

Classification of Clustered Microcalcifications in Mammograms using Particle Swarm Optimization and Least-Squares Support Vector Machine

Imad Zyout
Assistant Professor
Electrical Engineering Dept.
Tafila Technical University
Tafila, 66110, Jordan

ABSTRACT

Feature selection and classifier hyper-parameter optimization are important stages of any computer-aided diagnosis (CADx) system for mammography. The optimal selection for shape features, kernel parameter, and classifier regularization constant is crucial to achieve a good generalization and performance of least-squares support vector machines (LSSVMs). This paper presents a morphology-based CADx that uses a computationally attractive and unified scheme for accomplishing the model selection task. A heuristic parameter search based on particle swarm optimization (PSO) not only reduces the dimensionality of the input feature space but also optimizes hyper-parameters of the classifier. The performance of the proposed shape-based CADx including PSO-LSSVM parameter selection method is examined using 60 microcalcification clusters. Using different cross-validation procedures, the proposed PSO-LSSVM demonstrated a good generalization ability by producing classification accuracies higher than 92%. The best classification accuracy of 97% was obtained using the leave-one-out cross-validation procedure. Comparing the performance of PSO-LSSVM with PSO-SVM method that uses conventional SVM formulation, results demonstrated the attractive computational complexity and classification performance of PSO-LSSVM.

General Terms

Feature selection, Pattern recognition.

Keywords

Computer-aided diagnosis, Mammography, Microcalcification Clusters, Particle Swarm Optimization, Least squares support vector machines.

1. INTRODUCTION

Mammography continues to be the golden diagnostic imaging tool for breast cancer screening. However, human based interpretation of mammograms is a time-consuming, error prone-task, challenging, and irreproducible even when it is accomplished by an expert radiologist. Reading mammograms becomes more demanding and difficult when mammographic abnormalities being diagnosed are clustered microcalcifications (MCs). MCs are subtle mammographic findings shown on 30-50% of cancer diagnosed using mammography and are indistinguishable from surrounding when present in dense breast tissue. Presence of MCs is a key to an early detection

and diagnosis of breast cancer. These challenges and the nature of MCs lead to unsatisfactory positive predictive value (PPV) of mammography-based breast cancer diagnosis [1]-[2]. Hence, computer aided diagnosis (CADx) technology is very crucial to assist radiologist in reading mammogram, improving the PPV of mammography by providing radiologist with second opinion when double reading stage is not feasible most of the time [1].

Existing CADx systems [3]-[13] model the diagnosis of mammographic abnormalities including microcalcification clusters as a two-class pattern recognition problem accomplished in four steps, namely, preprocessing and region of interest selection, feature extraction, feature selection, differentiating between malignant and benign patterns using different supervised learning machines. Among these systems, CADx systems [4],[6]-[9],[11] combined shape features [3] of microcalcification clusters and kernel-based support vector machine (SVM) and neural network (NN) classifiers seem to be more effective. A common drawback of CADx systems [8]-[11], which used SVM for pattern classification, is the use of standard formulation of SVM learning [14]. A standard SVM classifier solves the learning problem that is a convex optimization with affine constraints using a quadratic programming (QP) method, which expected to be impractical for handling large-scale studies, optimization the classifier performance and estimating generalization ability using k-fold and leave-one-out cross-validation procedures, embedded feature selection, and hyper-parameter selection using exhaustive GA heuristic search [10].

Least-squares support vector machines (LSSVMs) introduced by Suykens and Vandewalle [15] convert convex optimization associated with the standard formulation of SVM learning into a linear system that can be solved by a matrix inversion [15]. This least square formulation [15,16] not only significantly reduces complexity of solving SVM learning problem but also provides an efficient mean for estimating the generalization ability of a nonlinear SVM classifier and for accomplishing feature selection using heuristic methods [17] more efficiently.

SVM learning, including both least square or standard formulations, guarantees a minimum generalization error on training examples but not necessary on unseen examples. Moreover, the feature extraction stage usually produces redundant and irrelevant features that causes a curse of dimensionality and so a poor generalization capacity of a

supervised learning machine. Therefore, it is very essential to select appropriate learning model, the best feature subset and classifier's parameters, to ensure an optimal generalization ability of the classifier. Instead of performing feature selection and classifier model selection task independently, an embedded feature selection approach efficiently incorporates feature selection during the classifier parameter optimization stage.

