2012 International Conference on Image, Vision and Computing (ICIVC 2012)
IPCSIT vol. 50 (2012) © (2012) IACSIT Press, Singapore
DOI: 10.7763/IPCSIT.2012.V50.47

Detection System of Microcalcification Based on a Hybrid and Intelligent Classifier

Ying KONG[1+], and Jie ZHAO[2]

[1]School of Information and Electronic Engineering, Zhejiang University of Science and Technology, Hangzhou, Zhejiang Province, China

[2]School of Experimentation and Practice Training Management Center, Zhejiang Police Vocational Academy, Hangzhou, Zhejiang Province, China

Abstract. One of the early signs of breast cancer is the presence of microcalcification clusters at the mammogram of asymptomatic women. However, a number of such findings especially the microcalcifications that have small size and low contrast could be missed or misinterpreted by doctors. Thus, we present a new Computer-aided detection system based on neural network classifier for the identification of microcalcification clusters in digital mammograms. The proposed method is based on a three-step procedure: preprocessing and segmentation; regions of interest (ROI) specification; feature extraction and classification. Results are presented as the receiver operating characteristic (ROC) performance and are quantified by the area under the ROC curve (Az). The system successfully combines intelligent methods and image processing techniques which contribute to the enhancement of mammographic diagnosis sensitivity and reduction of negative biopsies.

Keywords: Microcalcification clusters detection;neural network;Mammograph; Computer-aided detection (CAD)

1. Introduction

Breast cancer is currently one of the leading causes of death among women worldwide.Regular mammographic screening programs for women of certain age or high-risk groups are taking place in a number of countries on a nation-wide basis or as projects organized from several institutes[1].Although some researchers doubt about the real effectiveness of population screening programs the majority of them contribute to the mortality reduction .

Medical imaging has revolutionized the practice of medicine in the past century. Physicians are empowered to see through the human body for abnormalities non-invasively, and to make diagnostic decisions rapidly, which has impacted the therapeutic outcomes of the detected diseases. The several types of breast abnormalities that are visible in mammograms include asylmnetry between the breasts, distortion in the architecture of tile breast, increase in breast tissue density, masses, and calcifications[1-3]. Normally, structures of die breast converge toward the nipple ill a smooth pattern. Tissues usually become denser when diseased.

In this paper, we present an intelligent system (Fig. 1) for the identification of microcalcification clusters in digitized mammographic images. The Classification, as it consists of three modules: the preprocessing and segmentation, (ROI) specification and the feature extraction and the classification module. The latter is a hybrid classification schema composed of a rule-based and a neural network sub-system. The proposed systemis fast and accurate in the detection of ROIs.We employ an additional feature for ROI characterization that is related with the existence of a small ROI in the neighborhood of a large one. The proposed

+ Corresponding author.
E-mail address: Kongying-888@163.com

methodology could be an essential part of an integrated CAD technique, which could assist radiologists in mammogram analysis and diagnostic decision making. The classification successfully combines intelligent methods and image processing techniques which contribute to the enhancement of mammographic diagnosis sensitivity and reduction of negative biopsies.

2. Process of the classification

2.1. Designing method and database

The Classification, as it consists of three modules: the preprocessing and segmentation, (ROI) specification and the feature extraction and the classification module.

For the development and evaluation of the proposed model we used the Nijmegen [4]and theMIAS [5] databases. The first contains 40 mammograms of both craniocaudal and oblique views from 21 patients. The proposed system is implemented in three stages. The first is related to image segmentation, the second with the identification of candidate ROIs, and the third with the characterisation of each ROI as cluster of microcalcifications or not.

2.2. Preprocessing and segmentation module

At the beginning of preprocessing it is necessary to locate the breast region. For this reason we apply a skin-line segmentation procedure by setting equal to zero the image pixels with intensity less than 20 (for 0–255 Gy levels).

