International journal of imaging science and engineering

ISSN: 1934--9955 www.ijise.net

Vol-09 Issue-01 May 2014

A Bayesian Network-Based Approach to Digital Mammography Lesion Classification

Lavanya Germany

Article Info

Received: 17-02-2014 *Revised: 02 -03-2014* *Accepted: 06-04-2014* *Published:17/05/2014*

Abstract : Worldwide, breast cancer is a major concern for women. When cancer cells begin to grow in the breast, the resulting disease is known as breast cancer. Various screening methods are accessible for the purpose of detecting breast cancer. This publication presents many techniques for the detection of breast cancer. The first stage of this work was eliminating background noise from the images. To do this, the authors used a median filter. When searching the MIAS database, you may find images with pectoral muscles; to eliminate them, just calculate the image's thresholding value.We suggest using an entropy-based segmentation method to divide up a grayscale breast picture. This method determines the entropy value and histogram of a picture. The output displays the segmented picture after determining the image's thresholding value. An entropic technique for noisy cell picture segmentation is proposed in this research, and it is efficient and quick. Following the segmentation process, a variety of characteristics are extracted from the picture, including standard deviation, entropy, skewness, kurtosis, variance, energy, correlation, smoothness, and root mean square. These features are then fed into a Bayesian network, which uses them to categorize the image based on their values. In comparison to previous methods, the experimental findings demonstrate that the suggested approach is both efficient and noise-tolerant.

Keywords: Median filter, entropy, ROI, pectoral muscle, breast cancer, Bayesian network

1. **Introduction**
2. Among females across the globe, breast cancer is by far the most frequent kind of breast cancer. Cancers that originate in the breast cells are known as breast cancers. Worldwide, breast cancer accounts for over 18.2% of all cancer-related fatalities, regardless of gender. In industrialized countries, breast cancer is a much more significant problem than in less developed ones. Breast cancer is more frequent in older women for a number of reasons; one is that women in wealthy countries tend to live longer than those in poorer ones. One reason for this is that women in wealthy nations

International journal of imaging science and engineering
ISSN: 1934--9955 www.ijise.net
Vol-09 Issue-01 May 2014

tend to eat and live differently than women in disadvantaged ones. Annually, there are 232,380 cases of breast cancer in women and 2,240 cases in men in the United States, as reported by the National Cancer Institute. The World Health Organization (WHO) reports that seven lakh people in India lose their lives to cancer each year, with over ten lakh receiving a cancer diagnosis.

The origin of breast cancer is from the inner lining of milk ducts or the lobules that supply them with milk. Malignant tu- mor 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 inva- sive 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, progres- sively more and more than there should be. However, it is diffi- cult 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 mus- cles 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 pa- per 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.

3. Materials and Methods

3.1 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 un- wanted data median filter is used. Median filter preserves edges while removing noise. Image processing operations implemented with filtering include smoothing,

International journal of imaging science and engineering
ISSN: 1934--9955 www.ijise.net
Vol-09 Issue-01 May 2014

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 val- ues of the pixels in the neighborhood of the corresponding input pixel. The median filter is a sliding-window spatial fil- ter. It replaces the value of the center pixel with the median of the intensity values in the neighborhood of that pixel. A me- dian filter is more effective than convolution when the goal is to simultaneously reduce noise and preserve edges.

Pectoral muscle removal operation is important in medio- lateral 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- mograms, the pectoral muscle is visible as a triangular region of high-density at the upper posterior part of the image. Tex- ture of the pectoral muscle may also be similar to some ab- normalities and may cause false positives in the detection of suspicious masses. Pectoral muscles are the regions in mam- mograms that contain brightest pixels. These regions must be removed before detecting the tumor cells so that mass detec- tion can be done efficiently.

3.2 Segmentation

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

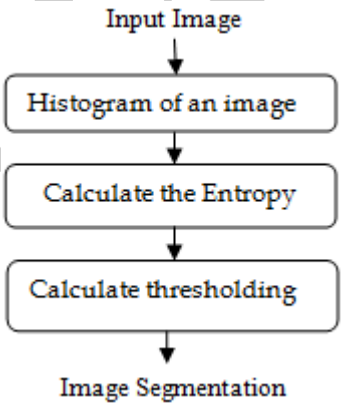

Figure 1: Segmentation algorithm

International journal of imaging science and engineering
ISSN: 1934--9955 www.ijise.net
Vol-09 Issue-01 May 2014

The simplest property that pixels in a region can share is in- tensity. So, a natural way to segment such regions is through thresholding, the separation of light and dark regions. Thre- sholding creates binary images from grey-level ones by turn- ing 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.

