Feature Selection and Mass Classification Using Particle Swarm Optimization and Support Vector Machine

Man To Wong[1], Xiangjian He[1], Wei-Chang Yeh[2], Zaidah Ibrahim[1,3], and Yuk Ying Chung[4]

[1] Faculty of Engineering and Information Technology, University of Technology Sydney, NSW 2007, Australia
[2] Department of Industrial Engineering and Engineering Management, National Tsing Hua University, Hsinchu 300, Taiwan
[3] Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Malaysia
[4] School of Information Technologies, University of Sydney, NSW, Australia
eemtwong@gmail.com, Xiangjian.He@uts.edu.au

Abstract. This paper proposes an effective technique to classify regions of interests (ROIs) of digitized mammograms into mass and normal breast tissue regions by using particle swarm optimization (PSO) based feature selection and Support Vector Machine (SVM). Twenty-three texture features were derived from the gray level co-occurrence matrix (GLCM) and gray level histogram of each ROI. PSO is used to search for the gamma and C parameters of SVM with RBF kernel which will give the best classification accuracy, using all the 23 features. Using the parameters of SVM found by PSO, PSO based feature selection is used to determine the significant features. Experimental results show that the proposed PSO based feature selection technique can find the significant features that can improve the classification accuracy of SVM. The proposed classification approach using PSO and SVM has better specificity and sensitivity when compared to other mass classification techniques.

Keywords: mass classification, support vector machine, particle swarm optimization, feature selection.

1 Introduction

Breast cancer is the most common cancer of women in America [1]. Mammography is the most effective method for early detection of breast cancers [2]. Masses are important early signs of breast cancer [3]. Mass detection in mammogram is difficult because the features of masses can be obscured and can be similar to normal breast parenchyma [4]. The results from a computer aided detection system can be used as a second opinion to a radiologist and improve the detection accuracy.

Many mass detection algorithms have the following two steps. In the first step, suspicious regions of interest (ROIs) are detected on the mammogram images by

C.K. Loo et al. (Eds.): ICONIP 2014, Part III, LNCS 8836, pp. 439–446, 2014.
© Springer International Publishing Switzerland 2014

using some image processing techniques such as segmentation or thresholding. In the second step, one typical approach is to extract features from the suspicious regions. Classifiers can then be applied on these features to classify the regions as mass or normal tissue. This will reduce the number of false positives. Sahiner et al. [6] used texture features and convolution neural networks in mass classification. He obtained 90% sensitivity and 69% specificity. Tourassi et al. [7] had applied template matching scheme based on the mutual information and obtained 90% sensitivity and 65% specificity. Christoyianni [8] used the GLCM [9] texture features and MLP and obtained 85% sensitivity and 83% specificity. Petrosian et al. [10] used the GLCM texture features and a modified decision tree classifier and obtained 76% sensitivity and 64% specificity. Angelini et al. [11] had tested and compared the performance of different image representations for mass classification. Instead of extracting features from the suspicious regions, the features are embodied by the image representation used to encode the suspicious regions. The best result was given by the pixel image representation, using SVM as classifier, with 90% sensitivity and 94% specificity.

The objective of this paper is to propose a novel feature selection and mass classification technique using SVM and PSO. The regions of interests (ROIs) are manually extracted from the MIAS Mini-Mammographic database [12]. The ROIs can contain mass or normal tissue. The ROIs will be classified as mass or non-mass regions using texture features calculated from the gray level co-occurrence matrix (GLCM) and statistical features from the gray level histogram. A PSO-based feature selection technique is proposed to select a smaller subset of significant features which can provide comparable or even better performance when compared to the full set of features.

2 Feature Selection Using PSO and SVM

Support Vector Machine (SVM) [13] is a classifier that has robust and accurate classification performance in many different applications. SVM finds the best hyperplane that separates the data by maximizing the margin between the hyperplane and the support vectors. The performance of SVM depends on the selection of kernel, the kernel's parameters, and cost parameter C. The RBF kernel is used in this paper. This kernel nonlinearly maps samples into a higher dimensional space and can handle the case when the relation between class labels and attributes is nonlinear. When RBF kernel is used, two parameters have to be properly chosen for good classification performance: the gamma (γ) parameter of the RBF kernel and the C parameter.

