Detection of cancer tumors in mammography images using support vector machine and mixed gravitational search algorithm

Fatemeh Shirazi
Department of electrical engineering
Graduate university of advanced technology
Kerman, Iran
Shirazi.fatemeh0@gmail.com

Esmat Rashedi
Department of electrical engineering
Graduate university of advanced technology
Kerman, Iran
e.rashedi@kgut.ac.ir

***Abstract*—In this paper, support vector machine (SVM) and mixed gravitational search algorithm (MGSA) are utilized to detect the breast cancer tumors in mammography images. Sech template matching method is used to segment images and extract the regions of interest (ROIs). Gray-level co-occurrence matrix (GLCM) is used to extract features. The mixed GSA is used for optimization of the classifier parameters and selecting salient features. The main goal of using MGSA-SVM is to decrease the number of features and to improve the SVM classification accuracy. Finally, the selected features and the tuned SVM classifier are used for detecting tumors. The experimental results show that the proposed method is able to optimize both feature selection and the SVM parameters for the breast cancer tumor detection.**

Keywords-breast cancer detection; support vector machine; parameter setting; feature selection; mixed gravitational search algorithm.

I. Introduction

Breast cancer is one of the most common cancers among women. Tumor detection in mammography images is a challenging problem for radiologists because tumors are located into complicated tissues; also, they are in different shapes and sizes. Recently, many computer-aided diagnosis (CAD) systems have been developed for breast cancer tumors controlling. Tumor detection in the early step can be helpful for patient. Image processing and machine learning methods are frequently used for this goal. Ref [1] utilized a tumor detection method based on complex texture features that extracted the information of discrete photometric distribution and local intensity relation. Ref [2] described a sphere template matching method to segment the tumors in mammogram images that areas from mammograms with a sphere were correlated on a two dimensional surface that yielded regions of tumors. Active contour based on curve evolution and Mumford-shah functional for segmentation was utilized in [3, 4, 5]. Ref [6] was proposed a feature extraction method for the early detection of tumors in mammograms that extracted discrete photometric features. Ref [7] was proposed an approach using histogram of oriented gradients (HOG) for tumor detection. This method was based on the distribution of differential intensity histogram in the image. Ref [8] described a concentric morphology model for tumor detection that was based on the existence of the concentric layers surrounding a central area. A set of statistical features based on wavelets was used for feature extraction in mammography images [9].

Gravitational search algorithm (GSA) was proposed in [10] and in some works was used for enhancing classification algorithms [11, 12] and image processing algorithms [13]. Ref [11] was proposed a feature selection (FS) method that its goal was improving the classification accuracy using improved binary gravitational search algorithm (IBGSA). Ref [12] represented the method to optimize the feature selection and the SVM parameter tuning, simultaneously. The combination of the mixed GSA (MGSA) and SVM were introduced as a hybrid system that was able to select the best features and increase the classification accuracy [12].

Some optimization techniques were used to enhance tumor detection, such as the genetic algorithm (GA) [14] and particle swarm optimization (PSO) [15]. Ref [16] proposed a neural-genetic algorithm for feature selection in combination of neural and statistical classifiers to classify in mammograms. Several types of features such as texture features and shape features were utilized for feature extraction in [17] and the combination of GA and SVM were used for optimization of classification process. In this paper, to reduce the human errors and improve detection accuracy of tumors, the mixed GSA is used to improve the SVM classifier accuracy in tumor detection.

This paper is organized as follows: part II describes the proposed method. The experiments and results are presented in part III and we finish with a conclusion in part IV.

II. THE PROPOSED METHOD

There are four main stages for tumor detection. “Fig. 1” shows the proposed algorithm for tumor detection. In this figure, at first, images are preprocessed by morphological filters. Then, ROIs are determined by sech template matching method. Extracted features fed to SVM to determine the class of each object. Details come next.

978-1-4673-8737-8/16/$31.00 ©2016 IEEE

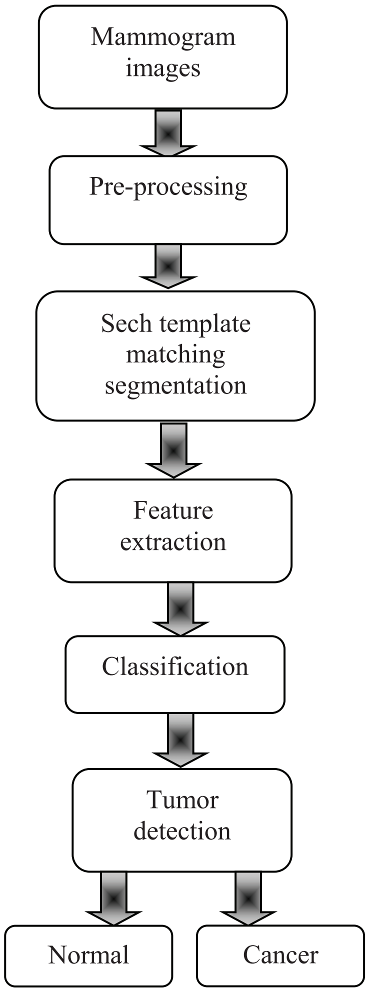

Figure 1: The block diagram of the tumor detection process.

