CAD SYSTEM FOR BREAST MICROCALCIFICATION DETECTION

G.G.Udaya¹, J.Preethi²

¹(Computer Science and Engineering, Anna University, Regional Centre, Coimbatore, India udayacse.be@gmail.com)
²(Computer Science and Engineering, Anna University, Regional Centre, Coimbatore, India preethi17j@yahoo.com)

Abstract--Breast cancer is the second leading cause of mortality among women. Over the past 50 years, it has become a major health issue in the world and its rate has increased in recent years. Early detection is the most effective way to diagnose and handle breast cancer. Computer-aided detection or diagnosis (CAD) systems can play a key role in the early detection of breast cancer and can reduce the death rate among women with breast cancer. The CAD system consists of preprocessing, detection and classification of microcalcification. Preprocessing of mammograms is done to improve the quality of the image and in separating the pectoral muscle from the whole breast area. Detection of microcalcification is done by segmentation. In this Particle Swarm Optimization (PSO) has been proposed to segment the suspicious regions from the enhanced image. Classification is done by three steps namely feature extraction, feature selection and by classification. Support Vector Machine (SVM) is used as a classifier, which classifies the microcalcification into benign and malignant. SVM improves classification accuracy and proved to possess the best recognition ability due to its ability to deal with nonlinearly separable data sets.

Keywords--Breast cancer; mammogram; pectoral muscle suppression; computer aided detection; microcalcification; segmentation.

1. INTRODUCTION
Breast cancer is a malignant growth of abnormal cells that begins in the tissues of the breast. It is the most frequently diagnosed cancers and is one of the foremost causes for mortality among women in the world at present. Mammography (x-ray examination of the breast) is currently the frequent procedure available for early detection of breast cancer. It facilitates the revelation of different types of abnormalities such as microcalcifications, masses, bilateral asymmetry and architectural distortions. The need for early detection of breast cancer is highlighted by the fact that incidence rates for breast cancer is one of the highest among all cancers according to the American Cancer Society which quotes a morbidity of 230000 and a mortality of 40000 according to the latest figures gathered for the American population. A group of hastyly dividing cells may form a knob, microcalcifications or architectural distortions which are usually referred to as tumors. Mortality rates due to breast cancer have been falling due to better diagnostic facilities and effectual treatments. One of the principal methods for diagnosing breast cancer is screening mammography. Screening mammography examinations are performed on asymptomatic women to detect early, clinically unsuspected breast cancer. Important signs of breast cancer are clusters of microcalcifications, masses and architectural distortions. Following the results of screening mammography, a follow up study is made for patients according to the level of suspicion of the abnormality. This stage is referred to as diagnostic mammography. Both screening mammography and diagnostic mammography are performed by radiologists who visually inspect the mammograms. Automated screening of mammograms or computer-aided diagnosis (CAD) of breast cancer is an immense field of research. Classifier systems have been widely used in medical diagnosis. Though the most important factor in diagnosis is the evaluation of data taken from patients by human experts, expert systems and various artificial intelligence techniques for classification aid radiologists to a greater extent. As yet, there is no definitive literature which focuses on a detailed discussion on the feature extraction, selection and classification methodologies that reveal the presence of breast cancer. The current study aims at filling this gap by documenting developments in that aspect.

2. RELATED WORK
Worldwide, more than 700,000 women die of breast cancer annually and it is estimated that eight to twelve percent of women will develop breast cancer in their lifetime. Every effort directed to improve early detection is needed [1]. Therefore, many computer vision techniques applied to analysis of digital mammograms have been proposed. Most of them involve an initial processing step that splits the image into remarkable areas, such as the breast region, background and patient markings [2]. One of the first attempts to separate the breast region was presented by Karssemeijer [4] and it was done using simple histogram thresholding. Yang et al [10] provide a scheme based on different thresholds to find the breast edge. The segmentation presented...
by Mendez et al [14] was another gradient based method. After subtracting the background via an initial threshold, an edge was found by a line-by-line gradient analysis. Kato et al [15] used a clustering approach to attain an initial region, calculates the real boundary extrapolating and linking those detected points. Guliato et al [12] use a scale-based fuzzy connectedness algorithm. McLoughlin et al [6] presents self-organizing artificial neural network model. The method applied by Laine et al [9] was a fuzzy segmentation and evaluates the results in terms of completeness and correctness comparing the images from the MIAS database with a gold standard manually generated. The traditional histogram based method has provided good and quick results [3]. This quality sometimes turns on weakness in difficult cases where can be enhanced with local histogram or gradient approaches. Feature selection is an important issue in building classification systems [7]. It is advantageous to limit the number of input features in a classifier to in order to have a good predictive and less computationally intensive model [8]. A computerized methodology for the automatic identification of clustered microcalcifications was proposed by Nishikawa et al [5]. A microcalcification detection algorithm that works on digital mammograms by merging morphological image processing with arithmetic processing was proposed by Murakami et al [11]. An ensemble classifier for the computer-aided diagnosis of breast microcalcification clusters that are tedious to characterize according to the radiologists and computer models alike was proposed by Giger et al [13]. The primary intent of the study was to aid radiologists in to recognize if doubtful calcification clusters are benign or malignant, so that they may potentially suggest fewer needless biopsies for actually benign lesions.