In this paper, the SVM software implementation in OpenCV [15] software library is used. The SVM in OpenCV is based on LIBSVM [16]. The C-Support Vector Classification (C-SVC) type and the RBF kernel of LIBSVM are used. According to the recommendation of [14], the feature values are linearly scaled to the range of [0,1]. The parameters C and γ (gamma) of SVM (using RBF kernel) are chosen by using PSO to search for C and gamma (γ) that can provide the best fitness function value of PSO. The fitness function used is the classification accuracy of SVM in the training set, using leave one out (LOO) cross validation.

PSO is a population based stochastic optimization technique modelled after the social behavior of bird flocks [17]. In PSO each particle represents a potential solution

to the optimization problem. Initially each particle is assigned a randomized velocity. Then the particles are flown through the problem space [17, 18]. The aim of PSO is to find the particle position with the best fitness function value.

Each particle keeps track of the following information in the problem space: x_i, the current position of the particle; v_i, the current velocity of the particle; and y_i, the personal best position of the particle which is the best position that it has achieved so far. This position yields the best fitness value for that particle. The fitness value of this position, called *pbest*, is also stored. In this paper, the *gbest* model of PSO is used. The best particle is determined from the entire swarm. The overall best value (*gbest*) obtained so far by any particle in the population and its location y_g are also tracked.

The velocity and position of the particle are given by equations (1) and (2) [18].

$$v_i(t+1) = wv_i(t) + c_1 r_1(t)\big(y_i(t) - x_i(t)\big) + c_2 r_2(t)(y_g(t) - x_i(t)) \quad (1)$$

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (2)$$

where w is the inertia weight, c_1 and c_2 are the acceleration constants, and $r_1(t)$ and $r_2(t)$ are random numbers generated in the range between 0 and 1.

Before feature selection, the parameters C and gamma (γ) of SVM, using the RBF kernel, are chosen by using PSO to search for values of C and γ that can provide the best fitness function value, using all the available features. The classification accuracy of SVM is used as the fitness function for PSO. In the training set, leave-one-out (LOO) cross validation is used. The LOO cross validation is especially suitable for small training set as it can maximize the use of training data. The two values $\log_2 C$ and $\log_2 \gamma$ are searched by PSO within the range from -10 to 10. Hence the actual range of C and γ that can be found in the search is from 2^{-10} to 2^{10}.

The original version of PSO described above operated in continuous space. The binary version of PSO (BPSO) has been developed for discrete problems [19] which can be used in feature selection. The velocity in BPSO represents the probability of an element in the position taking value 1. Equation (1) is used to update the velocity while x_i, y_i and y_g are restricted to 1 or 0. A sigmoid function $s(v_i)$ is used to transform v_i to the range of (0,1). BPSO updates the position of each particle according to the following formulae:

$$x_i = 1 \; if \; rand() < s(v_i), else \; 0 \; ; \quad s(v_i) = \frac{1}{1+e^{-v_i}} \quad (3)$$

rand( ) is a random number selected from a uniform distribution in [0,1].

In this paper, binary PSO (BPSO) is used to search for the feature subset in the training set. When x_i is 1, the feature corresponding to this bit position will be selected. When x_i is 0, the feature will not be selected. SVM classifier is used to evaluate the feature subset using LOO cross validation. The fitness function used in the proposed BPSO based approach is to maximize classification accuracy.