A. Preprocessing

Mammography preprocessing is used to decrease the effects of image noises. In this case, the two types of morphological filters, the opening filter and closing filter, are used in the mammography gray images

B. Segmentation

To separate regions of interest (ROIs) in each image, hierarchical template matching method [18] using the sech template is used. This template [1] is described as (1).

$$S(x,y) = \frac{2}{exp\left(\beta * \sqrt{x^2+y^2}\right) + exp\left(-\beta * \sqrt{x^2+y^2}\right)} \quad (1)$$

The parameter β controls the rate of change on the gray level in the template from the center to the boundary. The S(x,y) is the gray level of the template image at the situation (x,y). In this paper, the value of β is set to“0.08”.

C. Feature extraction

After segmentation of regions of interest, some features are extracted to determine the characteristics of the objects in the region. Gray-level co-occurrence matrix (GLCM) [19] describes texture features of the image. Texture features are determined as entropy, energy, contrast, homogeneity, difference variance and etc. The other features are mean [1] and variance.

D. Classification

In this stage, SVM is used as a classifier. Support vector machine (SVM) classifier works based on the linear function in a high dimensional feature space that finds an optimal separating hyper plane. The SVM classifier has high accuracy compared to other kinds of methods.

E. Optimization

In this paper, improving the classification accuracy is based on the combination of binary GSA and continuous-valued GSA that is named as MGSA. “Fig. 2” shows the block diagram of the proposed method. According to this figure, MGSA is simultaneously applied to the both parts of feature selection and setting the SVM parameters.

MGSA improves two parameters of the SVM. These parameters are the kernel Parameter, gamma (γ), for the radial basis functions (RBF) kernel and penalty parameter (C) [12]. MGSA also selects the best set of features. The feature selection is useful for enhancing the classification performance. In MGSA, each object is consisted of the two real-valued SVM parameters and the number of selected features.

If all of these extracted features (such as mean, variance, entropy, Correlation and etc.) from images are used, the processing time is increased and the training of the SVM would be complicated.

As mentioned before, the feature selection and the SVM parameter setting are important to have the high classification accuracy. Here, the proposed method finds the suitable features for tumor detection and simultaneously optimizes the SVM parameters.

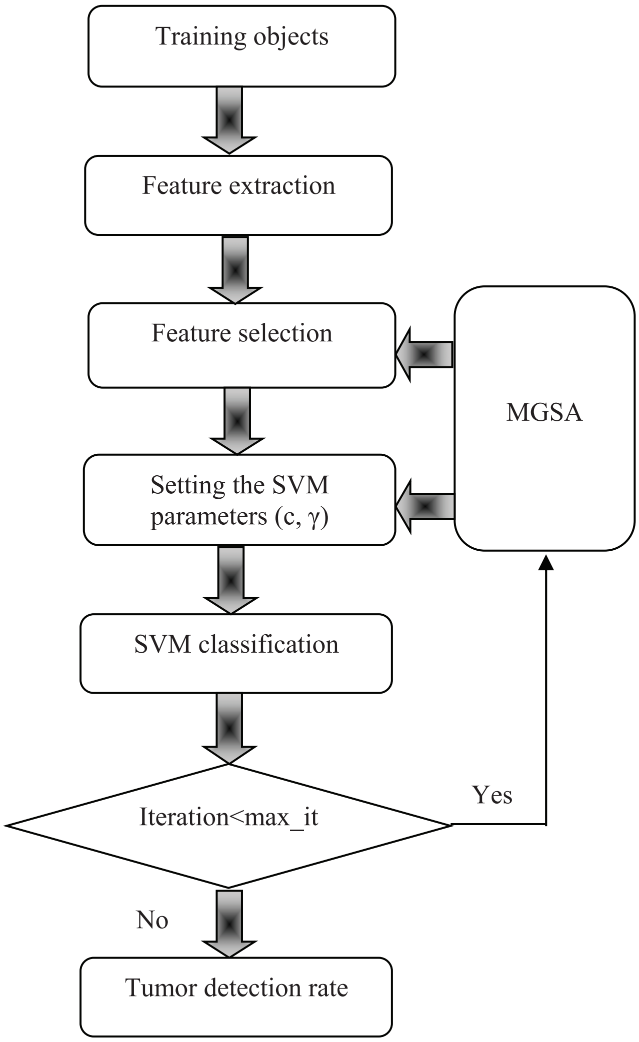

Figure 2: The block diagram of using MGSA for tuning SVM and selecting salient features.

