Automatic Classification of Breast Masses for Diagnosis of Breast Cancer in Digital Mammograms using Neural Network

Hitiksha Shah
PG Student
Department of Biomedical Engineering
Government Engineering College, Gandhinagar, India

Abstract

This paper presents a computer aided diagnosis (CAD) system for automatic classification of breast masses in digital mammograms. Initially, Digital mammogram is pre-processed by 2D-median filter, connected component labelling method, and morphological functions for breast extraction. Wavelet transform is used for enhancement of mammogram and triangular mask is used for pectoral muscle suppression. Morphological functions like opening, closing, erosion, dilation and reconstruction are used for the segmentation of mammogram to extract region of interest (ROI). From ROI, intensity histogram based texture features are extracted. Extracted features are fed into classifier algorithm. In this proposed work, concept of neural network is used for classification, which is applied for two levels. In the first level, neural network classify the segmented ROI into normal (without tumor) and abnormal (with tumor) ROI. Second level neural network classify abnormal ROI into malignant and benign masses. The proposed CAD system achieves 96.07% specificity and 94.73% sensitivity at first level classification, 91.66% specificity and 80% sensitivity at second level.

Keywords: Digital mammogram, Median filter, neural network, Texture feature, Wavelet transform

I. INTRODUCTION

Breast cancer with its increasing incidence rate and its carcinogenic cause still being unknown has been an area of concern for years. It is reported to be second most common type cancer in the population. According to recent statistics the ratio between the incidence and mortality in us is 4.5:1 (meaning about 1 death for 4.5 new cases detected) and in India it is 2:1 which is alarming. The world health organization stated that more than 1.2 million women were found with breast cancer and more than 700,000 women lost their life every year in the world [1]. Early detection is the key for improving breast cancer diagnosis. Various methods have been proposed for early detection and screening of breast cancers and the mammography is being considered as one of the most effective method [3]. One of the difficulties with mammography is that mammograms generally have low contrast. This makes it difficult for radiologists to interpret the results. Mammography is susceptible to a high rate of false positives as well as false negatives, causing a high proportion of women without cancer to undergo further clinical evaluation or breast biopsy, or miss the best time interval for the treatment of cancer [7].

In breast cancer diagnosis, the radiologists mainly use their eyes to discern cancer when they Screen the mammograms. However, in many cases, cancer is not easily detected by the eyes because of the bad imaging conditions. The distinction between malignant and benign tumours sometimes is not easy because the characteristics of areas of abnormality are significant and complex [4]. However the appearance of breast cancer is very subtle and unstable in their early stages, therefore doctors and radiologists can miss the abnormality easily if they only diagnose by experiences [2].So that the performance of the radiologists varies from 65% to 85%. Due to the above mentioned regions a variety of computer assisted detection techniques have been proposed. In order to improve the accuracy of interpretation CAD involves two major process computer aided detection (CADe) and Computer Aided Diagnosis (CADi) [5]. CADe is able to identify the Regions of Suspicion (ROS), but CADi can make a decision whether a ROS is benign or malignant. The general process of CAD for mammograms refers to image pre-processing, defining ROS, extracting features and classifying a ROS into benign, malignant or normal [6]. Since 65–90% of the biopsies of suspected cancers turned out to be benign, it is very important to develop CADs that can distinguish benign and malignant lesions. The combination of CAD scheme and experts’ knowledge would greatly improve the detection accuracy. The detection sensitivity without CAD is 80% and with CAD up to 90% [8].

In this paper, CAD system is developed for automatic diagnosis of breast cancer in digital mammogram by classifying the breast masses. The proposed CAD system consists of five major stages: pre-processing, segmentation, feature extraction, classification and evaluation. The database for the study is taken from mammographic image analysis society (MIAS). The database comprises a set of 322 digitized mammograms along with ground truth information.
II. LITERATURE REVIEW

Various algorithms have been proposed for preprocessing, segmentation and classification of the breast masses in digital mammogram. Jawed Nagi et al. 2010 [9] proposed algorithm for an automated technique for mammogram segmentation. The algorithm uses morphological pre-processing and seeded region growing (SRG) to remove digitization noises, suppress radiopaque artifacts and remove the pectoral muscle to accentuate the breast profile region for use in CAD algorithms.