This paper presents a morphology-based CADx that employs an embedded feature selection and classification approach, called PSO-LSSVM, to optimize the generalization ability of the classification scheme by selecting optimal learning model that consists of the smallest and most discriminative shape features and optimal parameters of a kernel-based LSSVM classifier. The proposed PSO-LSSVM approach employs a particle swarm optimization (PSO) [18,19] heuristic search for accomplishing feature selection and for selecting the optimal parameters of a kernel-based LSSVM classifier. PSO and LSSVM techniques have been proven efficient for handling different optimization and pattern recognition problems [17]. Further, both LSSVM and PSO are simple and easy to implement, which offer tremendous computational savings compared to their counterparts GA and standard SVM, respectively. The main contribution of this paper is the use of PSO-LSSVM scheme for accomplishing feature selection and classification stages needed for morphology-based CADx of microcalcifications clusters. Further, this paper compares complexity and performance of the heuristic search based on LSSVM approach with that uses a standard SVM classifier.

The remaining sections of this paper are organized as follows: The theoretical background of pattern classification using LSSVM and heuristic parameter selection will be presented in Section 2. The proposed morphology-based CADx system will be presented in Section 3. Sections 4 and 5 will present the experimental results and conclusions, respectively.

2. BACKGROUND

2.1 Least squares support vector machine

In principle, least-squares support vector machine (LSSVM) [15,16] is similar to standard SVM [14] but the former accomplishes learning task by using least-square empirical error as an objective function while Hinge loss is used for the standard formulation of SVM learning [14]. This different loss function significantly affects modeling and complexity of the learning process in each method. The least-square objective function converts inequality affine constraints into linear ones, which modeled and optimally solved as a linear system. On the other hand, standard SVM formulation [14] models the learning problem as convex optimization with equality constraints. Such convex formulation commonly solved using more sophisticated solver such as quadratic programming (QP) [14].

For further understanding of least-square SVM formulation, consider a binary classification problem where a set of L training pattern $\{(x_1, y_1), (x_2, y_2), \dots, (x_L, y_L)\}$ with input pattern $\mathbf{x}_i \in \mathcal{R}^n$ and a class label $y_i = \pm 1$.

Suykens and Vandewalle [15] reformulated SVM learning problem as follows:

$$\min_{w, e} J(w, e) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + \frac{1}{2} \gamma \sum_{i=1}^L e_i^2 \quad (1)$$

Subject to the following equality constraints

$$y_i(\mathbf{w}^T \phi(\mathbf{x}_i) + b) = 1 - e_i, \quad i = 1, 2, \dots, L \quad (2)$$

where $\phi(\mathbf{x}_i)$ is a none-linear function that maps a pattern vector \mathbf{x}_i into higher dimensions and e_i is the error corresponding to the misclassification of pattern \mathbf{x}_i .

Comparing least-square formulation [16] with standard SVM formulation [14], three main differences are found; 1) equality constraints are used instead of inequality ones, 2) squared misclassification error e_i used as cost function instead of Hinge loss function employed for conventional SVM, and 3) parameter γ is equivalent to the regularization parameter C used for processing of misclassification due to overlapping training data.

Now, combining the cost function $J(w, e)$ and the set of equality constraints expressed as

$$L(w, b, e, \alpha) = J(w, e) - \sum_{i=1}^L \alpha_i \{y_i[\mathbf{w}^T \phi(\mathbf{x}_i) + b] - 1 + e_i\} \quad (3)$$

α_i is Lagrange multiplier that can be either positive or negative while it must be positive in the standard SVM problem [14].