3.3 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, Smooth- ness, and Root Mean Square. These features are calculated by follo wing 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, esti- mates the value in the image in which central clustering occurs. The mean can be calculated using the formula:

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 in- formation. 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

(3)

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

International journal of imaging science and engineering
ISSN: 1934--9955 www.ijise.net
Vol-09 Issue-01 May 2014

in- formation. 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}b*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
a finite discrete pro-bability distribution that satisfies these conditions

p(t)>=0 ,where t=0,1,2...n
(4)

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)

$$h = -\sum \Pr_k (\log \Pr_k)$$

Segmentation involves separating an image into regions (or their contours) corresponding to objects[1]. Usually try to segment regions by identifying common properties. Or, simi- larly, we identify contours by identifying differenc between regions (edges).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 dis- tribution. The formula for finding Skewness is given in the below equation:

$$p(x_1,.... x_n) = \prod P(x_i \quad c_j)$$

The aim of supervised classification is to classify instances i given by certain characteristics xi={xi1,…,xin} into r class
Kurtosis, K measures the Peakness or flatness of a distribu- tion relative to a normal distribution. The conventional defini- tion of kurtosis is:labels, ci , i=1,…r. xil denotes the value of variable xl ob- served in instance i. The main principle of a Bayesian clas- sifier is the application of Bayes' theorem.

International journal of imaging science and engineering
ISSN: 1934--9955 www.ijise.net
Vol-09 Issue-01 May 2014

In this section, the results of the proposed approach are pre-

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

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,

sented. 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, fea- tures are extracted from segmented image and the output re- sult is shown with the help of Bayesian network. The Baye- sian network shows the image is abnormal image.

Correlation-

(9)

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 nega- tively 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 neigh- bor.

(12)

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.

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 tech- niques the unwanted part is removed in the image. Using En- tropy segmentation it shows the tumor area in the image. Then the Bayesian

ISSN: 1934--9955 www.ijise.net

Vol-09 Issue-01 May 2014

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] "The Pre-Processing Techniques for Breast Cancer De-tection in Mammography Images" (I.J. Image, Graphics and Signal Processing, 2013, 5, pp-47-54), written by R. Ramani, Dr. N. Suthanthira Vanitha, and S. Valarmathy.
"Digital Image processing using Mat-lab," published by Pearson in 2005, was written by R. C. Gonzalez.
In their 2013 paper "k9. Automatic segmentation of digital mammograms to detect masses," H. Abdellatif, t. E. Taha, o. F. Zahran, w. Al-nauimy, and f. E. Abd el-samie presented their findings at the 30th national radio science conference. The paper was pp. 557-565.
[4] "Automatic Mammography image Breast Region Extraction and Removal of Pectoral Muscle" (R. Subash Chandra Boss, K. Thangavel, and D. Arul Pon Daniel's work).
The paper "Local entropy-based transition region extraction and threshold-ing" was published in the Pattern Recognition Letters journal in 2003 and can be found on pages 2935–2941.
[6] "Statistical Texture Feature Extraction in Digitized Mammograms using Shannon and Non-Shannon Measures of Entropy," Amar Partap Singh Pharwaha and Baljit Singh,
Volume II of the 2009 World Congress on Engineering and Computer Science, held October 20-22, 2009, with the ISBN 978-988-18210-2-7, is cited accordingly.
Samuel Sadek and Sayed Abdel-Khalek "Medical Image Segmentation Based on Generalized α-Entropy," published in the Journal of Software Engineering and Applications in 2014, volume 7, pages 62–67.
"Feature extraction of mammograms," in Pradeep N., Girisha H., Sreepathi B., and Karibasappa K.'s 2012 publication in the International Journal of Bioinformatics Research, volume 4, number 1, pages 241–244.
10. "Classification of Mammograms by the Breast Composition," presented at the 2012 International Conference on Image Processing, Computer Vision, and Pattern Recognition by W.R. Silva, D. Menotti, and arwaha.
[11]Published in the International Journal of Advanced Science and Technology, "An Expert System for Detection of Breast Cancer Using Data Preprocessing and Bayesian Network," Amir Fallahi and Shahram Jafari's work may be found on pages 65–70 in the September 2011 issue.

International journal of imaging science and engineering

ISSN: 1934--9955 www.ijise.net

Vol-09 Issue-01 May 2014

[12] In the International Journal of Advanced Research in Computer and Communication Engineering, S. Kharya1, S. Agrawal2, and S. Soni published an article titled "Using Bayesian Belief Networks for Prognosis & Diagnosis of Breast Cancer" in February 2014, pages 5423–5427, Volume 3, Issue 2.

International journal of imaging science and engineering

ISSN: 1934--9955 www.ijise.net

Vol-09 Issue-01 May 2014

ISSN: 1934--9955 www.ijise.net
Vol-09 Issue-01 May 2014

ISSN: 1934--9955 www.ijise.net

Vol-09 Issue-01 May 2014

5. Experimental Result

[1]