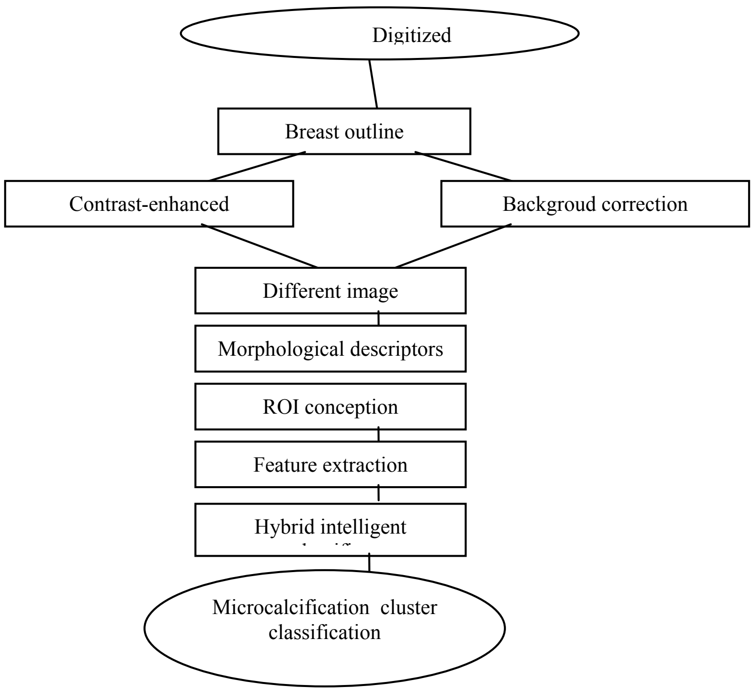

Fig.1: The microcalcification cluster detection modules.

Most of those pixels belong to the background area, although a small number exists belonging to the tissue area close to the breast surface. This thresholding procedure results in a binary image of white objects on a black background. Neighbouring white pixels with connectivity of eight are grouped together to form objects corresponding either to the breast region or to marks and film artifacts. The largest object corresponds to the breast region (Fig. 2) and close to the breast outline a number of very small objects appear. These are actually part of the breast region but, due to thresholding, they appear as distinct objects. To deal with this problem, we apply morphological dilation with a structure element radius of 30 pixels (1.5 mm) . This results in an expansion of breast region outline, which includes all the nearby located objects. All the pixels that do not belong to the expanded breast area are set to zero, resulting in the removal of background, marks and artifacts.

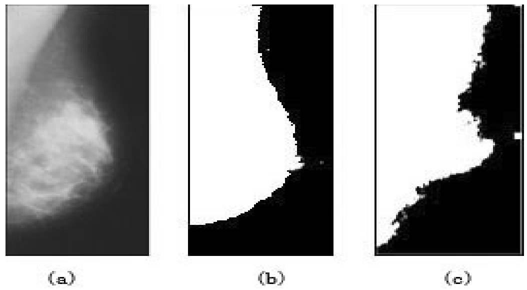

Fig.2: (a) An original mammogram, (b) the different objects appearing in the binary image and (c) a zoom view in the area of the breast skinline.

The artifacts located at the boundary of the breast region, at the chest side, forming a thick line are eliminated too. The minimum rectangle containing the breast region is automatically drawn and it is used in the subsequent processing stages.Using the structural element figer b gray corrode the image f, marked as $f\Theta b$,defined it as:

$$(f\Theta b)(s,t)=\min\left\{f(s+x,t+y)-b(x,y)(s+x),(t+y)\in D_f,(x,y)\in D_b\right\} \quad (1)$$

D_f and D_b is the individual definition of f and b 。
The definition of gray expansion:

$$(f\Theta b)(s,t)=\min\left\{f(s+x,t+y)-b(x,y)(s+x),(t+y)\in D_f,(x,y)\in D_b\right\} \quad (2)$$

In a typical image, the number of selected pixels is quite large and in subsequent processing a fraction of them will be removed.If the amount of the selected pixels is lower than 10% of the total number of pixels of the cropped mammogram, the pixels with intensity higher than half of the previously specified threshold are added. In such a way an adequate number of pixels are included in the obtained binary image fig3(a). This case occurs when the mammogram exhibits very low contrast usually due to erroneous exposure conditions. Next a contrast enhancement filter is applied with 9*9 kernel having central element equal to 80 and all the other elements equal to. Five percent of the pixels having the highest intensity are selected, producing a second binary image fig3(b). The outcome of the segmentation module is an image produced by the logical summation (AND) of the two binary images A and B. It contains the pixels that have high intensity values and, at the same time, quite high intensity values in comparison with the background intensity of their local neighbourhood. It is the image fig3(c).

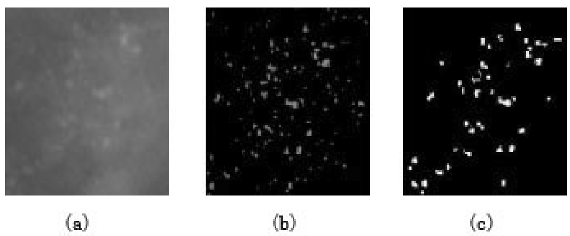

Fig.3: (a) A part of a mammogram (original image), (b) the output of the segmentation component and (c) the binary image after small object elimination.

