ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 2, Issue 5, May 2015

Detection and Classification of Masses in Mammograms Using a Hybrid GA-PSO-KNN Approach

Subashini Sundaravinayagam[1], Bhavani Sankari. S[2]
PG Scholar,Jerusalem College Of Engineering[1]
Associate Professor,Jerusalem College Of Engineering[2]

Abstract: A leading cause of high mortality rate in women above 40 years is the breast cancer. The above, combined with the fact that women have inhibition in undergoing a mammogram scan, increases the problem manifold. The method currently available for the early detection of breast cancer is screening mammography. The screening method if supplemented with computer aided diagnosis is very effective in early detection of breast cancer and its cure. The project proposes such an effective computer aided diagnosis technique using GAPSO-KNN approach, wherein the Region Of Interest is classified as mass and normal breast tissue regions. Reference data set has been collected from MIAS-mini mammographic database, LBP of each ROI is found from which twenty two features are extracted using Gray Level Co-occurrence Matrix. GAPSO is used for searching the best feature set and KNN classifier for classification.

Keywords: mammography, mass classification, particle swarm optimization, feature selection, GAPSO, KNN classifier, GLCM, LBP

I. Introduction

Incidence of breast cancer in India is on the rise and is rapidly becoming the number one cancer in females pushing the cervical cancer to the second spot [1]. In India, the death toll due to the breast cancer is increasing at a rapid pace [2].This warrants for early detection and diagnosis. Controlling the breast cancer has been a major challenge in India especially in case of marginalized women. Breast cancer is the most common diagnosed malignancy in women worldwide (22%) and in India (18.5%) it ranks second to cervical cancer.

The burden of breast cancer is increasing in both developed and developing countries; the peak occurrence of breast cancer in developed countries is above the age of 50 whereas in India it is above the age of 40 [4]. In India the age standardized incidence rate of breast cancer varies from 9 to 32 per 1, 00,000 women. To generate the reliable data on magnitude and pattern of cancer, India started National cancer registry program in 1981 [5]. Up to 2003 the program comprised of six population based cancer registry and one registry serving rural area covering the total population of 35.7 million (only 3.5% of the Indian total population) [6] and an increasing trend in incidence is reported from various registries of national cancer registry project and now India is a country with largest estimated number of breast cancer deaths worldwide [9].

Early detection and appropriate treatment of breast cancer can significantly increase the chances of survival. They have also shown that early detection of small lesions boosts prognosis and leads to a significant reduction in mortality. Mammography is in this case the best diagnostic technique for screening. , the interpretation of mammograms is not easy because of small differences in densities of different tissues within the image. This is especially true for dense breasts. An automatic early detection of breast cancer by analyzing mammographic images can ease this process. This analysis could provide radiologists a better understanding of stereotypes and provides, if it is detected at an early stage, a better prognosis inducing a significant decrease in mortality.

Evolutionary computation techniques like Genetic algorithms (GAs) and particle swarm optimization (PSO) have been used in feature selection due to their global search ability. PSO is easier to implement, computationally less expensive and can converge quickly [3]. The drawback of PSO is that the swarm may prematurely converge [7] Another reason is the fast rate of information flow between particles, resulting in the creation of similar particles, leading to loss in diversity that increases the possibility of being trapped in local optima [8].

Recently, a hybrid algorithm called GAPSO (Genetic algorithm based Particle swarm optimization) is being used as a feature selection method and has proven to give better results than standard PSO based on benchmark functions. The purpose of this paper is to use the hybrid GAPSO approach to select the significant texture features from ROI. The KNN classifier is then trained using the training set of images. The trained KNN classifier is then used to classify between normal and abnormal breast tissue.

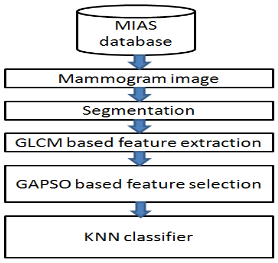


Figure1. Workflow of the GAPSO based breast mass classification

The main objective is to show that a small number of significant GLCM based texture features found by GAPSO-KNN feature selection can have better or comparable performance in classification accuracy when compared to the full set of features or other existing mass classification methods.

II. Generic Flow Of Mammogram Image Processing

The general of processing a mammogram image is shown in Figure 2. It summarizes a few methods currently available for mammogram image processing.

