International Journal of Scientific Engineering and Research (IJSER)
www.ijser.in
ISSN (Online): 2347-3878
Volume 3 Issue 2, February 2015

A Bayesian Network Based Classification of Breast Lesion in Digital Mammogram

Amruta V. Shelke

Savitribai Phule Pune University, Maharashtra, India

Abstract: ***Breast cancer is a serious issue in the worldwide females. Breast cancer is the type of cancer which develops from breast cells. For detection of breast cancer there are different types of screening techniques are available. For detection of breast cancer this paper includes several techniqus. In this paper, the first step was to remove noise from the image; median filter is used to remove the unwanted noise in the image. In the MIAS database pectoral muscles are available in the image, pectoral muscle are removed by calculating the thresholding value of an image.The entropy based segmentation approach is proposed to segment a gray-scale breast image. The approach calculates the histogram of an image also finds the entropy value of image. Then by finding the thresholding value of an image the segmented image is shown at the output. In this paper, an efficient and fast entropic method for noisy cell image segmentation is presented. Then the features like mean, standard deviation, Entropy, Skewness, Kurtosis, Variance, Energy, Correlation, Smoothness and Root mean square are extracted from a segmented image, this features are then given to the input of Bayesian network to classify the image according to the feature value. Experimental results show that the proposed method is efficient and much more tolerant to noise than other techniques.***

Keywords: Entropy, Breast cancer, Bayesian network, median filter, pectoral muscle, noise, ROI

1. Introduction

Breast cancer is the most common breast cancer in the worldwide females. The Breast cancer is a type of cancer that develops from breast cells. Around 18.2% of all cancer deaths worldwide, including both males and females, are from breast cancer. Breast cancer is the serious matter in developed nations comparing to developing ones. There are some reasons behind this breast cancer that is more common in elderly women, women in the richest countries live much longer than those in the poorest nations. Because of different lifestyle and eating habits of females in rich and poor countries is also a contributory factor. According to the National Cancer Institute, 232,380 female breast cancers and 2,240 male breast cancers are reported in the USA each year. According to the World Health Organisation (WHO), seven lakh Indians die of cancer every year, while over 10 lakh are newly diagnosed with some form of the disease.

The origin of breast cancer is from the inner lining of milk ducts or the lobules that supply them with milk. Malignant tumor spread to other parts of the body. The breast cancer that starts off in the lobules is known as lobular carcinoma, while one that developed from the ducts is called ductal carcinoma. There are billions of microscopic cells are available in the body. The cancer cells multiply in orderly fashion new cells are made to replace the ones that died.

The majority of breast cancer occurs in the females. The invasive breast cancer is spread over the body part such as bones, liver or lungs and the non-invasive breast cancer is still inside its place of origin and has not broken out. In cancer, the cells multiply uncontrollably and there are too many cells, progressively more and more than there should be. However, it is difficult for radiologists to provide accurate and uniform evaluation for the mammograms generated. The advances of digital image processing radiologist have an opportunity to improve their diagnosis with the aid of computer system.

In this paper the median filter is used to remove noise in the image. Filtering is the technique for modifying or enhancing an image. After removing noise from the image the pectoral muscles are removed from the image. The entropy segmentation is used to detect ROI which is present in the image. The features are extracted from the ROI part of image and then given to the classifier to classify the image is normal or abnormal. The paper is organized as follows: Section 2 presents the flow of the method, preprocessing phase, segmentation, feature extraction and classifier. Section 3 shows the implementation and result.

2. Materials and Methods

2.1 Preprocessing

The main goal of the preprocessing is to improve the image quality to make it ready to further processing by removing or reducing the unrelated and surplus parts in the background of the mammogram image. Mammograms are medical images that complicated to interpret. To remove the noise and unwanted data median filter is used. Median filter preserves edges while removing noise. Image processing operations implemented with filtering include smoothing, sharpening, and edge enhancement. In image processing filters are mainly used to suppress either the high frequencies in the image, *i.e.* smoothing the image, or the low frequencies, *i.e.* enhancing or detecting edges in the image. Filtering is a neighborhood operation, *in* which the value of any given pixel in the output image is determined by applying some algorithm to the values of the pixels in the neighborhood of the corresponding input pixel. The median filter is a sliding-window spatial filter. It replaces the value of the center pixel with the median of the intensity values in the neighborhood of that pixel. A median filter is more effective than convolution when the goal is to simultaneously reduce noise and preserve edges.

