Classification of Breast Cancer Tumor using Statistical Features

Kalyani Ohri
Department of ECE
CTITR
Jalandhar, India
kalyaniohri@yahoo.in

Harsukhpreet Singh
Department of ECE
CTITR
Jalandhar, India
harsukhpreet@gmail.com

Anurag Sharma
Department of ECE
CTITR
Jalandhar, India
er. anurags@gmail.com

Abstract—Breast Cancer is the most far-reaching cause of deaths among women. The most widespread technique used for diagnosis of breast cancer is Mammography. Malignant and benign masses had the different shape, size according to Breast Imaging Reporting and Data System (BIRADS), experts classify mammogram according to BIRADS. Determination of abnormalities is a very difficult task. In this paper, classification of masses is done using gray weighted function related to HSV color space based on statistical features. The mammogram images had been obtained from Digital Database for Screening Mammogram (DDSM) is used for conducting experiment.

Index Terms— Breast Cancer, Mammogram masses, Weight function, DDSM (Digital Database for Screening Mammogram).

I. INTRODUCTION

The Breast Cancer is thought to be one of most far-reaching causes of death among women and second highest cause of deaths among humans. It has been predicted that each single minute a woman is interpreted with breast cancer and every 13 minutes a woman expires due to disease. Many techniques and methods had been introduced for early detection and diagnosing of breast cancer such as Biopsy, Mammography, Fine Needle Aspiration, Magnetic Resonance Imaging but the mammography is the most widespread method for early prognosis of disease [1]. The objective of mammography is an early prognosis of breast cancer, especially through espionage of masses and micro calcification [2]. According to the report by Breast Imaging Reporting and Data System (BIRADS), malignant masses are irregular in shape whereas benign masses are oval or round. Medical experts classified masses as malignant or benign during diagnosis stage [1]. Medical experts read mammograms and try to identify the abnormalities present in the breast mass. It has been predicted that 10-30% of women interpret with breast cancer have false negative mammograms. Almost false negative mammograms are contributed to physicians radiologist failure to identify the breast cancer at an early stage due to lack of knowledge or experience or misinterpretation. According to the medical point of view, reading and interpretation mammograms are the very complex thing [2]. In recent years, computer-aided diagnosis methods are used to guide physicians in early diagnosis and detection of breast cancer. Computer aided diagnostic methods guided to speed up the diagnostic process proposed for various diseases such as diagnosis of various cancers classification, inclusive of breast and lung cancer [3]. Computer aided diagnostic tools uses computer technologies to determine the abnormalities present in the mammogram such as masses, calcification and with the help of these results medical experts can diagnose and detect breast cancer at an early stage[4]. Computer-aided techniques are used for pre-processing and evaluation of images as supporting or second reading. The first stage of the computer aided diagnostic tool is to capability to determine abnormalities in breast and the second stage is to diagnose the abnormalities found in masses identified in first stage. The most important step before implementing two stages of computer-aided diagnostic tool is pre-processing stage has to take place that is segmentation of breast part from the whole background. The evaluations of mammogram images can assist radiologists in early diagnosis and detection of breast cancer tumor and diagnose the treatment very effectively [2]. As many methods had been used by various researchers for classifying malignant and benign masses, but only a few methods are capable of classifying breast mass region into malignant and benign masses. This can be attained by proposing some constraints as by using a number of different features. Although by using a large number of features does not make reasonable classification rate but also make the breast cancer tumor classification more complex. The shape characteristics of masses have been introduced for classifying breast masses as malignant mass or benign mass with the use of shape properties. Most of the recent works done on mammography are almost based on mammography’s histogram and it has been found that histogram based on the mammograms are not much effective for classifying the breast masses. This is due to the fact that histograms based on mammograms pattern changes mainly due to noise and over-intensification of mammogram images [1].

II. RELATED WORK

In this part, we shall discuss some work related to the classification of masses based on statistical features. There are various works done in the literature that illustrate the classification of benign breast masses and malignant breast masses.