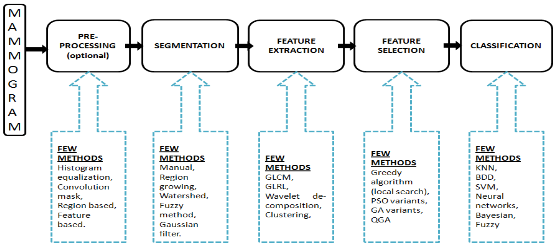

MAMMOGRAM
PRE-PROCESSING (optional)
SEGMENTATION
FEATURE EXTRACTION
FEATURE SELECTION
CLASSIFICATION
FEW METHODS Histogram equalization, Convolution mask, Region based, Feature based.
FEW METHODS Manual, Region growing, Watershed, Fuzzy method, Gaussian filter.
FEW METHODS GLCM, GLRL, Wavelet de-composition, Clustering,
FEW METHODS Greedy algorithm (local search), PSO variants, GA variants, QGA
FEW METHODS KNN, BDD, SVM, Neural networks, Bayesian, Fuzzy

Figure2.Generic flow of mammogram image processing

The images are obtained from the mini MIAS database and the ROI are manually segmented. Local Binary Pattern is used as a preprocessing step followed by GLCM feature extraction. The features are selected using a hybrid GAPSO-KNN approach and are classified into normal and abnormal mass.

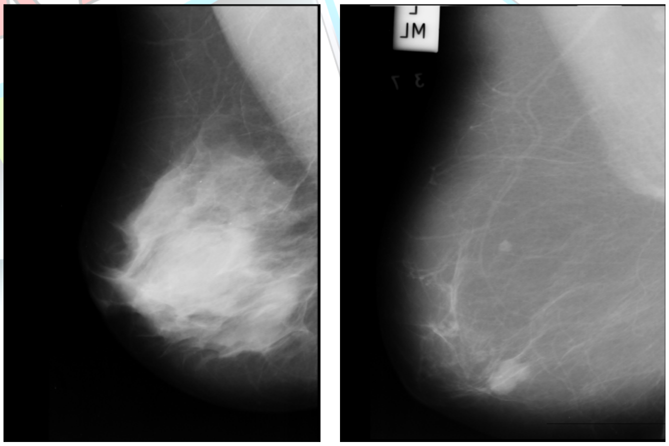
III. MIAS Database

The mammogram images were obtained from the mini MIAS-Mammographic Image Analysis Society database [12]. Every image is of 1024 pixels x 1024 pixels dimension.

The database has the following details

1. MIAS database reference number
2. Class of abnormality present
3. Severity of abnormality;
4. (X,Y) image-coordinates of centre of abnormality
5. Approximate radius (in pixels) of a circle enclosing the abnormality.

Figure 3 shows a normal and abnormal image obtained from the Mini MIAS database. The abnormality can be CALC – Calcification, CIRC - Well-defined/circumscribed masses, SPIC - Spiculated masses, MISC - Other, ill-defined masses, ARCH - Architectural distortion, ASYM – Asymmetry, NORM – Normal.



(a) **(b)**

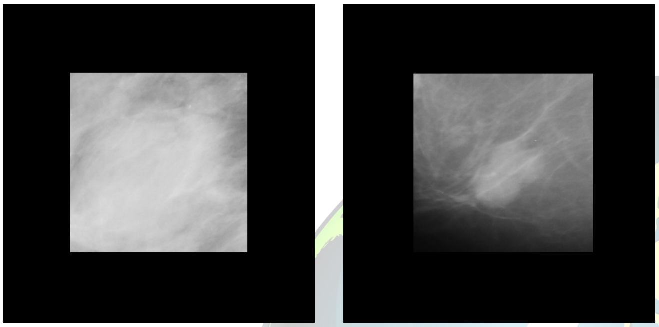
Figure 3. Images obtained from Mini MIAS database. (a)normal(mdb001.pgm) (b) abnormal(mdb005.pgm)

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

IV. Segmentation And Preprocessing

A. Segmentation

The ROIs of the mammograms are manually segmented using the description of image for abnormal (presence of tumor) patients. For normal mammograms, the ROI is randomly selected. Figure 4 shows the segmented ROI of normal and abnormal images shown in figure 3.