3 Texture Features

In Gray Level Co-occurrence matrix (GLCM), the texture context information is specified by the matrix of relative frequencies $P(i, j, d, \theta)$ with which two neighboring

pixels separated by distance d and along direction θ occur on the image; one pixel with gray level i and the other with gray level j [4,9]. After the number of neighboring pixel pairs R used in computing a particular GLCM matrix is obtained, the matrix is normalized by dividing each entry by R, the normalizing constant [9]. For each ROI, eight texture features were derived from each GLCM [5, 9, 10]. The notation $p(i,j)$ is used to represent the $(i, j)th$ entry in a normalized GLCM matrix and $p(i,j)$ is obtained by dividing each entry of the matrix $P(i, j)$ by R [9]. $\sum_{i,j}$ represents $\sum_{i=0}^{n-1}\sum_{j=0}^{n-1}$ where n is the number of gray levels per pixel.

$$Energy = \sum_{i,j} p(i,j)^2 . \tag{4}$$

$$Inertia = \sum_{i,j} (i-j)^2\, p(i,j) . \tag{5}$$

$$Entropy = -\sum_{i,j} p(i,j)\log\left(p(i,j)\right) . \tag{6}$$

$$Homogeneity = \sum_{i,j} \frac{1}{1+(i-j)^2}\, p(i,j) . \tag{7}$$

$$Max.probability = maximum\ of\ \ p(i,j) . \tag{8}$$

$$Cluster\ Shade = \sum_{i,j} (i+j-\mu_x-\mu_y)^3\, p(i,j) . \tag{9}$$

$$Cluster\ Tendency = \sum_{i,j} (i+j-\mu_x-\mu_y)^2\, p(i,j) . \tag{10}$$

$$Correlation = \frac{\sum_{i,j}(i-\mu_x)\left(j-\mu_y\right)p(i,j)}{\sigma_x\sigma_y} . \tag{11}$$

where μ_x , μ_y, σ_x and σ_y are the means and standard deviations of the marginal distributions associated with $P(i, j) / R$, and R is the normalizing constant [5, 9, 10].

In finding the GLCM, d is set to 1. Four directions are used for θ : 0, 45, 90 and 135 degrees. Then the average and range of the four values of each feature are calculated. The range is defined as the difference between the maximum and minimum of the four values. Hence a total of sixteen texture features are found for each ROI.

In addition to the GLCM features, seven statistical features are also derived from the gray level histogram of each ROI [8, 20]. The seven features are mean, standard

deviation, skew, entropy, smoothness, uniformity and kurtosis [8,20]. The equations for these seven features can be found in [20].

4 Experimental Result and Discussion

4.1 Mammogram Database and Test Method

The MIAS MiniMammographic Database is provided by the Mammographic Image Analysis Society in UK [12]. The mammograms are digitized at 200 micron pixel edge and have a resolution 1024 x 1024. The types of abnormality in the database include calcification, masses, architectural distortion and asymmetry. Mammograms which do not contain any abnormality (classified as normal) are also provided.

One hundred and twenty ROIs were manually extracted from the images in the MIAS database. The approach of extracting ROIs from the mammogram database is based on [11]. In the ground truth file of the MIAS database, the location of the center of the mass (if it exists) is given, together with the radius of circle which completely encloses the mass. A square crop centered on the location of each annotated mass is selected. The size of square crop is chosen so that the ratio between the crop area and the area of the annotated mass is approximately 1.3 . All the crops containing a mass are then resized to a fixed size of 128 x 128 pixels. The resizing of variable size ROI to a fixed size region has been used in other research paper on mass classification [11]. For the non-mass class (normal tissue), the 128x128 pixel regions are extracted randomly from the normal mammograms. 44 of the 120 ROIs contain mass and 76 of them contain normal tissue only. For ROIs which contain mass, the mass can be benign or malignant. Three types of masses were used in this paper: circumscribed, spiculated and ill-defined masses. For ROIs which contain normal tissue only, the ROIs are randomly chosen inside the breast body. Five-fold stratified cross validation is used in testing. The 120 ROIs are divided into five equal sets. Four sets are used as a training set and the remaining set as a test set. Hence there are 96 ROIs in the training set and 24 ROIs in the test set. Feature selection by BPSO-SVM is done using the training set only. Then only the significant features obtained from feature selection are used to train the classifier, using the training set only. The trained classifier is then used to classify the test set using the significant features. The above process is repeated by using another set of data as a test set and the other four sets as a training set. Every ROI is used in the test set once only. The average classification accuracy of the five test sets is calculated.