3. COMPUTER-AIDED DIAGNOSIS SYSTEM
Any computer-aided diagnosis system is based on artificial intelligence (AI) techniques. The CAD system for breast cancer detection is similar to any other AI-based system and consists of preprocessing, breast region segmentation, feature extraction and classification. A major difference between computer-aided detection of breast cancer and other AI-based technologies is that breast cancer detection using CAD systems requires human intervention for interpreting the final results. Preprocessing of mammograms is done to improve the contrast of mammograms which will be helpful in further stages of the detection pipeline. This step also includes denoising of the images. Segmenting the breast region from pectoral muscle and surrounding regions is carried out in order to make it easier to extract the suspicious tissues from breast segments. Feature extraction and classification steps are similar to other AI and pattern recognition systems with not much of a difference between commonly used methods. The proposed CAD scheme consists of preprocessing, segmentation, feature extraction, feature selection and microcalcification classification. Fig. 1 shows a standard CAD system.

A. Preprocessing of mammograms
A preprocessing operation suppresses the unimportant image features and an artifact simultaneously enhancing the features of interest often aids an accurate detection. Preprocessing is usually done by image processing methods or by filtering techniques. To remove the background noise preprocessing involves normalization, median filtering and pectoral muscle suppression.

1) Normalization and noise removal: The mammogram image contains noise which appears due to the presence of gray scale variations in the image which is removed by applying normalization and median filtering. Normalization of the image is done by mapping the image within fixed intensities between \( r1 \) and \( r2 \) \( (0 \leq r1 < r2 \leq 255) \). The purpose of it is usually to bring the image into the range that is more familiar or normal to the senses. Median filtering is a common image enhancement technique for removing noise without significantly reducing the sharpness of the image. Median filter first considers each pixel and its neighbors in the image. It replaces the pixel value with the median of the neighboring pixel values. Median is calculated first by sorting the entire pixel values.
from the neighborhood into numerical order and then replaces the pixel being considered with the middle pixel value.

2) Suppression of Pectoral muscle: Pectoral region is one of the main landmarks which have to be diagnosed. Segmentation of this region is the first step in analysis of mammogram. It always appears as a high intensity, triangular region across upper posterior margin of the image. Thresholding method is applied to segment the pectoral muscle region separately. There will be discontinuity between the histogram values of the breast region and the background. Based on the histogram of the mammogram, threshold is selected. The maximum and minimum intensity values are found. The values less than the threshold value are changed to zero.

B. Segmentation

Segmentation is a process which is used to distinguish object from background and is done by four popular approaches for intensity images namely, threshold techniques, edge, region based techniques and connectivity-preserving relaxation methods. Malignancy depends on the segmentation process and therefore it is essential that it should retain all the features of the surrounding tissue. Based on the images present segmentation is performed in two ways. The first one is comparison of left and right images while the second one is the extraction of suspected regions from a single image. Two important landmarks namely breast border and nipple have to be identified first.

1) Breast border detection and nipple identification: Based on the pectoral muscle region suppression, breast border is detected. Threshold method is adapted to detect the breast border. A threshold is used to extract the image from the background which is based on the minimum value between the two peaks. The values less than the threshold value are changed to zero. Morphological closing and opening operations are then used to smooth the boundary of the breast area and the enhancement of border is done by genetic algorithm. The next logical step is to locate the nipple on the mammogram. Usually the lowest point in the boundary will be considered as nipple. Particle Swarm Optimization (PSO) has been proposed for nipple identification. The goal of this method is to discover a pixel of the image on the border that exploits the posterior energy function value.

2) Segmentation of suspicious region: There are two types of segmentation procedures for mammogram images, used to extract the suspicious regions. In the case of pair of images the bilateral subtraction technique is used to extract the suspicious region from digital mammograms based on identity asymmetries between left and right breast image. Image segmentation by PSO is developed and applied to high-resolution digital mammograms with the aim of segmenting normal tissue from cancer tissue. The design of the proposed method for the extraction of suspicious areas from digital mammograms is critically based on two assumptions that,

i. Suspicious areas are brighter than their immediate surrounding tissues, and

ii. The pixels within a suspicious area have relatively uniform intensity.