(a) (b)

Figure 3: ROI segmented from mammograms. (a)normal(mdb001.pgm) (b) abnormal(mdb005.pgm)

B. Preprocessing

Before extracting the features from the ROI, a Local Binary Pattern (LBP) texture operator is used to highlight the textural features. LBP is an illumination invariant texture feature that is computed separately for every image pixel [13]. The Local Binary Pattern (LBP) texture operator is shown in Figure 5. The original 3x3 neighborhood is thresholded by the value of the center pixel. The values of the pixels in the thresholded neighborhood are multiplied by the weights given to the corresponding pixels. Finally, the values of the eight pixels are summed to obtain the LBP number for this neighborhood. In rotation invariant classification, it is advantageous to interpolate the values of the corner pixels in order to obtain a circular sampling of the neighborhood.

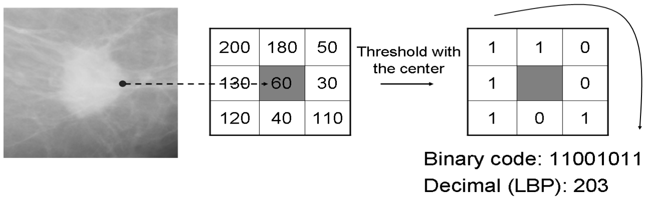
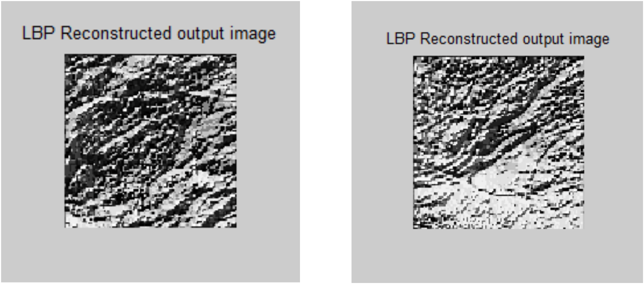


Figure 5: Example of basic LBP operator


LBP Reconstructed output image
LBP Reconstructed output image

(a) (b)

Figure 6: LBP applied to ROI. (a)normal(mdb001.pgm) (b)abnormal(mdb005.pgm)
Figure 6 shows LBP operator applied to ROI shown in figure 3

V. Feature Extraction Using GLCM

The texture features are extracted from the segmented ROI using Gray Level Co-occurrence Matrix method (GLCM) method. GLCM is the texture feature was proposed by Haralick *et al* in the 1970s [18]. It is well-established robust statistical tool for extracting secondary-order of texture information from image. The GLCM presents the joint frequencies of all pair-wise combinations of gray levels *i* and *j* in a specified direction θ and specified distance *d* from each other. The GLCM can be defined by equation (1) as where *(x1, y1)* and *(x2, y2)* are pixels in the ROI, *I(·)* is gray-level of pixels, and · is the number of the pixel pairs that satisfy the conditions. Given by equation,

$$C(i,j) = \left\| \left\{ \left[(x_1, y_1), (x_2, y_2) \right] \right\} \left| \begin{array}{l} x_2 - x_1 = d\cos\theta \\ y_2 - y_1 = d\sin\theta \\ I(x_1, y_1) = i \\ I(x_2, y_2) = j \end{array} \right. \right\| \qquad (1)$$

Each co-occurrence matrix is normalized by sum of all elements in matrix. Finally twenty two features are generated from each matrix. The features studied are autocorrelation, contrast, correlation I, correlation II, cluster prominence, cluster shade, dissimilarity, energy, entropy, homogeneity I, homogeneity II, maximum probability, sum of squares, sum average, sum entropy, sum variance, difference variance, difference entropy, information measure of correlation I, information measure of correlation II, inverse difference normalized and inverse difference moment normalized [14].

VI. Feature Selection Using GAPSO

The best feature subset is to be selected from the extracted features. Genetic algorithms (GAs) and particle swarm optimization (PSO) are population based heuristic search algorithms that have been used in feature selection. When compared to GA, PSO is easier to implement, and computationally less expensive and can converge quickly. Many PSO based feature selection techniques have been used with machine learning datasets. Recently the classification of mammogram micro-calcifications is done using PSO based feature selection methods. The use of PSO based feature selection in mammogram mass classification is rare. A hybrid approach called GA-PSO is expected to have merits of PSO with those of GA. To prevent the premature convergence, position update of the global best particles is changed. By applying crossover operation, information can be swapped between two particles to have the ability to fly to the new search area. The purpose of applying mutation to PSO is to increase the diversity of the population and the ability to have the PSO to avoid the local minima [18].