Tang J. et al. (2009) reviewed the recent changes made in the development of computer aided diagnostic tools and their related methods. Early detection of breast cancer is very
necessary to prevent millions of deaths occurring annually. Computer-aided diagnostic techniques play a very important role in detecting the breast cancer at an early stage and helps to decrease the death rate among women suffering from breast cancer. Authors concluded that CAD techniques reduces the burden of radiologists and can assist the physicians to detect and manage the breast cancer at an early stage [4]. Surendiran B. (2011) proposed Hue Saturation Value (HSV) weight function for classifying malignant masses and benign masses based on statistical features. The HSV weight function is robust against noise and entraps the gray content of the image. To effectively differentiate between malignant masses and benign masses, the gray weight value is used instead of gray weight function. The digital database for screening mammogram (DDSM) dataset had been used in which 233 mammograms are used. Authors concluded that classification rate is highly encouraging and are related to BIRADS classification [1]. Kowal M. et al. (2013) introduced computer-aided techniques to diagnose breast cancer which depends on fine needle biopsy images. Authors introduced four different clustering algorithms K-means, Fuzzy-c-means, neural networks and Gaussian models are tested and compared relative to nuclei segmentation. These techniques are an applicable to medical decision support system for breast cancer for classification of malignant or benign. Authors concluded that achieved classification is effective with 96-100% [5]. Abaddi N. et al. (2014) proposed some statistically features to effectively differentiate between malignant breast tumor and benign breast tumor. Authors introduced some features such as area, standard deviation, entropy, variance, mean so as to statistical differentiate between breast masses. Authors concluded that acquired classification rate is very encouraging with an accuracy of 100% [2]. Khezri R. et al. (2014) introduced a fuzzy expert system for identification of breast cancer which depends on ”nuclei segmentation. These techniques are an applicable to medical decision support system for breast cancer. The major advantage of the expert systems is that it can predict the risk of breast cancer even in healthy women also [3].

III. SYSTEM MODELING

This part explains the method that was adapted in implementing the overall framework for the classification of breast masses. The mammogram’s content can be accomplished by using four levels of gray. While describing the masses in mammogram they are enclosed with pixels with slight fluctuations in gray levels and configure smooth boundaries. Thus, it is very necessary to catch the boundary representation of the pixel values for segmentation of breast masses. The approach is conducted on spatial domain image processing with the use of HSV colour space properties and this color space is early relevant to human perception colors and for every pixel a weighted value is determined using a gray weighted function which seizes the gray part of the pixel and is very vigorous against noise and other noise parameters.

It is given as:

\[ W_{gray}(S,I) = 1 - \frac{r_1}{255} \] 

The extensity of \[ W_{gray}(S,I) \] is (0-1) and it evaluates the gray portion of the pixel by using together the intensity and saturation values. Thus in the case of R=G=B, the value of the saturation of pixel is zero and gray weighted value will also be zero regardless of any intensity values. So in order to seize the gray content of the pixel, closely agitate the either of the values of R, G or B, which effects the value of saturation of pixel. The gray weighted function is adjustable with saturation values and is found to be stable. It is very clear from equation1 that r1 must take value nearly higher than zero and r2 must hold value lesser than 1 for having stable and continuous gray weighted values. From recent works, it had been predicted that r1=0.1 and r2= 0.85.

![Figure 1: Variation of Partial Derivative of Wgray(S, I) with different values of Intensity](image)

For every pixel of concerned mass, a gray weighted value is determined and unwanted pixels are eliminated using thresholding approach as a pre-processing step. In gray thresholding, the average value of gray is determined as \[ Gray_{\text{threshold}} \] and the values of pixels lower than \[ Gray_{\text{threshold}} \] are removed in the pre-processing step. The pre-processing step is considered to be the necessary step as it eliminates the unwanted noise from the mass. The objective of applying gray thresholding is to produce an accurate mass which can be further used for processing. The figure below shows the images of mammogram after average thresholding.

![Figure 2(a) Extracted Mass](image)  ![Figure 2(b) After Average Thresholding](image)

<table>
<thead>
<tr>
<th>TABLE 1. STATISTICAL FEATURES OF EXTRACTED MASS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Features</td>
</tr>
<tr>
<td>Area</td>
</tr>
<tr>
<td>Standard Deviation</td>
</tr>
</tbody>
</table>
The various features which are used are shown in the table. The statistical feature mean is clearly used to differentiate between malignant masses and benign masses. As it is clear from the figure the range of mean for malignant masses is from 0.1 to 0.5 and for benign masses is from 0.4 to 0.8. Mean is considered to be best feature for discrimination of malignant and benign masses. In figure 4 standard deviation for malignant and benign masses is shown. Standard deviation is used to determine the deviation of weighted gray values of pixels present in the mass. Figure 5 displays the variance for malignant and benign breast masses. From the figures it is predicted that statistical features related to gray weighted function such as area, mean, standard deviation, variance and entropy are used very much effectively for classification of masses.

Figure 3 displays the statistical measure mean for malignant masses and benign masses. Mean is considered to best statistical for classification of benign masses and malignant masses. As it is clear from the figure 3, that if the mean of the mass is less than equal to 0.45, then the breast mass is classified as malignant mass and if the mean of the mass is greater than equal to 0.52, then the breast mass is classified as benign mass. Therefore, by using the mean of the masses it is much easy to classify breast masses as benign masses or malignant masses.

