered, time varying acceleration coefficients (TVAC) and time varying inertia weight(TVIM)[8]. Different challenges, issues, methods of CAD in feature analysis and mass detection are discussed by Patel.B.G and Karthikeyan Ganesan. Using SVM classification with a CAD system and combined with LDA classification was introduced by Aloffe.M.A. Machine learning techniques were proposed by Osareh.A to diagnose the breast cancer. GA is used to detect the geno-type frequencies were proposed by Cheng- Hong Yang. Jinshan Tang developed a system of computer aided diagnosis to detect and diagnosis the cancer, by using the techniques finding, locating, detecting and classification. Feature extraction method was used to check...
whether the cancer cells are malignant or benign proposed by Mohammad Sameti Rabab. Xiangjini shi used a fuzzy SVM method in CAD system for the detection of masses. Amir Fallahi used Bayesian network and data preprocessing, to reduce the dimensions of the database using the relief algorithm, after that pre-processing method is applied on that and use Bayesian to classify the breast cancer. Hossein Rabbani used speckle noise removal method to remove noise and to enhance the image for better classification and detection of masses. From the study of related works in this domain says, there are different techniques that are previously used on images related to medical to detect tumor. In those various classification methods applied and classified images with less no. of input attributes. Because of less number of classification features accuracy in classification also less and it depends on the type of input images that are used for the tumor detection and classification.

3 EXISTING SYSTEM
In the existing system the breast cancer detection of mammogram images have used edge-based, image processing threshold, watershed segmentation methods. In this system he also presented a case study which is based on simplicity of the algorithm and detection time. But there is a limitation in this system i.e. all the three methods that were proposed are using threshold value for the main functionality in detecting the cancer cells. The selected features, that were used, were not providing clarity and not using standard feature extraction method or classifier method. This is major drawback in the existing system, but in this paper we use different stand methods for the feature extraction and classification on multi database based images.

4 PROPOSED SYSTEM
In this system we have divided the tumor detection in mammogram images into three levels. The level-1 involves the image enhancement, the level-2 involves the segmenting the tumor area, and the level-3 involves extraction of features and classification. The images that we take as input contain noise which is removed by using SNR (speckle noise removal) method[9]. The area of the tumor is segmented using the iterative modified watershed segmentation method and some binary operations to that. After this the tumor area features are extracted using GLCM feature extractor [2] and it is used to measure the segmented image properties and in the final step the final step by using all these features the classification process is started by using SVM classifier. The complete functionality of proposed system is shown in below fig[1].

The images which are given as input may be a colored image or gray scale image or any other kind of images. If any necessary the image is converted from one color into another color space. [ex: RGB to GRAY or HSV to GRAY].

4.1 REMOVAL OF NOISE IN AN IMAGE
The mammogram image which is taken from the database is preprocessed by removing the noise from the image by using SNR method. For more accurate output, the image is upgraded using Expectation-Maximization (EM) algorithm. The image is converted into gray scale if the image is in any other color space.

4.2 REMOVAL OF SPECKLE NOISE
Speckle noise is a small particle of noise that naturally exists in and decreases the quality of images of the engineered radar, therapeutic and dynamic radar images. It is the challenging aspect to reduce the noise from the therapeutic images, satellite images for the experts in digital image processing (DIP). There are few methods to reduce the noise. An unassailable normal for MRI is the existence of speckle noise. Speckle noise is an unpredictable and deterministic in the image. It has negative effect on the ultrasound imaging which shows as the contrasted in MRI. Speckle noise is also called as texture in medicinal written works. Generalized method of the speckle is represented as,

\[ f(x,y)=g(x,y)\cdot m(x,y)+n(x,y) \rightarrow (1) \]

whereas \( f(x,y) \) is observed images, \( m(x,y) \) is the part of multiplicative image and \( n(x,y) \) is segment part of the speckle noise of added image. Here \( x,y \) means the axial and parallel indices of the image. For the ultrasound imaging, only multiplicative image part of noise is considered and ignored the added image part. So the mathematical equation (1) can be changed as;

\[ f(x,y)=g(x,y)\cdot m(x,y)+n(x,y)-n(x,y) \]

Therefore,

\[ f(x,y)=g(x,y)\cdot m(x,y) \quad \text{-------------------}(2) \]

In this the speckle noise could be cleared by then again select any alternative of wavelet filter, average filter and so on. Choosing the filter depends on the image and the information present in the image. Before the clearing the noise the size of the image is expanded, and soften as piece sort of sub-images, and connect filter, accordingly the noise get removed totally.