In MGSA method, subsets of extracted features from training data are evaluated based on fitness function. Therefore, the optimal set of features is selected.

The aim is to optimize the evaluation function that is the precision of the SVM classifier. After training, the selected features and tuned SVM are used for detecting target objects.

III. EXPERIMENTS AND RESULTS

The mini-MIAS [20] database is used in this paper. These images are 1024×1024 pixels. In this study, mammogram images are used including cases of both normal and masses. "Fig. 3" shows some of the target objects.

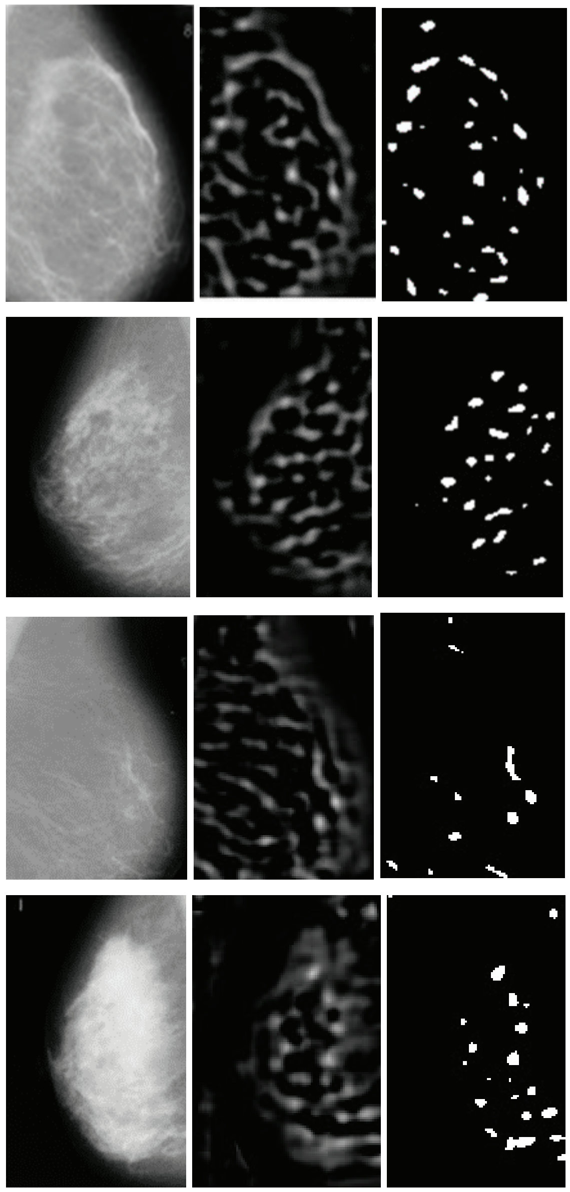

Figure 3: Some samples of target objects (normal and cancer regions), the original images are in the first column, target objects of images achieved by template matching are in the second column, and the binary of target objects are in the third column.

The dataset is separated into training data and test data. The dataset contains 100 ROIs of mass and normal tissues. These ROIs are produced by applying image segmentation in each image. Training data is randomly selected. 70% of dataset are used as training data and 30% are used as test data.

From each ROIs, 24 features have been extracted. These features are 22 features of GLCM, and mean and variance of the region. In our experiments, MGSA was applied to SVM and selected 12 of features as salient features. The values of optimized parameter γ and optimized parameter C were "9.18" and "6.84", respectively. Hence, the number of selected features and the two real-valued SVM parameters were attained, simultaneously.

In this study, once all of the features are applied to SVM and the recognition rate is calculated. Then, MGSA is applied to SVM and also the recognition rate is calculated.

Recognition rate on test data are reported in Table 1. According to Table 1, recognition rate of SVM classifier is lower than recognition rate of MGSA-SVM. Therefore, MGSA –SVM is suitable for tumor detection because the MGSA is more capable of tuning the classifier. By comparing SVM with MGSA-SVM, it could be found that the proposed method is efficient for tumor detection in mammography images.