2.3. Instruction of interest specification

In the segmented image obtained in the previous stage, neighbouring pixels with connectivity of eight are grouped together to create possible microcalcification objects. Objects containing one or two pixels are rejected since they are considered as artifacts [7].Since the diagnostic information is based on the existence of groups of objects, individual objects (possibly artifacts) should be removed. The elimination of these artifacts is achieved through the use of morphological operators. The application of the erosion operator (with structure element a 3* 3 kernel of unit value) results in the removal of all objects apart from those that have at least one innermost pixel that is not part of their boundary. In this way, only inner pixels that belong to large objects remain. These pixels correspond to the centres of ROIs, which are generated using the dilation operator with a

3*3 structure element of unit value. The dilation is repeated 50 times in order to produce a ROI with sufficient area around the object.

The smallest possible size of ROI is 101*101 pixels and appears when the central pixel of an object is isolated and no other central pixel is located at a distance smaller than 100 pixels.The set of ROIs is partitionned in two groups depending on their area. The first group contains those ROIs with areas lower than 20,000 pixels (2*100*100), which is a reliable threshold value discriminating ROIs that are generated from individual objects. The second group contains the remaining ROIs which contain at least two nearby objects. This discrimination of ROIs defines a novel feature that will be used at the classification stage.

2.4. Classification module

The objective of the classification module is to categorize the specified ROIs as true microcalcification clusters or not. The large number of false positive clusters that are identified by the segmentation process makes the characterization task difficult. In order to specify the features that will be used as inputs to the classification system, at first 54 features are identified and computed characterizing either an individual microcalcification (object) or a group of them in a specific ROI. Those features fall into three categories related with the intensity, shape and texture properties of each object. It should be noted that does not exit any particular feature indicating the relation of each ROI with the mammographic image of origin since each area is treated separately. The group features are computed as the mean value of the five largest objects included in a ROI. The term largest refers to the number of pixels each object is composed, in the binary-segmented image. The selection of the five largest microcalcifications is made since a very small microcalcifica-tion does not have enough pixels for reliable feature value computation [6].

In the next step of the classification module the selected features are fed into a hybrid intelligent classification system, which consists of two components (Fig. 4): a rule-based and a neural network component. The rule construction procedure consists of the feature identification step as well as the selection of the particular threshold value for each feature. First, visualization of all the calculated features in two-dimensional plots, in pairs, has been employed for the selection of suitable feature threshold values that lead to the categoriza-tion of a remarkable number of ROIs. For every feature, several threshold values are examined in the range of values corresponding to that feature. For each threshold value, the number of ROIs below and above the threshold value is recorded. The ratio of the number of ROIs that belong to a specific class (normal or pathological) over the total number of the ROIs that belong to the same class should be more than 6%. In addition, the number of the false negative ROIs must be equal or less than one.

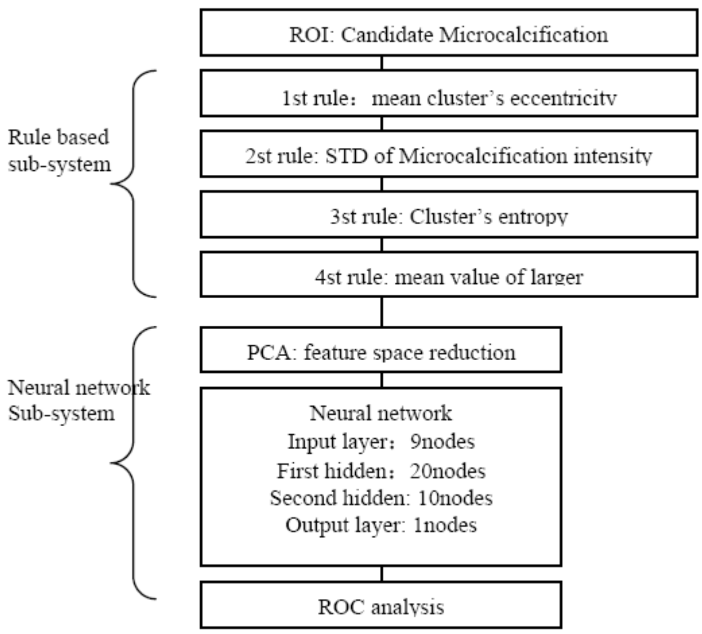

Fig.4: The hybrid classification system.