4.3 ENHANCEMENT OF AN IMAGE
In this paper, expectation – maximization (EM) procedure is proposed to get firmness improvement. The current proposed method provides a general framework for producing good quality images at the designated firmness of huge class of image development procedure. In the EM procedure, assumed, inconsiderable “complete data set” is working to ease the procedure of maximizing the likelihood function of the measured data. The actual procedure consists of sequence of irregular expectation steps and expansion steps. This repetitive process has the necessary possessions of maximizing the likelihood function defined on the careful data monotonically, and meeting to the worldwide maximum at a unique point. A graphic illustration of the EM algorithm to enhance the firmness faced by perusing process is shown in
fig[2]. From the apparatus a glowing light source pulled from top to bottom which produces light on the shallow of the scanned hardcopy. The array of light sensors notices the concentration of light which is reproduced. However, the intensity of specific sensor is affected by area restricted by scanner and as well as adjacent areas due to dispersal of light. To reduce above mentioned effect and to increase firmness & to reinstates the innovative density the proposed method EM technique can be used.

**Fig2: EM Algorithm results Image Enhanced**

In fig2 the brightness is increased and we can differentiate the normal portion and tumor portion clearly on the image. The fig[2] is result of the EM algorithm which is developed in Matlab and produced.

**4.4 SEGMENTATION OF A TUMOR**

Image segmentation means the making into parts of images or groups based on the similarities and other related properties. Every image contains pixel is belongs to one of the quantity of these groups. A good separation takes place based on which:

a) pixels in the same section or group have familiar grey scale values and form an associated group.

b) adjacent pixels in the image which are in dissimilar sections have dissimilar values.

The IWS technique is helpful when the described area in the image has neither gray level unity nor texture. Furthermore, this region should be problematic to human eyes and it should have unclear outlines. It is need to show inner indicators and outer indicators by using mouse clacks. Moreover the model is detailed; more precise and accurate in the final segmentation or separation. For the work that we do, the minimum number of mouse clacks was of 5 outside and 3 inside, but it should depend on your image type. This may vary from one image type to another image type. The images that we are dealing with real tumor borders are unclear. This method has a more practical approach by adding an ambiguous portion with real tumor borders are unclear. This method has a more practical approach by adding an ambiguous portion with real tumor borders are unclear. This method has a more practical approach by adding an ambiguous portion with real tumor borders are unclear. This method has a more practical approach by adding an ambiguous portion with real tumor borders are unclear. This method has a more practical approach by adding an ambiguous portion with real tumor borders are unclear.

The IWS technique is helpful when the described area in the image has neither gray level unity nor texture. Furthermore, this region should be problematic to human eyes and it should have unclear outlines. It is need to show inner indicators and outer indicators by using mouse clacks. Moreover the model is detailed; more precise and accurate in the final segmentation or separation. For the work that we do, the minimum number of mouse clacks was of 5 outside and 3 inside, but it should depend on your image type. This may vary from one image type to another image type. The images that we are dealing with real tumor borders are unclear. This method has a more practical approach by adding an ambiguous portion with real tumor borders are unclear. This method has a more practical approach by adding an ambiguous portion with real tumor borders are unclear. This method has a more practical approach by adding an ambiguous portion with real tumor borders are unclear. This method has a more practical approach by adding an ambiguous portion with real tumor borders are unclear. This method has a more practical approach by adding an ambiguous portion with real tumor borders are unclear. This method has a more practical approach by adding an ambiguous portion with real tumor borders are unclear. This method has a more practical approach by adding an ambiguous portion with real tumor borders are unclear. This method has a more practical approach by adding an ambiguous portion with real tumor borders are unclear. This method has a more practical approach by adding an ambiguous portion with real tumor borders are unclear. This method has a more practical approach by adding an ambiguous portion with real tumor borders are unclear. This method has a more practical approach by adding an ambiguous portion with real tumor borders are unclear. This method has a more practical approach by adding an ambiguous portion with real tumor borders are unclear.
feature extraction method, which are later used to classify the images based on these feature values. These features are properties of the type of mammogram images and the relationship among the nearest pixels on the surface of the images. There are many textural features like mean, entropy, standard deviation of the pixels are computed by this approach. The spatial distribution features are obtained by using GLCM. The table indicates the features that are extracted which are used to classify the tumor portions in the image.

<table>
<thead>
<tr>
<th>FEATURE</th>
<th>DESCRIPTION</th>
<th>FORMULA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region size(area)</td>
<td>The region size of the segmented portion from the mammogram image is the no. of picture elements (pixels) inside of the region</td>
<td>( a = \sum_{i=1}^{n} i )</td>
</tr>
<tr>
<td>Mean(radii)</td>
<td>The sum of RGB values divided by 3</td>
<td>( \mu_R = \frac{\sum (R(i) - \mu_R)^2}{a} )</td>
</tr>
<tr>
<td>Standard deviation(texture)</td>
<td>The standard deviation of the RGB values inside the region. They are computed as</td>
<td>( \sigma_R = \frac{\sum (R(i) - \mu_R)^2}{a} )</td>
</tr>
<tr>
<td>Mean RGB</td>
<td>The average value of the mean of the RGB values</td>
<td>( \mu_R = \frac{\sum \text{R}(i)}{a} )</td>
</tr>
<tr>
<td>Uniformity</td>
<td>Similarity or the uniqueness among the pixels</td>
<td>( U = \frac{\sum p(i) \log_2 p(i)}{\sum p(i)} )</td>
</tr>
<tr>
<td>Entropy</td>
<td>Unpredictable information in the image is entropy</td>
<td>( p(i) )</td>
</tr>
</tbody>
</table>