Figure 4 displays the statistical measure standard deviation for benign and malignant masses. Standard deviation is the deviation of weighted gray values of pixels of the mass. As it is observed from the figure 4, if the standard deviation of mass is greater than equal to 0.45, then mass is classified as malignant mass. If the standard deviation of the mass is less than equal to 0.49, then the mass is classified as benign mass. By using measure such as standard deviation, classification rate is good so it can help experts during diagnosis stage.

The figure 5 displays variance for malignant masses and benign masses. The statistical feature variance is used where the variations in pixel belong to the same particular class. As it is clear from the figure 5, if the variance of mass is less than equal to 0.25, then the mass is classified as malignant mass. If the variance of mass is greater than equal to 0.12, then mass is classified as benign mass. Variance is thought to be best measure in classification if variation in pixels is too large.

The figure 6 displays entropy for malignant masses and benign masses. Entropy is a statistical feature that is used to describe the texture of input image. As it is clear from the figure 6, if the entropy of the image is greater than equal to 0.96, then the mass is classified as malignant mass. If the entropy of the mass is less than equal to 0.97, then the mass is classified as benign mass.

The images of mammogram that has been used are obtained from Digital Database for Screening Mammography (DDSM) for classification of breast masses. The experiment was conducted by using different images of the mammogram. For classification of masses, an experiment was conducted on so many samples of mammogram out of which some are

---

**Mean**

\[ \frac{\sum_{a=1}^{M} \sum_{b=1}^{N} I(a, b)}{M \times N} \]

**Entropy**

\[ \frac{1}{M \times N} \sum_{a=1}^{M} \sum_{b=1}^{N} I(a, b) \ln I(a, b) \]

**Variance**

\[ \frac{\sum_{a=1}^{M} \sum_{b=1}^{N} [I(a, b) - \mu]^2}{M \times N} \]
malignant masses and some are benign masses. The statistical measures such as mean, area, standard deviation, variance and entropy of three mammogram samples are shown below

### Table 2: Classification of Breast Masses

<table>
<thead>
<tr>
<th>Images</th>
<th>Standard Deviation</th>
<th>Area</th>
<th>Variance</th>
<th>Mean</th>
<th>Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M1</td>
<td>0.5000</td>
<td>3.965e+04</td>
<td>0.2500</td>
<td>0.4999</td>
<td>1.0023</td>
</tr>
<tr>
<td>M2</td>
<td>0.4927</td>
<td>2.446e+04</td>
<td>0.24278</td>
<td>0.4150</td>
<td>0.9790</td>
</tr>
<tr>
<td>M3</td>
<td>0.4889</td>
<td>2.331e+04</td>
<td>0.2390</td>
<td>0.3953</td>
<td>0.9681</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B1</td>
<td>0.3680</td>
<td>1.310e+04</td>
<td>0.1354</td>
<td>0.5227</td>
<td>0.6379</td>
</tr>
<tr>
<td>B2</td>
<td>0.4906</td>
<td>1.702e+04</td>
<td>0.2407</td>
<td>0.5517</td>
<td>0.9730</td>
</tr>
<tr>
<td>B3</td>
<td>0.4641</td>
<td>4.951e+04</td>
<td>0.2154</td>
<td>0.6858</td>
<td>0.8978</td>
</tr>
</tbody>
</table>

### Table 3: Statistical Analysis of Breast Masses

<table>
<thead>
<tr>
<th>Statistical Feature</th>
<th>Malignant Tumor</th>
<th>Benign Tumor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>If mean &lt;= 0.49, then its malignant tumor</td>
<td>If mean &gt;0.52, then its benign tumor</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>If Standard Deviation=0.45,then its malignant tumor</td>
<td>If standard deviations&lt;=0.49,then its benign tumor</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variance</td>
<td>If Variance=0.25, then its malignant tumor</td>
<td>If Variance=0.12,then its benign tumor</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entropy</td>
<td>If Entropy =&gt;0.90,then its malignant tumor</td>
<td>If Entropy&lt;=0.97,then its benign tumor</td>
</tr>
</tbody>
</table>

### V. CONCLUSION

The classification of malignant and benign masses is considered to be a very complex task. A gray weighted function related to HSV color space had been used for determining statistical measures like mean, area, standard deviation, variance and entropy. It has been found that statistical features had been effectively used for classification of malignant and benign masses. The classification rates are found to be very interesting. In future work, these statistical features can be used in the fuzzy expert system by using fuzzy membership functions for classification of breast masses.

### REFERENCES

   http://www.nationalcancer.org/breastcancer.