Mohammed J. Islam et al. 2010 [10] proposed ANN-based classifier with three layers for mass classification and statistical and texture features are used as an input to the classifier. It was evaluated on 69 images containing malignant and benign masses with different size, shape and contrast. Author achieved 90.91% sensitivity and 83.87% specificity, which is very much encouraging compare to the radiologist's sensitivity 75%.

R.Nithya, B.Sanithi. 2011 [11] proposed CAD (Computer Aided Diagnosis) system using GLCM feature and neural network. Authors achieved the maximum accuracy rate for normal and cancer classification is 96%.

Aswini Kumar Mohanty et al. 2011 [12] reviewed different methods for mammogram enhancement, segmentation, features extraction and classifications. They explored the further research directions for next generation CAD for mammograms.

K.vaidehi et al. 2013 [13] proposed a new method for pectoral muscle detection and elimination. The proposed work is done in three steps. In the first step, the mammogram is oriented to the left to minimize computations. In the second step the top left quadrant of the mammogram which contains the pectoral muscle is extracted and the pectoral muscle contour is computed using algorithm.

Pratishtha Shrivastava And K.G.Kirar 2014 [14] presented a method for detecting tumor in breast is based on edge segmentation using Canny Edge Detection Technique. It provides the strong edges as compared to other edge detectors and has overcome the problem of threshold adjustment, by which extraction of the tumor using morphological operations more easily and appropriately.

III. PROPOSED METHODOLOGY

A. Mammogram Database:

The database for the study is taken from Mammographic Image Analysis Society (MIAS). The Mammographic Image Analysis Society (MIAS) is an organization of UK research group interested in understanding of mammogram and has generated a database of digital mammogram. Each pixel is described as an 8-bit word. Every image is of size 1024 x 1024.Database comprise of 322 MLO mammograms with ground truth information for normal, benign and malignant breast masses.

B. Mammogram Preprocessing:

Mammogram Pre-processing steps include: (a) noise removal, (b) radiopaque artifact suppression, (c) image enhancement, and (d) pectoral muscle removal. Following steps are involved of proposed algorithm for pre-processing:
Automatic Classification of Breast Masses for Diagnosis of Breast Cancer in Digital Mammograms using Neural Network

2D-median filter is used to remove the digitization noise in the mammogram.

- Convert gray scale image into binary image using global thresholding.
- Using connected component labelling method find out the largest object in mammogram. So breast profile is separated and radiopaque artifact is suppressed.
- Wavelet transform is used to enhance the mammogram.
- Check the orientation of mammogram. If mammogram is left oriented flip to right.
- Divide image into four quadrant and upper left quadrant is processed.
- Triangular mask is applied to upper left quadrant and pectoral muscle is suppressed. Resulting image is subtracted from the original image to get mammogram without pectoral muscle.

C. Mammogram Segmentation:

The second stage of mass detection CAD system is to separate the suspicious regions that may contain masses from the background parenchyma, then extract regions of interests (ROIs), and locate the suspicious mass candidates from ROIs. The suspicious area is an area that is brighter than its surroundings, has almost uniform density, has a regular shape with varying size, and has fuzzy boundaries. This is a very essential and important step that determines the sensitivity of the entire system. Following steps are involved of proposed algorithm for segmentation:

- Create a morphological disk shaped structuring element with radius=25 using strel function. Open image by structuring element and perform erosion. Close image by structuring element
- Dilate the gray scale image. Complement the image.
- Find the regional maxima of image(binary image)
- Get gray scale image by multiply the binary image with transformed image.

D. Feature Extraction:

The third stage of mass detection by CAD system is the feature extraction. Texture features using intensity histograms are extracted from ROI. Table 1 provides explanation and equation of these features. Where $z$ is a random variable indicating intensity, $p(z)$ is the histogram of the intensity levels in a region, $L$ is the number of possible intensity levels.

