ijournals International Journal of Software & Hardware Research in Engineering
ISSN No: 2347-4890 Volume 2 Issue 1/ January 2014

Efficient Segmentation and Classification of Micro calcification Using PSO and SVM in Mammographic Images

G.G.Udaya[1]
PG Scholar[1]
Department of Computer Science and Engineering[1]
Anna University Regional Centre Coimbatore, Coimbatore, Tamil Nadu, India[1]

ABSTRACT

Breast cancer takes a tremendous toll on women, next to skin cancer; it is the most common cancer in women and the leading cause of mortality in women. Screening program is based on x-ray examination of the breasts which is the best method for early detection. The development of computer-aided diagnosis system is to assist the radiologist and not to replace the radiologist. The proposed CAD system consists of three stages namely preprocessing, detection and classification of microcalcification. Preprocessing of mammograms is done by normalization, noise removal and pectoral muscle suppression to improve the quality of the image and in separating the pectoral muscle from the whole breast area. Detection of suspicious regions 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. The classifier proved to possess the best recognition ability due to its ability to deal with nonlinearly separable data sets. Receiver Operating Characteristic (ROC) analysis is presented to evaluate the classification performances.

General Terms
Breast Cancer, digital mammography, microcalcification, normalization, segmentation, classification, Pectoral muscle suppression, computer-aided diagnosis (CAD).

1. INTRODUCTION

BREAST cancer is the most recurrently diagnosed cancers and is one of the foremost causes of 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 enables the exposure of different types of abnormalities such as masses, microcalcifications, bilateral asymmetry, and architectural distortions. A group of hastily dividing cells may form a knob, microcalcifications or architectural distortions which are usually referred to as tumors. Breast cancer is any form of malignant tumor which develops from breast cells [3]. Mortality rates due to breast cancer have been falling due to better diagnostic facilities and effectual treatments [5]. 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 [2]. 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 230 000 and a mortality of 40000 according to the latest figures gathered for the American population [6]. Important signs to look for in the case of breast cancer are clusters of microcalcifications, masses and architectural distortions[2]. 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.

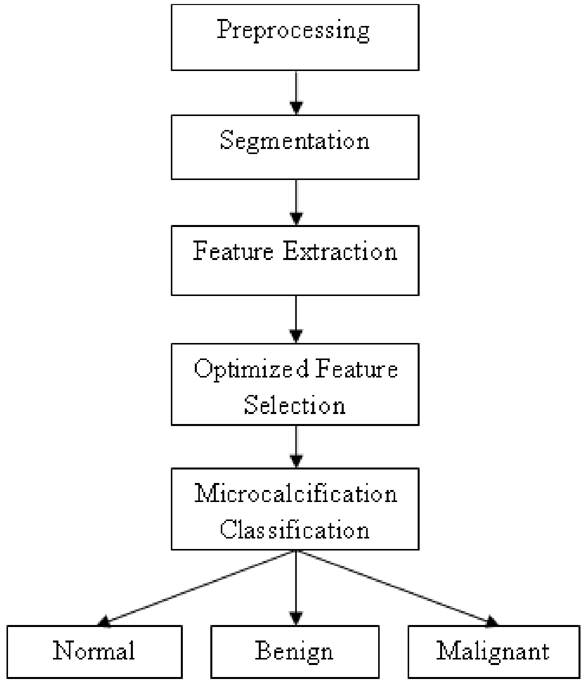

Fig 1: Proposed Architecture

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 a immense field of research. Sampat *et al.*, and Rangayyan *et al.*, provide a widespread review on different stages of a CAD methodology for breast cancer.

Classifier systems have been widely used in medical diagnosis [3]. Though the most important factor in diagnosis is evaluation of data taken from patients by human experts, expert systems and various artificial intelligence techniques for classification aid radiologists to a great extent [3]. As yet, there is no definitive literature which focuses on a detailed discussion on the feature extraction, selection and classification methodologies reveal the presence of breast cancer. The current study aims at filling this gap by documenting developments in that aspect.