Applying Karush-Kuhn-Thucker (KKT) conditions [16] for optimality leads to the following

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{w}} = 0 &\Rightarrow \mathbf{w} = \sum_{i=1}^L \alpha_i y_i \phi(\mathbf{x}_i) \\ \frac{\partial L}{\partial b} = 0 &\Rightarrow \sum_{i=1}^L \alpha_i y_i = 0 \\ \frac{\partial L}{\partial e_i} = 0 &\Rightarrow \alpha_i = \gamma e_i \\ \frac{\partial L}{\partial \alpha_i} = 0 &\Rightarrow y_i[\mathbf{w}^T \phi(\mathbf{x}_i) + b] - 1 + e_i = 0, \quad i = 1, 2, \dots, L \end{aligned} \quad (4)$$

By re-arranging KKT conditions, the following linear system of equations obtained

$$\begin{bmatrix} 0 & \mathbf{y}^T \\ \mathbf{y} & \Omega + \gamma^{-1} \mathbf{I} \end{bmatrix} \begin{bmatrix} b \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{1} \end{bmatrix} \quad (5)$$

where $\Omega = \mathbf{Z} \mathbf{Z}^T$ with $\mathbf{Z} = [\phi(\mathbf{x}_1) y_1, \dots, \phi(\mathbf{x}_L) y_L]$.

Using a kernel trick with a kernel function $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ that is a positive definite function satisfies mercer conditions, Ω is rewritten as $\Omega = y_i y_j K(x_i, x_j)$ $\mathbf{y} = [y_1, \dots, y_L]$,

$\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_L]$, and $\mathbf{1}$ is a L -length vector of all 1's.

Solving for dual space variables α_i and bias b [15], the

decision function of LSSVM learning machine classifies an input pattern \mathbf{X}_p as follows

$$y_p = \text{sign}\left(\sum_{i=1}^L \alpha_i K(\mathbf{X}_p, \mathbf{X}_i) y_i + b\right) \quad (6)$$

Several kernel functions are used in literature, the most popular kernel function is Gaussian or radial basis function (RBF) [11]. An RBF kernel function with bandwidth or control parameter σ is given by

$$K_{RBF}(x, y) = \exp\left(-\frac{(x - y)^2}{2\sigma^2}\right) \quad (7)$$

2.2 Embedded feature selection

An embedded feature selection approach includes feature selection stage as another hyper-parameter to be selected during optimizing the generalization performance of a classifier. Although conventional grid search is straightforward method to accomplish model selection or parameter tuning task, such approach is impractical and computationally expensive when applied to a real valued search space or when more than two parameters need to be optimized. More practical parameter optimization methods are analytic and heuristic techniques. The objective function of the model selection task (i.e. generalization error) is discontinuous with many local optima, which does not allow applying analytical methods directly. Instead, a smooth version of the objective function involved in the model selection commonly optimized using iterative gradient-based algorithms, which limits the optimality of the analytic solution.

2.3 Particle swarm optimization

Population based heuristic search techniques including genetic algorithms (GAs) and particle swarm optimization (PSO) [18, 19] is more suitable for solving parameter selection problems. Generally, a heuristic optimization technique does not guarantee an optimal solution in most cases. However, such approaches are still more attractive than analytical one because a heuristic search can escape from local minima [2], produce a near optimal solution, and perform very well even when search space is very large. Several studies have reported the efficacy of PSO over GAs and other evolutionary algorithms. Advantages of PSO based search are mainly the simplicity of the search process, ease of implementation that is demonstrated by the small number of parameters to be adjusted during initialization and search process. PSO also accomplishes an optimization task without mutation and cross-over operations, which are essential steps of GAs [17].

For a swarm of M particles, the velocity of the k th particle is updated using the best personal fitness \mathbf{x}_k^{pBest} and the best global fitness \mathbf{x}^{gBest} achieved so far as follows

$$\mathbf{v}_k(t+1) = w \cdot \mathbf{v}_k(t) + c_1 \cdot r_1 (\mathbf{x}_k(t) - \mathbf{x}_k^{pBest}) + c_2 \cdot r_2 (\mathbf{x}_k(t) - \mathbf{x}^{gBest}) \quad (8)$$

where w representing the inertia of the movement and is typically in the interval $[0, 1]$, r_1 and r_2 are random numbers between $[0, 1]$, and c_1 and c_2 are non-negative constants are the learning rates.

