Computer-Aided Diagnosis of Microcalcification Clusters Using Morphology Based Features and PSO-SVM Parameter Selection Approach

Imad M. Zyout

Department of Communications, Electronics and Computer Engineering, Tafila Technical University,
Tafila, Jordan
e-mail: izyout@ttu.edu.jo

Abstract— Computer-aided diagnosis (CADx) of mammographic microcalcifications (MCs) is designed to reduce the number of false biopsies by increasing the positive predictive value of the radiologist's interpretation. Although several algorithms were developed in the last two decades, the performance of these systems remains unsatisfactory and the need for developing efficient automated feature extraction and selection techniques is still high. In our attempt to address this demand, we propose a morphology-based CADx system for which we extract a set of 44 morphological features that describe the shape and the distributions of microcalcifications. In this paper, we present a heuristic model selection algorithm using a PSO-SVM framework that combines feature selection and SVM performance optimization steps. We also compare the performance of the feature selection using a binary PSO method against the outweighed nested-subsets method. To validate the proposed feature extraction and model selection methods, two datasets of microcalcification (MC) clusters have been used: the mini-MIAS and a digital mammography dataset from Bronson Methodist Hospital in Kalamazoo, Michigan. The obtained results demonstrate the effectiveness of the proposed CADx and indicate that a PSO-SVM framework using a binary PSO feature search method is more powerful than using an outweighed nested-subsets method.

Keywords— Feature selection, Mammogram, Microcalcifications, Morphology, Particle swarm optimization, Support vector machines.

I. INTRODUCTION

According to World Health Organization (WHO), breast cancer causes more than 500,000 women deaths each year. In Jordan, breast cancer, accounting for 36.7% of all female cancers, is not only ranked first among cancers afflicting women but also it is a leading cause of cancer's death. Mammography, an X-ray based medical imaging modality, is the most effective imaging tool for breast cancer screening [1]. A key sign of an early stage of breast cancer is the presence of granular clustered MCs, which are tiny deposits of calcium. Compared to other breast abnormalities, MCs appear more frequently on mammograms and represent an early sign to 30-50% of breast cancers diagnosed using mammography. However, discrimination between malignant and benign clustered microcalcifications is a challenging and error-prone task. This leads to a relatively low positive predictive value (PPV) of mammography interpretations [2], [3]. According to the literature, 10-30% of breast tumors are misclassified during routine mammography screening; and only 15-30 % of the cases recommended for invasive breast biopsies are found positive.

Mammography CADx systems are intended to help radiologists differentiate between malignant and benign breast abnormalities. That is to characterize the malignancy of clustered microcalcifications, which are tiny deposits of calcium and significant and common signs of the disease. Although numerous CADx approaches have been developed in the recent years, the development of CADx scheme with a satisfactory high PPV remains an open research question. In general, CADx approaches perform an automated diagnosis of mammographic MCs through four steps: mammographic region selection, feature extraction, feature selection,

and pattern classification. The malignancy of MCs is commonly characterized by using their morphology [2], [4], [6]-[12] or analyzing the texture of mammographic regions [8]-[10], [13], [14].

Feature extraction techniques may produce redundant or inadequate features, which result in a complex feature space and poor discrimination among different patterns. Hence, a feature selection process is a necessary stage to select a small subset of features that are more discriminating. Previous studies have selected the best feature subset and reduced the dimensionality of the original feature space using exhaustive heuristic search methods such as Genetic algorithms (GAs) [8]-[10],[13], linear discriminate analysis [8], sequential forward and backward selection methods [4], [12], and components analysis [15]. Other studies also presented a semi-automated feature selection method that eliminates weak features using univariate ranking and a rule-based expert system to search for additional discriminative features [11].

Microcalcification clusters CADx systems have been commonly modeled and solved as a binary classification problem accomplished using the supervised learning approach. The most popular classifiers used in previous CADx schemes are artificial neural network (ANN) [2], [4], [7], [11], [13], k-nearest neighbor (kNN) [9], [10], [13], [14], and the state of the art kernel based SVM [2], [11], [15], [16]. The other learning machines which were used in previous studies include statistical Bayesian [13], linear discriminate analysis (LDA) [6], [8], kernel fisher discriminate (KFD), relevance vector machine (RVM), and ensemble methods [2].