Pectoral muscle removal operation is important in mediolateral oblique view (MLO), where the pectoral muscle, slightly brighter compared to the rest of the breast tissue, can appear in the mammogram. In properly imaged MLO mam-

Licensed Under Creative Commons Attribution CC BY

mograms, the pectoral muscle is visible as a triangular region of high-density at the upper posterior part of the image. Texture of the pectoral muscle may also be similar to some abnormalities and may cause false positives in the detection of suspicious masses. Pectoral muscles are the regions in mammograms that contain brightest pixels. These regions must be removed before detecting the tumor cells so that mass detection can be done efficiently.

2.2 Segmentation

Segmentation is the process of partitioning a digital image into multiple regions. For this purpose the entropy segmentation is used to segment the image. The following diagram shows the flow of entropy segmentation.

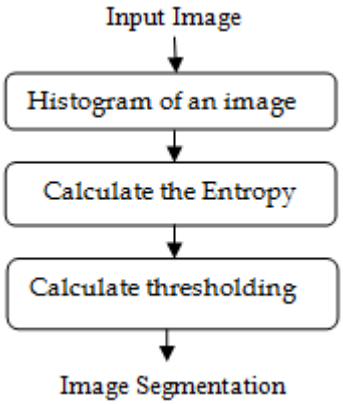

Figure 1: Segmentation algorithm

Histograms show the distribution of data values across a data range. They do this by dividing the data range into a certain number of intervals, tabulating the number of values that fall into each interval and plotting the values in the bins using bars or wedges of varying height. Entropy is a concept of information theory. It is used to measure the amount of information. It is defined in terms of the probabilistic behavior of a source of information. Entropy can best represent the information containing in the image. The approach of image segmentation based on entropy algorithm is used to segment foreground and background image. Suppose p={p1,p2,,,pn}be a finite discrete pro-bability distribution that satisfies these conditions

p(t)>=0 ,where t=0,1,2...n

The amount of uncertainty of the distribution, is called the entropy of the distribution, *P*. The Shannon entropy of the distribution, *P*, a measure of uncertainty and denoted by E(P) , is defined as

$$\mathrm{E(P)} = -\sum_{t=1}^{n} P_t \log_2 P_t \quad (1)$$

Additive entropy is

$$\mathrm{E(E1+E2)=E1(t)+E2(t)} \quad (2)$$

Segmentation involves separating an image into regions (or their contours) corresponding to objects[1]. Usually try to segment regions by identifying common properties. Or, similarly, we identify contours by identifying differenc between regions (edges).

The simplest property that pixels in a region can share is intensity. So, a natural way to segment such regions is through thresholding, the separation of light and dark regions. Thresholding creates binary images from grey-level ones by turning all pixels below some threshold to zero and all pixels about that threshold to one.

Thresholding method in image segmentation that yields yields all the pixels and assumes the algorithm in two cases i.e. darkness and brightness.

2.3 Feature Extraction

After the segmentation part, the features are extracted from the segmented images which are used for the classification. Then by extracting the feature from segmented image the classifier gives output normal or abnormal image. There are various features like mean, standard deviation, Entropy, Skewness, Kurtosis, Variance, Energy, Correlation, Smoothness, and Root Mean Square. These features are calculated by following formulas: There are various features like mean, standard deviation, Entropy, Skewness, Kurtosis, Variance, Energy, Correlation, Smoothness and Root mean square (rms). These features are calculated by following formulas:

Mean-

The mean, μ of the pixel values in the defined window, estimates the value in the image in which central clustering occurs. The mean can be calculated using the formula:

$$\mu = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N} p(\mathrm{i,j}) \quad (3)$$

Where *p(i,j)* is the pixel value at point *(i,j)* of an image of size MxN.