Using velocity just computed in (8), the particle location is updated as

$$\mathbf{x}_k(t+1) = \mathbf{x}_k(t) + \mathbf{v}_k(t+1) \quad (9)$$

2.4 Performance evaluation

Performance measures commonly used for evaluating CADx systems are classification accuracy, sensitivity (or true positive rate), specificity (true negative rate), and area under a receiver operating characteristic (ROC) curve. However, for small datasets the accuracy measure is more suitable. The classification accuracy as given by (10) is the ratio of correctly classified (true positive and true negative) samples to the total number of samples.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

Also, the false positive FPF and true positive TPF fractions are defined as follows

$$\begin{aligned} FPF &= \frac{TN}{TN + FP} \\ TPF &= \frac{TP}{TP + FN} \end{aligned} \quad (11)$$

where TN and TP are the numbers of true positive and true negative classifications, respectively. Also, FP and FN are the numbers of false positive and false negative classifications, respectively.

To estimate these measures, an independent test data must be available. Taking into consideration that such test data is unseen and have not been used in the feature selection and classifier training stages. A k-fold cross-validation procedure is usually used to partition the dataset into test and training data. Common values of k are 5 or 10. An extreme case is obtained when k is equal to the size of the dataset. In this case, k-fold is denoted as a leave-one-out (LOO) cross-validation.

3. MORPHOLOGY BASED CADx USING PSO-LSSVM

In this paper, malignancy of microcalcification clusters is characterized using a three-stage CADx scheme presented in Figure 1. A preprocessing step for the automated diagnosis of mammographic abnormalities (microcalcification cluster) is the region selection stage, where a region enclosing each MC cluster is extracted from an input image utilizing a ground truth data provided with each digital mammogram. As shown by Figure 1, the first stage is a morphology-based feature extraction stage that involves microcalcification segmentation [11]-[12] and morphology-based region analysis steps, which used 34 shape descriptors to represent each MC cluster. This study employs a shape feature extraction process developed and applied in our previous work [12]. In [12], shape feature extraction involves a microcalcification segmentation that was done using a top-hat morphological transform followed by shape analysis of segmented MCs. This process produces a set of 34 shape features used to characterize the malignancy of each MC cluster. In this study, an index F_i with the index i takes values between 1 and 34 is used to refer to shape

features from [12]. The set of shape attributes including region descriptors of individual microcalcifications (area, extent, solidity, compactness, equivalent diameter), boundary descriptors such as Fourier descriptor and normalized shape moments, and distribution of MCs in the cluster such as distance from a cluster centroid. The number of MCs and the number of single-pixel calcifications in MC cluster are also used to discriminate between benign and malignant one.

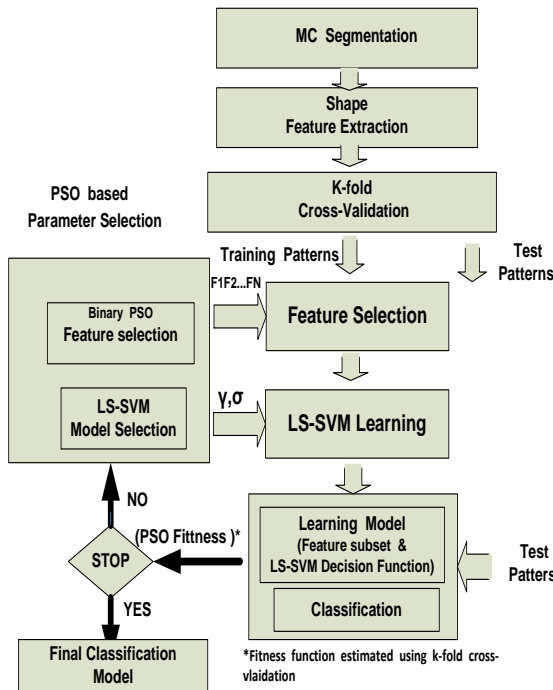


Fig 1: Characterization of MC clusters using morphology features and PSO-LSSVM hyperparameter selection

The shape feature extraction stage is followed by a heuristic parameter selection stage based on PSO in which feature selection, dimensionally reduction, and LSSVM parameter optimization are accomplished. The outcome of the parameter selection stage is the learning model that consists indexes of the selected shape features and optimal parameters of nonlinear LSSVM classifier. The discrimination between malignant and benign MC clusters is accomplished using a supervised a nonlinear least-squares support vector machines (LSSVM). Indeed, this paper uses a PSO-LSSVM embedded feature selection (or model selection) approach to present a unified and an efficient approach for accomplishing feature selection and classifier optimization steps, which are very important for improving generalization ability of the proposed CADx.