CADx systems that combine shape-based feature extraction and kernel-based SVM learning have proven to be more effective than the popular ANN. Several studies have demonstrated this result by applying both SVM and ANN learning machine to classify MCs within the same experiments (i.e. similar extracted features and mammograms) [2], [11]. However, the previous SVM based MCs diagnosis methods have several shortcomings and limitations, which include employing semi-automated techniques to perform segmentation of the individual MCs and feature selection [2] and [11]. The performance of SVM classifier was optimized using the conventional grid search selection [11], k-fold cross-validation [16], and exhaustive and computationally expansive heuristic search method using GA [15]. In addition, shape features extracted in [2], [11] were limited to the geometrical (e.g. region and distribution) descriptors and have not included other mathematical boundary descriptors such as normalized shape moment and Fourier descriptor. Yet, mathematical descriptors have demonstrated to be very effective in discriminating benign and malignant MCs [6].

This paper presents a four-stage CADx scheme. In the first stage of our shape-based CADx scheme, we segment the individual MCs using a morphological filtering scheme with dual filtering scales. In the second stage, 44 shape descriptors including measures of the region (e.g. area, compactness, eccentricity, and extent), distribution, and shape boundary are used to characterize each MC cluster. A heuristic model selection, or more specifically an embedded feature selection, using a PSO heuristic search method is mainly intended to integrate both processes of feature selection and the SVM classifier's model selection. Moreover, in this paper, we compare two methods to achieve the feature search process. The first method is based on a heuristic search using binary PSO technique to find an optimal feature subset, whereas the second method constructs a search space and feature subsets using an outweighed univariate-based nested subsets method. It is worth noting that the previous applications of PSO-SVM algorithm, presented in [1], [17] and [18], are different from this work. In [1], the PSO-SVM approach was used for accomplishing model selection (parameter and hyper-parameter selection) and reducing false positive results in computer-aided detection. In [17],

textural descriptors were used to characterize the malignancy of microcalcifications. The PSO-SVM approach in [18] used the least-squares SVM classifier, a fast and simple algorithm for solving the classification problem, instead of the conventional SVM classifier. The remainder of this paper is organized as follows: a theoretical background of the methods used in this study is presented in section 2; the proposed shape-based CADx is introduced in section 3; experimental results, discussion and conclusions are presented in sections 4 and 5, respectively.

II. BACKGROUND

A. Heuristic Parameter Search using PSO

Particle swarm optimization (PSO) [19], introduced by Eberhart and Kennedy in 1995, is a population based heuristic search approach inspired by the social behavior of the flocks of birds and the schools of fish, where a group of individuals (particles) located in the parameter space of an objective function search for the optimal solution. PSO search strategy uses the location of both best personal fitness x_k^{pBest} achieved by k^{th} particle and global fitness x^{gBest} to compute i^{th} dimension velocity and the new position of kth particle as follows:

$$v_{ki}(t+1) = w.v_{ki}(t) + c_1.r_1.(x_{ki}(t) - x_{ki}^{pBest}) + c_2.r_2.(x_{ki}(t) - x_i^{gBest}), \quad i = 1, 2, \ldots d \tag{1}$$

where d is the dimensionality of the kth particle; w is a constant, typically in the interval [0, 1], representing the inertia of the movement; r_1 and r_2 are random numbers between [0, 1]; and c_1 and c_2 are non-negative constants representing learning rates. To control the search speed, the ith velocity $v_{ki}(t)$ is constrained by the user to be in the range $[v_{min}, v_{max}]$. During the search process, the location of each particle is updated using the velocity computed in (1) as:

$$x_{ki}(t+1) = x_{ki}(t) + v_{ki}(t+1), i = 1, 2, \ldots d \tag{2}$$

B. Feature Selection Methods

Candidate features or feature subsets produced by various feature search techniques are commonly evaluated using feature filters, wrappers [3], [20] and embedded methods [21]. In this work, we opted to use an embedded feature selection technique that uses a heuristic feature search based on the particle swarm optimization (PSO) algorithm. PSO based heuristic search method is used instead of a genetic based algorithm since the former is proven to be a more computationally efficient and a very competitive alternative for GAs based methods [22], [23] and [24].

1) *Feature selection using binary PSO:* PSO based feature selection method [23], [25] is similar in principle to GA based method proposed by Seidlecki and Skalanski [27]. Each particle in the swarm represents a candidate feature subset coded as N-dimensional binary string with each component randomly assigned a value 0 or 1 [26]. Coordinates of each particle are assumed to be real valued random variables uniformly distributed between zero and one. Hence, this study converted the real representation of each particle into a binary string by assigning a binary 1 to all components larger than a statistical mean of all coordinates that is also a real number between 0 and 1. A binary 0 is also assigned to all coordinates less than this statistical mean. This binary conversion of the coordinates is different from the original binary PSO [25], [26], which compared a logistic transformation of

the new coordinate velocity $v_{ki}(t+1)$ with a random number between 0 and 1 to determine the new location $x_{ki}(t+1)$ of the corresponding coordinate.