TABLE I. PRECISION RESULTS

	PERFORMANCE	
Method	***SVM***	***MGSA-SVM***
Classification Rate (in percent)	86	93.1
Number of features	24	12

IV. CONCLUSIONS

In this paper, breast cancer tumor detection using the combination of support vector machine and MGSA has been represented. The target data are the cancer and normal regions. The main purpose of this paper is optimizing the both SVM parameters and feature selection, effectively. The proposed method is compared with SVM classifier in MIAS database. Our experimental results indicate that the recognition rate in the proposed method is reasonable for the breast cancer detection.

ACKNOWLEDGMENT

The authors would like to thank Dr. Soroor Nazemi and Dr. Kambiz Bahaadini from Kerman University of medical sciences for their guidance and encourage.

REFERENCES

[1] S.C Tai, Z.S Chen, W.T Tsai, "An Automatic Mass Detection System in Mammograms Based on Complex Texture Features," IEEE journal of biomedical and health informatics, vol. 18, pp. 618-627, March 2014.

[2] G. M. Te Brake, N. Karssemeijer, "Single and multiscale detection of masses in digital mammograms," Medical Imaging, IEEE Transactions on, vol. 18, pp. 628-639, July1999.

[3] T. F. Chan and L. Vese, "Active contours without edges," Image processing, IEEE transactions on, vol. 10, pp. 266-277, February 2001.

[4] D. Mumford and J. Shah, "Optimal approximation by piecewise smooth functions and associated variational problems," Commun. Pure Appl. Math, vol. 42, pp. 577–685, July 1989.

[5] V. Caselles, F. Catté, T. Coll, and F. Dibos, "A geometric model for active contours in image processing," Numer. Math., vol. 66, pp. 1–31, December 1993.

[6] M. Sameti, R. Ward, J. Morgan-Parkes, and B. Palcic, "Image feature extraction in the last screening mammograms prior to detection of breast cancer," IEEE J. Sel. Topics Signal Process, vol. 3, pp. 46–52, February 2009.

[7] S. Ergin, O. Kilinc, "A new feature extraction framework based on wavelets for breast cancer diagnosis," Computers in Biology and Medicine, vol. 51, pp. 171–182, August 2014.

[8] N. H. Eltonsy, G. D. Tourassi, A. S. Elmaghraby, "A Concentric Morphology Model for the Detection of Masses in Mammography," IEEE transactions on medical imaging, vol. 26, June 2007.

[9] S. Liu, C. F. Babbs, and E. J. Delp, "Multiresolution detection of spiculated lesions in digital mammograms," Image Processing, IEEE Transactions on, vol. 10, pp. 874-884, January 2001.

[10] E. Rashedi, H. Nezamabadi-pour, and S. Saryazdi, "GSA: a Gravitational Search Algorithm," Information Sciences. J, vol. 179, pp. 2232–2248, June 2009.

[11] E. Rashedi, H. Nezamabadi-pour, "Feature subset selection using improved binary gravitational search algorithm," Journal of Intelligent & Fuzzy Systems, vol. 26, pp. 1211–1221, May 2014.

[12] S. Sarafrazi, H. Nezamabadi-pour, "Facing the classification of binary problems with a GSA-SVM hybrid," Mathematical and Computer Modelling,vol. 57,pp. 270–278, January 2013.

[13] E. Rashedi, H. Nezamabadi-pour, S. Saryazdi, "A simultaneous feature adaptation and feature selection method for content-based image retrieval systems," Knowledge-Based Systems, vol. 39, pp. 85–94, February 2013.

[14] A. Bhardwaj, A. Tiwari, "Breast cancer diagnosis using genetically optimized neural network model," Expert Systems with Applications, vol. 42, pp. 4611-4620, June 2015.

[15] J. Dheeba, N. A. Singh, and S. T. Selvi, "Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach," Journal of biomedical informatics, vol. 49, pp. 45-52, June 2014.

[16] P. Zhang, B. Verma, and K. Kumar, "A neural-genetic algorithm for feature selection and breast abnormality classification in digital mammography," in Neural Networks, 2004 IEEE International Joint Conference on,vol. 3, pp. 2303-2308, July 2004.

[17] N. Azizi, N. Zemmal, M. Sellami, and N. Farah, "A new hybrid method combining genetic algorithm and support vector machine classifier: Application to CAD system for mammogram images," in Multimedia Computing and Systems (ICMCS), 2014 International Conference on, Marrakech, , pp. 415-420, April 2014.

[18] X. Shengzhou and P. Chengdan, "Hierarchical matching for automatic detection of masses in mammograms," in Electrical and Control Engineering (ICECE), 2011 International Conference on, Yichang, pp. 4523-4526, September 2011.

[19] R. Parekh, "Using texture analysis for medical diagnosis," IEEE Multi-Media, vol. 19, pp. 28 –37, February. 2012.

[20] http://peipa.essex.ac.uk/info/mias.html.