4. EXPERIMENTAL RESULTS

4.1 Mammographic test data

The proposed morphology-based CADx and PSO-LSSVM hyper parameter selection methods are tested using 60 MC clusters. This dataset is a mixed one that consists of two sets of MC clusters. The first set included 25 clusters (12 malignant and 13 benign) from a mini-mammographic

database provided by Mammographic Image Analysis Society (min-MIAS) [20]. The second group contains 35 MC clusters (17 malignant and 18 benign), which obtained from 30 digital mammograms produced by a full field digital mammography [12]. This paper presents a morphology-based characterization of microcalcification clusters that uses only shape features to discriminate a benign from malignant cluster. Hence, differences in spatial and contrast resolutions between MC clusters from different datasets will not impact of the classification result but such heterogeneous nature of the patterns make the test data more realistic and could improve the generalization ability of the classifier.

4.2 Parameter setup

The proposed heuristic parameter selection based on the PSO-LSSVM approach has adopted an objective function [12] that optimized the generalization ability (classification accuracy) of the proposed CADx by selecting the best and smallest subset of shape features and by selecting the optimal learning model of the LS-SVM classifier. The learning model of the kernel LSSVM classifier consists of the classifier regularization constant (γ) and the RBF kernel function control parameter (σ). As for the setting of PSO algorithm, the size of swarm was chosen to be 100 particles. The dimension of each PSO particle is 36 coordinates of which 34 coordinates allocated for the feature selection process and two coordinates for classifier parameter selection task. The population of the swarm was properly initialized assuming that each parameter belongs to a random variable that is uniformly distributed in the corresponding search space. As for the control parameters of heuristic search process using PSO, c_1 and c_2 were both set to 2, inertia constant ω monotonically decreased from 1.2 to 0.4 as the number of iteration increases. PSO search process was terminated if a maximum number of iteration 50 was reached or a predefined fitness value of 1.0 was achieved, whichever met first. Further, the CADx scheme including features extraction, PSO-LSSVM parameter selection and classification were implemented in MATLAB software.

4.3 Results analysis

Results of applying PSO-LSSVM with RBF kernel for differentiating between benign and malignant mammographic MC clusters using their morphology are presented in Table 1. As shown in Table 1 using a leave-one-out cross-validation procedure the proposed PSO-LSSVM heuristic parameter selection approach has achieved a classification accuracy of 100%. This perfect performance is obtained using 13 shape features, which are also listed in Table 1. By examining the learning model that produces the perfect performance, two weaknesses are observed. These challenges are a very high regularization constant ($\gamma = 3,634,116$) and relatively complex feature space that consists of 13 shape features.

Results also demonstrate that some learning models are expected to be more robust and may provide better generalization ability of the classification scheme. For instance, a learning model with $\gamma = 481.13$ and 7 shape features achieves a classification accuracy of 95%, which equivalent to true positive fraction of 1.0 and a true negative fraction (TNF) of 0.9. Further, results indicate that a classification accuracy of 90% is obtained using only 4 features (the smallest feature space) and reasonable values of the kernel and classifier parameters. Clearly, shape features

Table 1. Results of MC cluster classification using PSO-LSSVM and LOO cross-validation

γ	σ	N	Shape Features	TPF	TNF	Accuracy	Fitness
604.63	5.55	6	F9,F18,F23,F27,F31,F33	1.0	0.87	0.93	0.2481
481.13	5.67	7	F9,F18,F23,F24,F27,F31,F33	1.0	0.90	0.95	0.2429
928.68	4.83	9	F4,F7,F8,F15,F16,F23,F26,F31,F32	0.93	1.0	0.97	0.2642
3634116	13.28	13	F3,F6,F7,F8,F13,F17,F19,F23,F26,F29,F30,F31,F32	1.0	1.0	1.0	0.2974
229.84	3.19	4	F9,F18,F23,F27	0.90	0.90	0.90	0.2586

Table 2. Best classification results of PSO-LSSVM using different cross-validation procedures

Cross-validation	γ	σ	N	Shape Features	TPF	TNF	Accuracy
LOO	481.13	5.67	7	F9,F18,F23,F24,F27,F31,F33	1.0	0.90	0.95
10-Fold	6529.11	9.09	7	F9,F14,F18,F23,F27,F31,F33	1.0	0.87	0.93
5-Fold	40500.84	13.10	7	F7,F9,F16,F23,F27,F31,F33	0.97	0.87	0.92

for this model are also included in the feature space of the learning model that produced the best classification performance.