Table 1: Features Extracted Using GLCM

6 SVM CLASSIFICATION
We have extracted several features and we take only 6 features to classify the tumor. From that one is area and the rest all are textural features. The SVM classifier is supervised learning models in which we give input then it will give the outputs. It is used to classify the tumor and non tumor area in the mammogram images. The features that we extracted are taken as input to the parameters of SVM classification[4] and it classifies into categories.

Proposed algorithm()

1. Take input picture or image I
2. I= noise_add(I, type of noise)
3. I= clear_noise(I,type of filter)
4. [x,y]= image_segmentation(I, threshold value)
5. BC= feature_Extraction(x)
6. Train= SVM - classifier(BC);

7. RESULT = Train
8. Result is tumor area or non – tumor area

7 EXPERIMENTAL RESULTS
To find out the efficiency of the approach that we used is programmed in the MATLAB software and made it to run on entire datasets. We have the tumor detection was gathered from the obtained result set. The results are shown in the figures given below in fig[4] and fig[5].

7.1 PERFORMANCE EVALUATION
To run a MATLAB software[1] takes a minimum time of 14 sec using personal computer with core i5 processor and 4GB RAM. The images are individually called for testing and training the detection as well as classification processes. The datasets is divided into two categories, half of them are for testing and half of them are for training. At last the SVM classifier is used to train the data set. Then compare each and every testing image with the training image and to classify. The results are shown in the fig [4], fig[5],[fig6] & fig[7].

Fig4: input image, segmentation using watershed, negative image, detected tumor area for color image

Fig 5: Breast cancer detection for mammogram images

The following values are computed using evaluation metrics such as specificity, sensitivity and accuracy by the following equations

Sensitivity (%) = \( \frac{PT}{PT+NF} \times 100\% \)

Specificity (%) = \( \frac{NT}{NT+PF} \times 100\% \)

Accuracy (%) = \( \frac{PT+NT}{N} \times 100\% \)

Where PT\( \rightarrow \) positive true; NT\( \rightarrow \) negative true; PF\( \rightarrow \) positive false ; NF\( \rightarrow \) negative false ;
N→ is the total number of images

In the fig[4] we have shown the enhanced input image and gray scale images which are used in the experiment. The features of the images are represented in the form of GLCM matrix obtained from the images.

### 7.2 CO-OCCURRENCE MATRIX

The gray level, contrast, homogeneity, correlation and energy are the textural features which are calculated from the GLCM which is shown in Table(2) and Table[3].

<table>
<thead>
<tr>
<th>Total no. of images</th>
<th>PT</th>
<th>PF</th>
<th>Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>78</td>
<td>2</td>
<td>97.50%</td>
</tr>
</tbody>
</table>

**Table 2: classification of sensitivity results**

\[
\text{Sensitivity (\%)} = \frac{PT}{PT + NF} \times 100\% = \frac{78}{(78+2)} \times 100 = 97.5\%
\]

\[
\text{Specificity (\%)} = \frac{NT}{NT + PF} \times 100\% = \frac{20}{(20+20)} \times 100 = 100\%
\]

\[
\text{Accuracy (\%)} = \frac{PT + NT}{P + N} \times 100\% = \frac{(78+20)}{(50+50)} \times 100 = 98\%
\]

Where P = PT + NF, N = PF + NT

<table>
<thead>
<tr>
<th>Total no. of images</th>
<th>NT</th>
<th>NF</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>20</td>
<td>0</td>
<td>100%</td>
</tr>
</tbody>
</table>

**Table 3: classification of specificity results**

8 CONCLUSION

In this paper, we have discussed the various steps in detecting the tumor automatically. The approach that we used with enhancement of image [SNR and Expectation-Maximization (EM) algorithm], repetitively modified segmentation using watershed, extraction of features using gray level co-occurrence matrix and classification using support vector machine which proved its performance by showing some performance metrics such as sensitivity of the image is 97.5%, specificity of the image is 96%, and classification of accuracy is 97%. The proposed method’s when compared to existing system provides the better performance, which is very useful for the medical people in tumor detection. If this is used in rural area then it helps in finding out the occurrence of the tumor in MRI. Further, it may be focused on detection of cancer automatically by classifying the MRI images.

REFERENCES


[4] Robust breast cancer prediction system based on rough set theory at National Cancer Institute of Egypt, Saeed Khodary M. Hamouda , Mohammed E. Wahed , Reda. Abo Alez , Khaled Riad, ELSEVIER: 2018