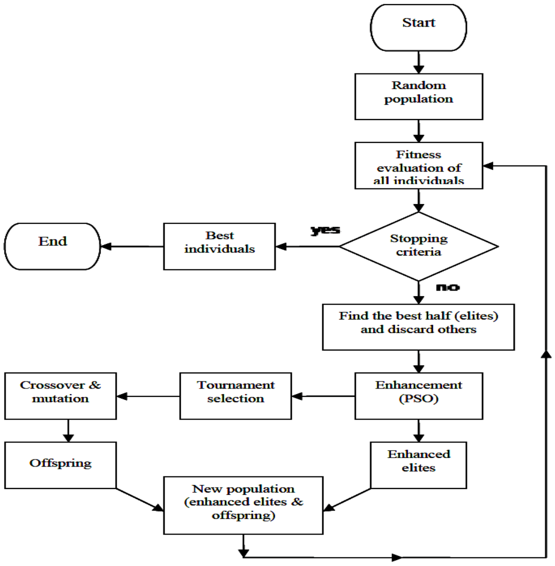


Figure 7: Workflow of hybrid GA-PSO algorithm

The working of the GA-PSO algorithm is shown in figure 7. The fitness values of individuals in the current population are found [19].

Enhancement: In each generation, after the fitness values of all the individuals in the population are calculated, the top-half best performing ones are marked. These individuals are regarded as elites. Instead of reproducing the elites directly to the next generation as elite GAs do, we first enhance the elites. The enhancement operation tries to mimic the maturing phenomenon in nature, where individuals will become more suitable to the environment after acquiring knowledge from the society. Furthermore, by using these enhanced elites as parents, the generated offspring will achieve better performance than those bred by original elites. The enhancement of the elites is performed by the velocity and position update procedures in PSO.

Selection: In GAPSO, the GA operations are performed on the enhanced elites achieved by PSO. In order to select parents for the crossover operation, the tournament selection scheme is used. Two enhanced elites are selected randomly, and their fitness values are compared to select the one with better fitness as a parent and place it in the mating pool. This scheme is used as the selection operator in the GA as well.

Crossover: Parents are selected randomly from the mating pool in groups of two and two offspring are created by performing crossover on the parent solutions.

Mutation: The final genetic operator is The ROIs are divided into three equal sets. Two sets are used as a training set and the remaining set as a test set. Feature selection by GAPSO is done using the training set only.

Then only the significant features obtained from feature selection are used to train the K-Nearest Neighbor (KNN) classifier, using the training set only. The trained classifier is then used to classify the test set, using the significant features only. The above process is repeated by using another set of data as a test set and the other two sets as a training set. Every ROI is used in the test set once only. The average classification accuracy of the three test sets is calculated. In GAPSO-KNN based feature selection, KNN is used to evaluate the feature subset in the training set. The classification accuracy of the feature subset on the training set is evaluated using KNN.

VII. Results And Discussion

In TABLE I, classification accuracy is defined as the number of correctly classified samples (to the class mass or non-mass) in the test set divided by the number of samples in the test set. The average number of features in the table is the average of the number of significant features found in the three different training set partitions

TABLE I. COMPARISON OF VARIOUS FEATURE SELECTION METHODS

Feature selection method	*Average Number of features*	*Average classification accuracy (%) in test set*

All features: No feature selection	18	79.71
GAPSO + KNN	6	86.81
PSO+KNN	6.3	81.16
Genetic algorithm+ KNN	6	82.07

From TABLE I, the GAPSO-KNN feature selection method has better classification accuracy than the methods and without feature selection (using all 18 features).

VIII. Conclusions

The experimental results show that the GAPSO-KNN feature selection method used in this paper can have comparable or better result than other widely used feature selection methods when it is applied to mammogram mass classification. By using texture features from GLCM alone, a small number of significant features found by GAPSO-KNN can have better performance in classification accuracy than the full set of features in mass classification.

References

[1]. Medindia. Breast cancer in India rising rapidly (2006). [Online] Availablefrom http://www.medindia.net/news/view_news_main. Jan 22, accessed on September 26, 2014.

[2]. Gajalakshmi V, Mathew A, Brennan P, Rajan B, Kanimozhi VC, Mathews A, et al. "Breast feeding and breast cancer risk in India: A multicenter case control study", Int'l Journal for Cancer, 125:662–5, 2009,.