Although the leave-one-out cross-validation procedure represents unbiased estimator of the generalization and it is more suitable for small-scale studies, it has a risk of over-fitting the dataset. To further examine the generalization power of the proposed PSO-LSSVM classification approach, the results are evaluated using k-fold cross-validation procedures. Evaluation results using 10-fold and 5-fold cross-validation are presented in Table 2, which demonstrates a good level of robustness and generalization that using PSO-LSSVM produces classification accuracy of 92% or more using different training-testing schemes.

The paper presents a multivariate feature selection approach that uses an embedded framework to evaluate candidate features based on their joint discrimination rather than the individual power of each feature. In this work, PSO-LSSVM heuristic parameter selection was executed 280 times. Features selected by 146 runs, which produce classification accuracies higher than 90% were examined to determine the most discriminant shape features. This analysis demonstrated that some shape features more frequently and significantly contributed to the best learning models. These shape descriptors are: standard deviation of region equivalent diameter (F7), standard deviation of region solidity (F9), standard deviation of region compactness (F23), average and standard deviation of distances from a cluster centroid (F26 and F27), and maximum of the fourth normalized boundary moments (F32).

Using the same test data, this study has compared performance and computational complexity of the training process using PSO-LSSVM with those of PSO-SVM approach. As for parameter selection and classification performance, results using different cross-validation methods indicate the effectiveness and good generalization ability of both PSO-LSSVM and PSO-SVM approaches. That is both

classification approaches have achieved classification accuracies higher than 90%.

The computational complexity of SVM and LSSVM classifier based on the execution time required for accomplishing the training stage of the classifiers is compared and presented in Table 3. As shown in Table 3, the less computation time required for LSSVM training demonstrate the power of least-square formulation of SVM over the conventional SVM formulation. The efficiency of LSSVM was clearer in case of leave-one-out cross-validation procedure where the size of the training set is the largest and equal to 59 samples. In case of LOO cross-validation, training of SVM took about 69.7 milliseconds compared to 8.3 milliseconds. It is worth noting that the power of LSVM training becomes very visible and more attractive when accomplishing PSO-based heuristic search where SVM training must to be accomplished for every particle of the swarm.

Table 3. Average execution time of SVM training (in milliseconds)

Classifier	Cross-validation		
	LOO	10-fold	5-fold
SVM	69.7	39.5	33
LSSVM	8.3	6.2	5.2

5. CONCLUSIONS

This paper has discriminated between malignant and benign MC clusters in mammograms using shape features and PSO-LSSVM feature selection and classification approach. The proposed PSO-LSSVM employed a heuristic search via PSO to accomplish the feature selection process in parallel with appropriate selection of the kernel and regularization parameters of a kernel-based LSSVM classifier. The proposed parameter selection using heuristic PSO-LSSVM method provided an efficient and a reliable framework for optimizing the performance and generalization ability of kernel-based

LSSVM classifier. Compared to a conventional SVM method, the LSSVM not only simplifies the computational complexity of the learning process, but also it produces a very competitive classification performance.

Results of evaluating the proposed PSO-LSSVM method using 60 MC clusters, described by 34 shape features, indicated that PSO-LSSVM effectively selected the best shape features, reduced the dimensionality of the input feature space, and determined the optimal classifier parameters that produced the desired classification performance. Using different training-testing (cross-validation) scenarios, the good generalization ability of PSO-LSSVM approach was demonstrated by producing classification accuracies higher than 90%.

Further improvement of this work includes validating the proposed CADx including feature selection and classification methods using a large dataset. The extension of PSO-LSSVM parameter selection method to other feature spaces and pattern recognition problems is also another potential future research direction.