The ROIs that have been classified by the rule-based system are easily identified regions that are subsequently removed from the set that is used for training and testing of the neural network component. The remaining ROIs constitute the dataset that will be used for the construction of the neural network. The input feature vector of the latter contains the total number of features (22 features per ROI) and includes those features used by the rule-based system. The two components (rule-based and neural network) are sequentially applied in the classification scheme. Only if a ROI remains uncharacterized by the rule-based system, it is subsequently fed to the neural network module for characterization.The neural network (Fig.5) that is used for ROI characterisation is a feedforward neural network with sigmoid hidden nodes (Multiplayer Perceptron—MLP). In order to select an appropriate architecture .

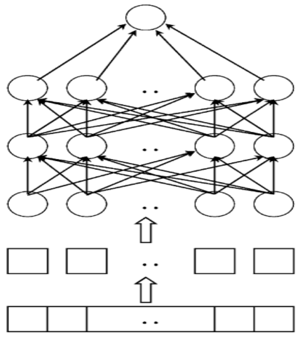

Fig.5: The neural network sub-system.

3. Result

This paper has used MIAS data set to test. The processing of image segmentation modules in this set has given rise to 193 candidate ROIs, among which 34 are true. The sub-systems based on neural network have characterized 116 ROIs that correspond to 25 is true positivities, 79 true negativities, 12 false positivities and 0 false negative event.

Table 1 lists the test results achieved with classification approaches based on the system of hybrid neural network classifier. As it is revealed by the table, the false positive nanoclusters of each image have significantly decreased at the same level of sensitivity in detection system based on intelligent classifier. For example, at the sensitivity level of 0.9, there are 0.65 false positive nanoclusters in each image when using intelligent classifier while there are 3.00 and 1.15 false positive nanoclusters in each image when using neural network classification system.

TABLE I. EXPERIENT RESULTS

Classification Approach	**Sensitivity**	**False Positive Nanoclusters of Each Image**	**Number of Tumors Undetected**
Using neural network classification system	0.85	2.55	3
	0.93	3.05	
	0.98	3.37	
Using Detection system of Microcalcification based on a hybrid and intelligent classifier	0.82	0.95	1
	0.89	1.14	
	0.97	1.34	
	0.98	1.36	
Using Cross validation	0.79	0.86	2
	0.84	0.82	
	0.90	0.77	
	0.96	0.60	

4. conclusion

The experimental results show that the classifier can effectively achieve the detection of tumors in a mammary X-ray picture with hybrid neural network classifier. The proposed system exhibits high performance in the detection of microcalcification clusters since it is able to identify more that 90% of the total number of clusters with a rather small number of false positive findings. Therefore, this approach is more useful for the diagnosis of early breast cancer. In addition, more efforts shall be made to perfect detection system and determine the attribute of micro-calcification clusters, benign or malignant.

5. References

[1] Claridge E, Richter JH (1994) Characterization of mammographic lesions. In: Proceedings of the 2nd international workshop on digital mammography, York, England, July 1994,pp 241-250.

[2] Zhou X, Gordon R (1989) Detection of early breast cancer: an overview and future prospects. Crit Rev Biomed Eng 17(3):203-255.

[3] Skinner M, Swain M, Simmons R, McCarty K, Sullivan D,Iglehart D (1988) Nonpaipable breast lesions at biopsy. AnnSurg 208(2):203-208.

[4] Smaranda Belciug,Florin Gorunescu.Assessing performances of unsupervised and supervised neural networks in breast cancer detection[J]. The7th international Conference on Informatics and Systems(INFOS), 2010(3):1-8.(in Cairo)

[5] R.R.Janghel,Anupam Shukla.Breast cancer diagnosis using Artificial Neural Network models[C].3th International Conference on Information Sciences and Interaction Sciences(ICIS), 2010(6):89-94.(in China)

[6] Suckling J, Parker J, Dance D, Astley S, Hutt I, Boggis C, et al. The mammographic images analysis society digital mammogram database. Exerpta Medica 1994;1069:375–8.

[7] Betal D, Roberts N, Whitehouse GH. Segmentation and numeral analysis of microcalcifications on mammograms using mathematical morphology. Br J Radiol 1997;70(9):903–17.