In BPSO-SVM based feature selection, SVM 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 SVM and LOO cross validation. Once the significant features have been found by the BPSO-SVM technique, only the significant features are used in the training set to train the classifier. Note that 5-fold cross validation is used to calculate the classification accuracy of the SVM on the test set while LOO cross validation is used to evaluate the feature subset found by BPSO-SVM in the training set. The PSO based parameters tuning for SVM and the BPSO-SVM feature selection method were implemented using C++ language and OpenCV software library [15]. The BPSO

based feature selection method is compared with other wrapper based feature selection methods which are all available in the WEKA machine learning workbench [13]. The wrapper subset evaluation technique used is SVM. The three different search techniques in WEKA library used to find feature subsets include stepwise forward selection, stepwise backward selection and best first search [13].

4.2 Experimental Result and Discussion

In Table 1, 2 and 3, the values of specificity, sensitivity and overall accuracy are all measured in the test set, using 5-fold cross validation. The notation "BPSO-SVM" refers to the proposed method in this paper. Sensitivity, specificity and overall accuracy are defined as follows [13]:

$$Sensitivity = \frac{TP}{TP + FN} \tag{12}$$

$$Specificity = \frac{TN}{FP + TN} \tag{13}$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{14}$$

where TP is the number of true positives, FN is the number of false negatives, TN is the number of true negatives and FP is the number of false positives. In Table 1, the proposed BPSO-SVM feature selection method has the best sensitivity, specificity and overall classification accuracy. In Table 2, except the proposed method, all the other classifiers shown were used to classify the test set without using feature selection. For the MLP, J48 and KNN classifiers, their implementations in the WEKA machine learning software library [13] are used. From Table 2, the proposed method BPSO-SVM gives the highest sensitivity and overall accuracy while its specificity performance is very close to KNN.

Table 1. Comparison of feature selection methods using SVM as classifier

Feature Selection Method	Specificity (%)	Sensitivity(%)	Accuracy (%)
BPSO-SVM	97.33	97.78	97.50
All Features	96.05	88.64	93.33
Stepwise forward search	96.10	85.84	92.50
Stepwise backward search	94.76	88.34	92.50
Best first search	96.10	88.06	93.32

Table 3 compares the performance of the proposed BPSO-SVM method with other existing mammogram mass classification techniques. The specificity and sensitivity of the proposed method in this paper are better than other existing methods.

Table 2. Comparison of classification methods using BPSO-SVM (with feature selection) and other classifiers without feature selection

Classifier	Specificity (%)	Sensitivity(%)	Accuracy (%)
BPSO-SVM + SVM	97.33	97.78	97.50
SVM (all features)	96.05	88.64	93.33
MLP	94.76	83.12	90.82
J48 (decision tree)	89.58	88.34	89.16
KNN (K=3)	97.42	86.40	93.34

Table 3. Comparison of proposed BPSO-SVM based classification and other existing mammogram mass classification techniques

Classification method	Specificity (%)	Sensitivity(%)
BPSO-SVM + SVM	97.33	97.78
Angelini et al. [11]	94.00	90.00
Christoyianni et al. [8]	83.05	86.66
Sahiner et al. [6]	69.00	90.00
Petrosian et al. [10]	64.00	76.00
Tourassi et al. [7]	65.00	90.00

5 Conclusion

The objective of this paper is to demonstrate the good performance of the proposed feature selection and mass classification approach using BPSO and SVM. PSO is used to search for the optimal parameters C and gamma of SVM, using the RBF kernel. Then BPSO-SVM feature selection technique is used to find the significant features in the training set. Finally SVM is used to classify the test set, using the significant features only. The experimental results show that the proposed BPSO-SVM feature selection method can have better result than other widely used feature selection methods when it is applied to mammogram mass classification. By using features from GLCM and gray level histogram, a small number of significant features found by BPSO-SVM can have better performance in classification accuracy than the full set of features in mass classification. Also the proposed mass classification approach has better performance when compared to other existing mass classification techniques. The proposed classification approach using PSO and SVM can achieve 97.78% sensitivity and 97.33% specificity on the test set using 5-fold cross validation.