Standard deviation-

The Standard Deviation, σ is the estimate of the mean square deviation of grey pixel value p (i, j) from its mean value m. Standard deviation describes the dispersion within a local region

$$\sigma = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(p(\mathrm{i,j})-\mu)^2} \quad (4)$$

Entropy-

Entropy, *h* can also be used to describe the distribution variation in a region. Overall entropy of the image can be calculated as:

Entropy is defined as

$$h = -\sum_{k=0}^{L-1} \mathrm{Pr}_k(\log \mathrm{Pr}_k) \quad (5)$$

Where, Pr_k is the probability of the k^{th} grey level, which can be calculated as $Z_k / M*N$, Z_k is the total number of pixels with the k^{th} grey level and L is the total number of grey levels.

Skewness-

Skewness, S characterizes the degree of asymmetry of a pixel distribution in the specified window around its mean. Skewness is a pure number that characterizes only the shape of the distribution. The formula for finding Skewness is given in the below equation:

Licensed Under Creative Commons Attribution CC BY

International Journal of Scientific Engineering and Research (IJSER)
www.ijser.in
ISSN (Online): 2347-3878
Volume 3 Issue 2, February 2015

$$s = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(\frac{g(i,j)-\mu}{\sigma}\right)^3 \quad (6)$$

Kurtosis-

Kurtosis, K measures the Peakness or flatness of a distribution relative to a normal distribution. The conventional definition of kurtosis is:

$$K = \left\{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(\frac{g(i,j)-\mu}{\sigma}\right)^4\right\} - 3 \quad (7)$$

Variance-

Variance is the square root of standard deviation or it is the average of the squared differences from the Mean.

$$\text{variance} = \sqrt{SD} \quad (8)$$

Energy-

Energy returns the sum of squared elements in the Grey Level Co-Occurrence Matrix (GLCM). Energy is also known as uniformity. The range of energy is [0 1].

Mathematical equation,

$$E = \sum_i \sum_j g(i,j)^2 \quad (9)$$

Correlation-

Correlation returns a measure of how correlated a pixel is to its neighbor over the whole image. The range of correlation is [-1 1]. Correlation is 1 or -1 for a perfectly positively or negatively correlated image. Correlation is NaN (Not a Number) for a constant image. The below equation shows the calculation of Correlation. Evaluates how a pixel is related to its neighbor.

$$correlation = \sum_{i,j}\frac{(i-\mu_i)(j-\mu_j)P(i,j)}{\sigma_i\sigma_j} \quad (10)$$

Smoothness-

Relative smoothness, R is a measure of grey level contrast that can be used to establish descriptors of relative smoothness. The smoothness is determined using the formula:

$$R = 1 - \frac{1}{1+\sigma^2} \quad (11)$$

Root Mean Square-

The RMS (Root Mean Square) computes the RMS value of each row or column of the input, along vectors of a specified dimension of the input, or of the entire input. The RMS value of the jth column of an M-by-N input matrix u is given by below equation:

$$y = \sqrt{\frac{\sum_{i=1}^{M}|u_{ij}|}{M}} \quad (12)$$

2.4 Bayesian Network

Bayesian networks are a statistical method for Data Mining, a statistical method for discovering valid, novel and potentially useful patterns in data. Bayesian networks are used to represent essential information in databases in a network structure. The network consists of edges and vertices, where the vertices are events and the edges relations between events. The networks can be used to represent domain knowledge, and it is possible to control inference[11]. A simple usage of Bayesian networks is denoted naive Bayesian classification.

$$p(x_1,\ldots x_n) = \prod_{i=1}^{n} P(x_i \mid c_j) \quad (13)$$

The aim of supervised classification is to classify instances i given by certain characteristics xi={xi1,…,xin} into r class labels, ci , i=1,…r. xil denotes the value of variable xl observed in instance i. The main principle of a Bayesian classifier is the application of Bayes' theorem.

3. Experimental Result

In this section, the results of the proposed approach are presented. First the preprocessing is done by the median filter and also pectoral muscles are removed. Then this image is segmented by using the entropy segmentation method, features are extracted from segmented image and the output result is shown with the help of Bayesian network. The Bayesian network shows the image is abnormal image.