The mammogram to be segmented is converted into a matrix of pixels and label is given for each matrix. The algorithm is initialized with a swarm of n particles randomly distributed over the search area. The swarm is then set loose in the sense that for a number of iterations each particle moves and updates its velocity in an autonomous manner striving for the optimal position.

Algorithm for segmentation

1. Initialize the particles randomly
2. For each particle create random vectors
3. Update particle positions, particle velocities
4. Calculate the fitness for each particle and find the global best solution
5. If (pbest < gbest) assign pbest to gbest
6. Repeat the procedure until the stopping criteria is related.

C. Feature extraction

Feature extraction is the key for the classification of benign and malignant patterns in digital mammograms. The next step after detection and segmentation of the lesion is the extraction of features that would depict the class to which it belongs. Feature extraction methods examine objects and images to extract the most important features that corresponds to various classes of objects and these features are used as inputs to classifiers that assign them to the class they represent. The extracted features are then used by a classifier to extract the different regions.

1) Mammogram feature construction: Feature construction is a process that aims at discovering hidden relationships between features, inferring new composite features. Identified suspicious regions are further analyzed and statistical features are constructed from these suspicious regions. This step is achieved by checking for each considered pixel pair or group of pixels in the image, whether the considered pixel is located inside the specific region. The check is
Feature construction is achieved by second order and higher order methods. Second-order statistics describe the spatial properties of the texture. It is considered as a very useful characterization that uses subsets of pixels to obtain information on the image texture. The main criterion in the separation of textures that belongs to distinct classes is the difference in their second-order statistics. Textures differing in third or higher order statistics seem to surpass the capabilities of the human perceptual system, it is appropriate to consider the higher-order analysis of mammogram information. A statistics is called higher order if it involves three or more pixels. Two different higher order statistics methods were used in this work for the first time in the context of mammogram image feature analysis.

D. Feature selection

Feature selection is used to improve the efficiency of learning algorithms by finding an optimal subset of relevant features. A successful choice of features make classifier to increase its accuracy, to save the computation time and also to simplify its results. The idea of Swarm Optimization is now used for the optimal feature selection problem.

E. Classification of Microcalcification

The selected features are provided as inputs to the classifier, with the known values. Support Vector Machine (SVM) is used as a classifier. SVMs are a set of related supervised learning methods that explore data and recognize patterns, employed for classification and regression analysis.

Classification phase implements two phases. In the first one, the classifier is useful to classify mammograms into normal and abnormal cases. The mammogram is considered abnormal, if it contains tumor (Microcalcification). Finally, the abnormal mammogram is classified into malignant or benign in the second stage. In this classification stage, SVM classifier in every phase is trained at definite number of training sets in each class. Receiver Operating Characteristics (ROC) analysis is presented to evaluate the classification performance of the textural features extracted by texture-analysis method.

ROC curve illustrates the relationship between the true-positive probability and the false-positive probability. Fig 2 shows the classification of mammogram images into normal, benign and malignant.

4. CONCLUSION AND FUTURE WORK

Mammography is one of the best methods in breast cancer analysis, but in some cases, radiologists cannot analyze tumors despite their experiences. Computer-aided methods could assist radiologist and improve the accuracy of detection. Preprocessing of input image can be done to remove background noise and in separating the pectoral muscle from the whole breast area. Then segmentation can be done by using Optimized swarm method. Then feature can be extracted and selected by using PSO. This selects the best features of the image to be classified. Then classification of microcalcification can be through Support Vector Machine. SVM classify mammograms into normal and abnormal cases. Then the mammogram is considered as abnormal if it contains tumor (Microcalcification). Finally, the abnormal mammogram is classified into malignant or benign in the second stage. The classifier proved to possess the best recognition ability due to its ability to deal with nonlinearly separable data sets. SVM leads to excellent classification accuracy.

Larger datasets can be used that would permit separate training, validation and testing databases with statistically significant results. Building hybrid models by combining models of different types like Hidden Markov model with SVM systems to obtain the best result. Algorithms are to be generalized for applying in other image modalities such as the MRI and the CT.

ACKNOWLEDGEMENT

It’s my immense pleasure to thank my guide and mentor for this earnest efforts and incessant support throughout my project accomplishment. I also thank my family members for their support and all my friends for their encouragement.

REFERENCES