6. REFERENCES

- [1] Nishikawa, R. M., "Current status and future directions of computer aided diagnosis in mammography", *Computerized Medical Imaging and Graphics*, 2007, 31, 224-235.
- [2] Elter, M. and Horsch, A., "CADx of mammographic mass and clustered microcalcifications: A review". *Medical Physics*, 2009, 36(6), 2052-2068.
- [3] Shen, L., Ranayyan, R. M., and Desautels, J. E. L., "Application of shape analysis to mammographic calcifications", *IEEE Transactions on Medical Imaging*, 1994, 13(2), 263-274.
- [4] Jiang, Y., Nishikawa, R. M., Wolverton, Metz, C. E., Giger, M. L., Schmidt, R. A., Vyborny, C. J., and Doi, K., "Malignant and benign clustered microcalcifications: automated feature analysis and classification", *Radiology*, 1996, 198, 671-678.
- [5] Chan, H. P., Sahiner, B., Lam, K. L., Petrick, N., Helvie, M. A., Goodsitt, M. M., and Adler, D. D. 1998. Computerized analysis of mammographic microcalcifications in morphological and texture feature spaces, *Medical Physics*, 2007-2019.
- [6] Zadeh, H. S., Nezhad, P. S., and Rad, F. R. 2001. Shape based and texture-based feature extraction for classification of microcalcifications in mammograms. In *Proceedings of SPIE Medical Imaging*, 4322, 3010-310.
- [7] Kallergi, M., "Computer-aided diagnosis of mammographic microcalcification clusters", *Medical Physics*, 2004, 31(2), 314-326.
- [8] Wei, L., Yang, Y., Nishikawa, R. M., and Jiang, Y., "A study on several machine-learning methods for classification of malignant and benign clustered microcalcifications", *IEEE Transactions on Medical Imaging*, 2005, 24(3), 1278-1285.
- [9] Papadopoulos, A., Fotiadis, D. I., and Likas, A., "Characterization of clustered microcalcifications in digitized mammograms using neural networks and support vector machines," *Artificial Intelligence in Medicine*, 2005, 4(2), 141-150.
- [10] Wang, C., Jiang, W., and Dong, X., 2006. Characterization of clustered microcalcifications in mammogram based on support vector machine with Genetic algorithms. *IEEE International Conference Biophotonics, Nanophotonics and Metamaterials*, 114-117.
- [11] Zyout, I. 2010. Toward automated detection and diagnosis of mammographic microcalcifications. Doctoral dissertation, Dept. of Elect. & Comp. Eng., Western Michigan University.
- [12] Zyout, I., Abdel-Qader, I., and Jacobs, C. 2011. Embedded feature selection using PSO-KNN: application to shape-based diagnosis in mammography", *International Journal of Ubiquitous Systems and Pervasive Networks (JUSPN)*, 3 (1/2), 7-11.
- [13] Zyout, I. and Abdel-Qader, I. 2012 An improvement of texture-based classification of Microcalcifications in Mammography using PSO-SVM approach, 5th International conference on Communications, Computers, and Applications (MIC-CCA2012), Istanbul, Turkey.
- [14] Vapnik, V. 1995. *The Nature of Statistical Learning Theory*, Springer.
- [15] Suykens, J. A. K. and Vandewalle, J. 1999. Least Squares Support Vector Machine Classifiers", *Neur. Proc. Lett.*, 9(3), 293- 300.
- [16] Suykens, J. A. K., Gestel, T. Van, Brabanter, J. De, Moor, B. De, Vandewalle, J., *Least Squares Support Vector Machines*, 2002, World Scientific Publishing Co., Singapore.
- [17] Guo, X. C., Yang, J. H., Wu, G. C., Wang, C. Y., and Liang, Y. C. 2008. A novel LS-SVMs hyperparameter selection based on particle swarm optimization. *Neurocomputing*, 71, 3211- 3215.
- [18] Kennedy, J. and Eberhart, R. 1995. Particle swarm optimization. In *Proceedings of the IEEE International Conference on Neural Networks*, Perth: IEEE Service Center, Piscataway, NJ, 4, 1942-1948.
- [19] Kennedy, J. and Eberhart, R. C. 1997. A discrete binary version of the particle swarm algorithm. In *Proceedings of the Conference on Systems, Man, and Cybernetics*, Piscataway, NJ, 4104-4109.
- [20] Suckling, J., Parker, J., Dance, D., Astley, S., Hutt, I., Boggis, C., Ricketts, I., Stamatakis, E., Cerneaz, N., Kok, S., Taylor, P., Betal, D., and Savage, J. 1994. The mammographic image analysis society digital mammogram database. *Excerpta Medica*, 1069, 375-378.