Table – 1:

<table>
<thead>
<tr>
<th>Moment</th>
<th>Expression</th>
<th>Measure of texture</th>
<th>Texture features using intensity histogram [15]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>$m = \sum_{i=0}^{L-1} z_i p(z_i)$</td>
<td>A measure of average intensity</td>
<td></td>
</tr>
<tr>
<td>Standard deviation</td>
<td>$\sigma = \sqrt{\mu_2(z)} = \sqrt{\sigma^2}$</td>
<td>A measure of average contrast</td>
<td></td>
</tr>
<tr>
<td>Smoothness</td>
<td>$R = 1 - 1/(1 + \sigma^2)$</td>
<td>Measures the relative smoothness of the intensity in a region.</td>
<td></td>
</tr>
<tr>
<td>Third Moment</td>
<td>$\mu_3 = \sum_{i=0}^{L-1} (z_i - m)^3 p(z_i)$</td>
<td>Measures the skewness of a histogram</td>
<td></td>
</tr>
<tr>
<td>Uniformity</td>
<td>$U = \sum_{i=0}^{L-1} b^2(z_i)$</td>
<td>Measures the uniformity of intensity in the histogram</td>
<td></td>
</tr>
<tr>
<td>Entropy</td>
<td>$e = -\sum_{i=0}^{L-1} p(z_i) \log_2 p(z_i)$</td>
<td>A measure of randomness</td>
<td></td>
</tr>
</tbody>
</table>

E. Classification using neural Network:


The proposed structure consists of 6 units in input layer, 20 units in hidden layer, and 1 unit in output layer. The feature values are extracted from the ROI is fed into input layer of neural network, which is applied for two levels. In the first level, neural network classify the segmented ROI into normal (without tumor) and abnormal (with tumor) ROI. Second level neural network classify abnormal ROI into malignant (harmful) and benign (harm less) masses. Neural network are trained by experience, when fed an unknown input into neural network, it can generalize from past experience and produce a result [11]. Six texture features fed to input layer of neural network. The output layer produces either normal or abnormal at first level. If ROI classify as abnormal then neural network perform second level classification and produce either benign or malignant masses.
F. Simulation Results:

1) Case 1: Normal Mammogram

Proposed CAD algorithm applied on MIAS image mdb006 which is a normal mammogram. Simulation result indicating normal at first level classification as shown in figure 2.

![Figure 2: Correct classification of mdb006 as normal](image)

2) Case 2: Abnormal Mammogram

Proposed CAD algorithm applied on MIAS image mdb134 which is a abnormal (malignant) mammogram. Simulation result indicating abnormal at first level & malignant at second level shown in fig 3.

![Figure 3: Correct classification of mdb134 as abnormal (malignant)](image)

3) Case 3: Abnormal Mammogram

Proposed CAD algorithm applied on MIAS image mdb021 which is a abnormal (benign) mammogram. Simulation result indicating normal at first level shown in fig 4.
4) Case 4: Abnormal Mammogram

Proposed CAD algorithm applied on MIAS image mdb105 which is a abnormal (malignant) mammogram. Simulation result indicating abnormal at first level and benign at second level shown in fig 4.

G. Performance Evaluation

Based on ground truth information of MIAS database sensitivity and specificity are measured. Sensitivity and specificity have high values at both level which shows better performance of proposed CAD system.

<table>
<thead>
<tr>
<th>classification (normal and abnormal) 1st level</th>
<th>classification (malignant and benign) 2nd level</th>
</tr>
</thead>
<tbody>
<tr>
<td>TN    FP        SPECIFICITY      ACCURACY</td>
<td>TN    FP        SPECIFICITY      ACCURACY</td>
</tr>
<tr>
<td>49    2          96.07%          95.37%</td>
<td>22    2          91.66%          85.18%</td>
</tr>
<tr>
<td>TP    FN        SENSITIVITY</td>
<td>TP    FN        SENSITIVITY</td>
</tr>
<tr>
<td>54    3          94.73%</td>
<td>24    6          80.00%</td>
</tr>
</tbody>
</table>
IV. CONCLUSION

In this paper, CAD system is proposed for automatic classification of breast masses using neural network. It is implemented using MATLAB and tested on digital mammogram taken from MIAS database for normal – abnormal and malignant – benign classification. Performance evaluation shows that accuracy obtained from the first level classification is 95.37% and second level classification is 85.18%. It is noticed that classification accuracy is reduced at second level due to misclassification in the first level affecting the second level classification.

For future work, accuracy may be improved efficiently by extracting new features and proper feature selection method.

ACKNOWLEDGEMENT

Author expresses her sincere gratitude to Mrs. Meghna Patel (Assistant Professor Dept. of Biomedical Engineering, G.E.C.G, Gandhinagar) for her genuine guidance and valuable suggestions during the course of the work.

REFERENCES