[3]. Man To Wong, Xiangjian He, Hung Nguyen (), 'Particle Swarm Optimization Based Feature Selection in Mammogram Mass Classification', 2013.

[4]. Population based cancer registries consolidated report (1990-96) [Online]. Available from: http://www.icmr.nic.in/ncrp/pbcr.pdf . accessed on September 06, 2014

[5]. National cancer registry programme report (1981-2001) [Online]. Available from http://www.icmr.nic.in/ncrp/cancer regoverrview.htm . accessed on october 26, 2014

[6]. Siddiqi M, Sen U, Mondal SS, Patel DD, Yeole BB, Jussawala DJ, et al.(2001), 'Cancer statistics from non-ICMR registries: Population based registries'. CRAB (Cancer registry Abstract) Newsletter of the National Cancer Registry Project of India. 47–59.

[7]. Van den Bergh F. and Engelbrecht A.P., 'A Cooperative Approach to Particle Swarm Optimization', IEEE Transactions on Evolutionary Computation, 2004, pp. 225-239

[8]. K. Premalatha and A.M. Natarajan, "Hybrid PSO and GA for Global Maximization", Int. J. Open Problems Compt. Math., Vol. 2, No. 4, December 2009

[9]. Nandkumar A, Gupta PC, Gangadharan P, Visweswara RN, Parkin DM.(2005), 'Geographic pathology revisited: Development of an atlas of cancer in India'. IntJCancer.; 116:740–54.

[10]. H.D. Cheng, X. Cai, X. Chen, L. Hu, X. Lou (2003) 'Computer-aided detection and classification of microcalcifications in mammograms: a survey', Pattern Recognition 36, pp 2967 – 2991.

[11]. A.Oliver, J. Freixenet, J. Marti, E. Perez, J. Pont, E. Denton, R. Zwiggelaar (2010), 'A review of automatic mass detection and segmentation in mammographic images', Medical Image Analysis 14, 87–110.

[12]. J Suckling et al (1994) 'The Mammographic Image Analysis Society Digital Mammogram Database', Exerpta Medica. International Congress Series 1069 pp375-378.

[13]. Vinh Dinh Nguyen, Dung Duc Nguyen, Thuy Tuong Nguyen, Vinh Quang Dinh, and Jae Wook Jeon, 'Support Local Pattern and Its Application to Disparity Improvement and Texture Classification', IEEE Transactions On Circuits And Systems For Video Technology, VOL. 24, NO. 2, FEBRUARY 2014

[14]. A. Markkongkeaw, A. Phinyomar, P. Boonyapiphat, P.Phukpattaranont, (2013), 'Preliminary Results of Breast Cancer Cell Classifying Based on Gray-Level Co-occurrence Matrix', The 2013 Biomedical Engineering International Conference.

[15]. J. Tang, R. M. Rangayyan, J. Xu, I. E. Naqa, Y. Yang, (2009), 'Computer-Aided detection and diagnosis of breast cancer with mammography: recent advances,' IEEE Trans. on Information Technology in Biomedicine 13(2), 236-251.

[16]. Gao Jin ; Peng Jin-ye ; Li Zhan , 'Application of Improved PSO-SVM Approach in Image Classification' (2010), Photonics and Optoelectronic (SOPO) Symposium

[17]. Hela, B., Hela, M., Kamel, H., Sana, B., Najla, (2013) 'Breast cancer detection: A review on mammograms analysis techniques': Systems, Signals & Devices (SSD), 10th International Multi-Conference.

[18]. R. M. Haralick, K. Shanmugam, and I. Dinstein, (1973), 'Textural Features of Image Classification', IEEE Transactions on Systems, Man and Cybernetics, vol. SMC-3, no. 6.

[19]. Hybrid Genetic Algorithm And Particle Swarm Optimization For The Force Method-Based Simultaneous Analysis And Design Iranian Journal of Science & Technology, Transaction B: Engineering, Vol. 34, No. B1, pp 15-34,Printed in The Islamic Republic of Iran, 2010

[20]. J. Kennedy and R. Eberhart, (1995), 'Particle swarm optimization', Proceedings of the IEEE International Joint Conference on Neural Networks, Perth, Australia, vol. 4, pp. 1942-1948,.

[21]. I. Zyout, I.Abdel-Qader, (2011) 'Classification of microcalcification clusters via PSO-KNN heuristic parameter selection and GLCM features', International Journal of Computer Applications (0975 – 8887) Vol. 31 No.2.