References

1. Garfinkel, L., Catherind, M., Boring, C., Heath, C.: Change trends: an overview of breast cancer incidence and mortality. Cancer 74(1), 222–227 (1997)
2. Bovis, K., Singh, S., Fieldsend, J., Pinder, C.: Identification of masses in digital mammograms with MLP and RBF nets. In: Proc. of the IEEE-INNS-ENNS International Joint Conference in Neural Networks, pp. 342–347 (2000)

3. Cheng, H., Cai, X., Chen, X., Hu, X., Lou, X.: Computer aided detection and classification of microcalcifications in mammograms:a survey. Pattern Recog. 36, 2967–2991 (2003)
4. Cheng, H., Shi, X., Min, R., Hu, L., Cai, X., Du, H.: Approaches for automated detection and classification of masses in mammograms. Pattern Recog. 39, 646–668 (2006)
5. Eisa, M.M., Ewees, A.A., Refaat, M.M., Elgamal, A.F.: Effective medical image retrieval technique based on texture features. International Journal of Intelligent Computing and Information Science 13(2), 19–33 (2013)
6. Sahiner, B., Chan, H.P., Petrick, N., Wei, D., Helvie, M.A., Adler, D.D., Goodsitt, M.M.: Classification of mass and normal breast tissue: a convolution neural network classifier with spatial domain and texture images. IEEE Trans. Med. Imaging 15, 598–610 (1996)
7. Tourassi, G.D., Vargas-Voracek, R., Catarious, D.M., Floyd, C.E.: Computer-assisted detection of mammographic masses: a template matching scheme based on mutual information. Med. Phys. 30, 2123–2130 (2003)
8. Christoyianni, I., Dermatas, E., Kokkinakis, G.: Neural classification of abnormal tissue in digital mammography using statistical features of the texture. In: Proc. of the 6th IEEE Int'l Conf. on Electronics, Circuits & Systems, vol. 1, pp. 117–120 (1999)
9. Haralick, R.M., Shanmugam, K., Dinstein, I.: Texture features for image classification. IEEE Trans. Syst. Man Cybernet, SMC 3(6), 610–621 (1973)
10. Petrosian, A., Chan, H.P., Helvie, M.A., Goodsitt, M.M., Adler, D.D.: Computer-aided diagnosis in mammography: classification of mass and normal tissue by texture analysis. Physics in Medicine and Biology 39(12), 2273–2288 (1994)
11. Angelini, E., Campanini, R., Iampieri, E., Lanconelli, N., Masotti, M.: Testing the performances of Different Image Representations for Mass Classification in Digital mammograms. Int'l Journal of Modern Phys. C 17(1), 113–131 (2006)
12. The Mini-MIAS Database of Mammograms, `http://peipa.essex.ac.uk`
13. Witten, I.H., Frank, E.: Data mining: practical machine learning tools and techniques, 2nd edn. Morgan Kaufmann (2005)
14. Hsu, C.W., Chang, C.C., Lin, C.J.: A practical guide to support vector classification, `http://www.csie.ntu.edu.tw/~cjlin` (last updated on April 15, 2010)
15. Bradski, G., Kaehler, A.: Learning OpenCV, 1st edn. O'Reilly (September 2008)
16. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines. ACM Trans. on Intelligent Systems and Technology 2(3), Article No. 27 (2011)
17. Kennedy, J., Eberhart, R.: Particle swarm optimization. In: Proc. of the IEEE International Joint Conf. on Neural Networks, Australia, vol. 4, pp. 1942–1948 (1995)
18. Eberhart, R., Shi, Y.: Particle swarm optimization: developments, applications and resources. In: Proc. of the Congress on Evolutionary Computation, pp. 81–86 (2001)
19. Kennedy, J., Eberhart, R.: A discrete binary version of the particle swarm algorithm. In: IEEE Int'l Conf. on Syst., Man, and Cybernetics, vol. 5, pp. 4104–4108 (1997)
20. Islam, M.J., Ahmadi, M., Sid-Ahmed, M.A.: An efficient automatic mass classification method in digitized mammograms using artificial neural network. International Journal of Artificial Intelligence and Applications 1(3), 1–13 (2010)