2. COMPUTER-AIDED DIAGNOSIS SYSTEM PIPELINE

Any computer-aided diagnosis system is based on artificial intelligence (AI) techniques. The pipeline used in a 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 [7], [2], [4]. Fig 1 shows a standard CAD system pipeline.

3. PREPROCESSING OF MAMMOGRAMS

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

3.1 Normalization and Median Filtering

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 ≤.r1<r2≤ 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.

3.2 Pectoral Muscle Suppression

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.

Steps in Suppression of Pectoral Muscle Region

1. Obtain the mammogram image and assign it to a pixel matrix M of size x*y
2. Store the maximum threshold pixel value from histogram to the variable PK
3. Compare every pixel in the pixel matrix with the threshold value PK, and from the binary matrix BM, with value 1 in corresponding pixel positions of M, when M_{ij} > PK. Else store 0 in Mij.
4. Perform erosion and dilation function on the binary matrix BM.
5. Compare BM with M then substitute M_{ij} with 0 wherever BM_{ij} is 1.

4. DETECTION OF MICRO CALCIFICATION

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.

4.1 Breast Border Detection

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.

4.2 Nipple Identification

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 find a pixel of the image on the border that maximizes the posterior energy function value. Initially assign the values for number of iterations (N), number of particles (k).

Algorithm for Nipple Identification

1. Assign original image to Mij
2. Assign the border pixel in Mij to Bij
3. Assign the kernel of border pixel matrix of size 3*3 from original image to G.
4. Calculate fitness value and assign to U.
5. Calculate G, U for all border pixel in Bij, similarly.
6. Assign values to N,K.
7. For each particles assign a random different kernel which is not same for any other particle.
8. Update velocity
9. For each pixel in Mij, repeat step 7 and 8
10. Calculate pbest
11. If (pbest< gbest) assign pbest to gbest
12. If the kernel generator gbest, then the coordinate of this border pixel from matrix can be considered as nipple position.

4.3 Segmentation

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. In mammographic images, different tissues correspond to different intensity levels. It is necessary to distinguish among background, fat, parenchymal tissues and masses based on intensity levels which generally increase in the same order. The mean intensity of each mammographic region may vary depending on its location and surroundings. 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.

5. 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 represent various classes of objects. 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.

5.1 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 repeated for all the identified regions in the image. Feature construction is achieved by second order and higher order methods.

Second-Order Statistics

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.

The Feature Construction Methods are

- Gray Level Dependence Matrix (GLDM) are matrices whose elements are the probabilities of finding a pixel
- Gray Level Difference Vector (GLDV) was used as a short form of the GLDM method. This method is expected to retain the important information present in the GLDM method.

Higher-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.

The feature contractions methods are

- Neighborhood Grey Tone Difference Matrix (NGTDM) attempts to mimic the mechanism with which the human visual system discriminates texture. It was built on the basis that the properties, which humans use to discriminate between different textural patterns like coarseness, contrast, complexity, busyness (fitness), shape, directionality and texture strength.
- Gray tone Run Length Matrix (GTRLM) is based upon the analysis of higher-order information content

6. 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 provided to a classifier can increase its accuracy, save the computation time and simplify its results. The idea of Swarm Optimization is now used for the optimal feature selection problem.

7. CLASSIFICATION OF MICRO CALCIFICATION

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 analyze data and recognize patterns, used for classification and regression analysis. The standard SVM is a non-probabilistic binary linear classifier, i.e. it predicts, for each given input, which of two possible classes the input is a member of. Classification phase executes two phases.

In the first one, the classifier is applied to classify mammograms into normal and abnormal cases. Then 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 specific number of training set in each category. Receiver Operating Characteristics (ROC) analysis is presented to evaluate the classification performance of the textural features extracted by texture-analysis method.

8. 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 done by using 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.