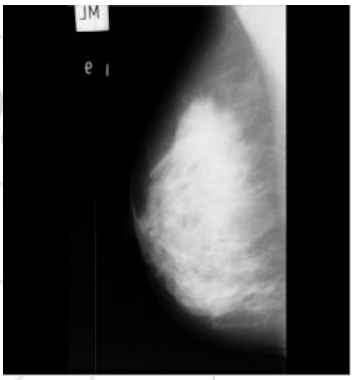

Figure 2: Original image

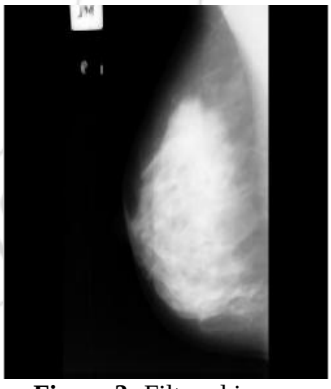

Figure 3: Filtered image

Figure 4: Pectoral muscles detected

Licensed Under Creative Commons Attribution CC BY

Figure 5: Image after removing pectoral muscle

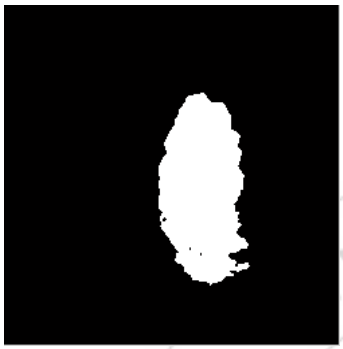

Figure 6: Segmented image

4. Conclusion

Although In this study, an automatic diagnosis system to detect the brast cancer by using Bayesian neural network is presented. In this study by using several preprocessing techniques the unwanted part is removed in the image. Using Entropy segmentation it shows the tumor area in the image. Then the Bayesian classifier is used to show the image result on the basis of features thet are extracted from the image. The accuracy of this method is high upto certain extent for the MIAS database. The results of this method can be improved by taking the combination of Bayesian network with another classifier.

References

[1] R. Ramani, Dr. N.Suthanthira Vanitha, S. Valarmathy. "The Pre-Processing Techniques for Breast Cancer Detection in Mammography Images", *I.J. Image, Graphics and Signal Processing*, 2013, 5, pp- 47-54.
[2] R. C. Gonzalez. "Digital Image processing using Matlab" Pearson publication, 2005.
[3] H. Abdellatif, t. E. Taha, o. F. Zahran, w. Al-nauimy, f. E. Abd el-samie, "k9. Automatic segmentation of digital mammograms to detect masses," 30th national radio science conference, 2013, pp-557-565.
[4] R. Subash Chandra Boss, K. Thangavel, D. Arul Pon Daniel, "Automatic Mammogram image Breast Region Extraction and Removal of Pectoral Muscle".
[5] Chengxin Yan a, Nong Sang a, Tianxu Zhang, "Local entropy-based transition region extraction and thresholding," *Pattern Recognition Letters*, 24 (2003), pp 2935–2941.
[6] Amar Partap Singh Pharwaha, Baljit Singh, "Shannon and Non-Shannon Measures of Entropy for Statistical Texture Feature Extraction in Digitized Mammograms," Proceedings of the World Congress on Engineering and Computer Science 2009 Vol II., October 20-22, 2009, ISBN: 978-988-18210-2-7.
[7] Samy Sadek, Sayed Abdel-Khalek "Generalized α-Entropy Based Medical Image Segmentation" *Journal of Software Engineering and Applications*, 2014, 7, pp- 62-67.
[8] 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.
[9] W.R. Silva, D. Menotti, arwaha, "Classification of Mammograms by the Breast Composition," The 2012 International Conference on Image Processing, Computer Vision, and Pattern Recognition.
[10] Amir Fallahi, Shahram Jafari, "An Expert System for Detection of Breast Cancer Using Data Preprocessing and Bayesian Network," *International Journal of Advanced Science and Technology*, Vol. 34, September, 2011, pp 65-70.
[11] S.Kharya1, S.Agrawal2, S. Soni, "Using Bayesian Belief Networks for Prognosis & Diagnosis of Breast Cancer", *International Journal of Advanced Research in Computer and Communication Engineering*, Vol. 3, Issue 2, February 2014, pp-5423-5427.
[12] Alok Sharma and Kuldip K. Paliwal, "A Gene Selection Algorithm using Bayesian Classification Approach", *American Journal of Applied Sciences*, 9 (1), pp- 127-131

Licensed Under Creative Commons Attribution CC BY