9. REFERENCES

[1] Karthikeyan Ganesan, U. Rajendra Acharya, Chua Kuang Chua, Lim Choo Min, K. Thomas Abraham, and Kwan-Hoong Ng, "Computer-Aided Breast Cancer Detection Using Mammograms: A Review", IEEE Reviews In Biomedical Engineering, Vol. 6, 2013 77.

[2] C. J. Vyborny, M. L. Giger, and R. M. Nishikawa, "Computer aided detection and diagnosis of breast cancer," *Radiologic Clinics N. Amer.*, vol. 38, no. 4, pp. 725–740, Jul. 1, 2000, 10.1016/S0033-8389(05)70197-4, 0033-8389.

[3] G. Boccignone, A. Chianese, and A. Picariello, "Computer aided detection of microcalcifications in digital mammograms," *Computers Biol. Medicine*, vol. 30, no. 5, pp. 267–286, Sep. 1, 2000, 10.1016/S0010- 4825(00)00014-7, 0010-4825.

[4] A. Papadopoulos, D. I. Fotiadis, and L. Costaridou, "Improvement of microcalcification cluster detection in mammography utilizing image enhancement techniques," *Computers Biol. Medicine*, vol. 38, no. 10, pp. 1045–1055, Oct. 2008, 10.1016/j.compbiomed.2008.07.006, 0010- 4825.

[5] A. Laine, J. Fan, and W. Yang, "Wavelets for contrast enhancement of digital mammography," *IEEE Eng. Medicine Biol.*, vol. 14, no. 5, pp. 536–550, Sep./Oct. 1995, 10.1109/51.464770.

[6] H. Jing, Y. Yang, and R. M. Nishikawa, "Detection of clustered microcalcifications using spatial point process modeling," *Phys. Med. Biol.*, vol. 56, no. 1, 2011, 10.1088/0031-9155/56/1/001.

[7] F. F. Yin, M. L. Giger, K. Doi, C. J. Vyborny, and R. A. Schmidt, "Computerized detection of masses in digital mammograms: Automated alignment of breast images and its effect on bilateral-subtraction technique," *Med. Phys.*, vol. 21, no. 3, pp. 445–452, 1994.

[8] A. J. Mendez, P. G. Tahoces, M. J. Lado, M. Souto, and J. J. Vidal, "Computer-aided diagnosis: Automatic detection of malignant masses in digitized mammograms," *Med. Phys.*, vol. 25, no. 6, pp. 957–964, 1998.

[9] T. Nakagawa, S. Kasai, T. Hara, H. Fujita, and T. Endo, "Automated extraction of pectoral-muscle region on mammogram," *Med. Imag. Inform. Sci.*, vol. 18, no. 1, pp. 39–45, 2001.

[10] S. Pohlman, K. A. Powell, N. A. Obuchowski, W. A. Chilcote, and S. Grundfest-Broniatowski, "Quantitative classification of breast tumors in digitized mammograms," *Med. Phys.*, vol. 23, no. 8, pp. 1337–1345, Aug. 1996.

[11] A. Petrosian, H. P. Chan, M. A. Helvie, M. M. Goodsitt, and D. D. Adler, "Computer-aided diagnosis in mammography: Classification of mass and normal tissue by texture analysis," *Phys. Med. Biol.*, vol. 39, pp. 2273–2288, 1994.

[12] H. Jing, Y. Yang, and R. M. Nishikawa, "Detection of clustered microcalcifications using spatial point process modeling," *Phys. Med. Biol.*, vol. 56, no. 1, 2011, 10.1088/0031-9155/56/1/001.

[13] R. M. Haralick, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. 3, pp. 610–621, Dec. 1973.

[14] N. Karssemeijer, "Adaptive noise equalization and recognition of microcalcification clusters in mammograms," *Int. J. Pattern Recog. Artif. Intell.*, vol. 7, pp. 1357–1376, 1993.