2) *Outweighed univariate-based nested subsets method:* This method does not rely entirely on PSO method to create candidate feature subsets, but it adopts the nested subset method to generate *N* candidate feature subsets from the *N* features ranked individually using ROC analysis method. We follow this by an embedded feature selection procedure using PSO-SVM algorithm. A general shortcoming of forming different feature subsets using a single variable evaluation is the fact that truly redundant or highly correlated features may exist within subsets. Therefore, this study uses an average cross-correlation between a candidate feature and features already included as an additional criterion to control the redundancy level among selected features [28]. Such a process uses a real-valued u weight constant that can be set between 0 (discard the redundancy) and 1 (highest penalty) to penalize the ranking score of a potential feature if this feature shows a high correlation with others already in the subset.

C. Support Vector Machine (SVM)

The basic principle of pattern recognition using SVM is based on finding an optimal hyper-plane in the input feature space that maximizes separation (geometric margin) among the patterns from different classes [28]. Given input patterns $\mathbf{x} \in \Re^n$ that are two classes with a class label $y \in \{-1, 1\}$, the SVM learning problem is formulated as a convex optimization problem that is subject to a set of inequality and linear constraints [29], [30] written as follows:

$$\min_{w,\xi} J(\mathbf{w},\xi) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{L}\xi_i \tag{3}$$

subject to:

$$\mathbf{w}^T\mathbf{x}_i + b \geq 1 - \xi_i, \; i = 1,2,...,L$$

where ξ_i is a positive slack variable as introduced by Cortes and Vapnik [31]; *C* is a positive regularization or penalization parameter, which corresponds to a training error that must be adjusted during a model selection process; and *L* is the number of training samples.

Data, in general, is either nonlinearly separable or linearly separable in the original feature spaces. However, it can be linearly separated in a higher dimensional feature space. This higher dimensional feature space is usually obtained using a nonlinear mapping called kernel function $\Phi(\mathbf{x})$ [30], known as "kernel trick", which maps the original feature space into a higher dimensional feature space. The most common kernel functions are the Gaussian or RBF kernel $K(\mathbf{x},\mathbf{y}) = \exp[-(\mathbf{x}-\mathbf{y})^2/2\sigma^2]$ and the polynomial kernel, $K(\mathbf{x},\mathbf{y}) = (1+\mathbf{x}^T\mathbf{y})^P$ [30]. The real and integer control parameters σ and *P* need to be adjusted during SVM learning to optimize their generalization abilities.

Solving SVM dual optimization results in the decision function described by a set of Lagrange multipliers α_i and a bias constant b, which can be used to compute the class label $\tilde{y}_p$ of an input test pattern $\mathbf{x}_p$ as follows:

$$\widetilde{y}_p = sign(\sum_{i}^{L} \alpha_i K(\mathbf{x}_p, \mathbf{x}_i) y_i + b) \qquad (4)$$

III.   MORPHOLOGY BASED CADX OF MICROCALCIFICATIONS

Computer-aided diagnosis of MCs, if approved for the clinical use, can have a significant impact on the performance of the entire diagnosis process. This requires a careful design of the CADx scheme that produces almost a perfect diagnosis performance. Hence, one should not underestimate the impact of various components of the CADx scheme including shape feature extraction, feature selection and classification. The shape based diagnosis scheme as proposed in this study and illustrated in Fig. 1 segments MCs via a multiscale morphological filtering scheme. It also employs the radiologist's input (location and size of MC cluster) to automate region selection and improve the segmentation of MCs. We also employ several groups of shape descriptors to characterize the region, distribution, and boundary of individual MCs and their entire cluster. Our scheme also employs a PSO heuristic search technique to accomplish a model selection of the SVM classifier to optimize classification performance and generalization ability. This study also compares the performance of feature selection using univariate-based nested subset methods and heuristic search using a binary PSO method.

A.   *Morphological Based Segmentation*

Mathematical morphology is recognized to be a very effective tool in digital image processing; and is employed by many researchers for pre-filtering, enhancement, segmentation, and shape feature extraction [3], [33].
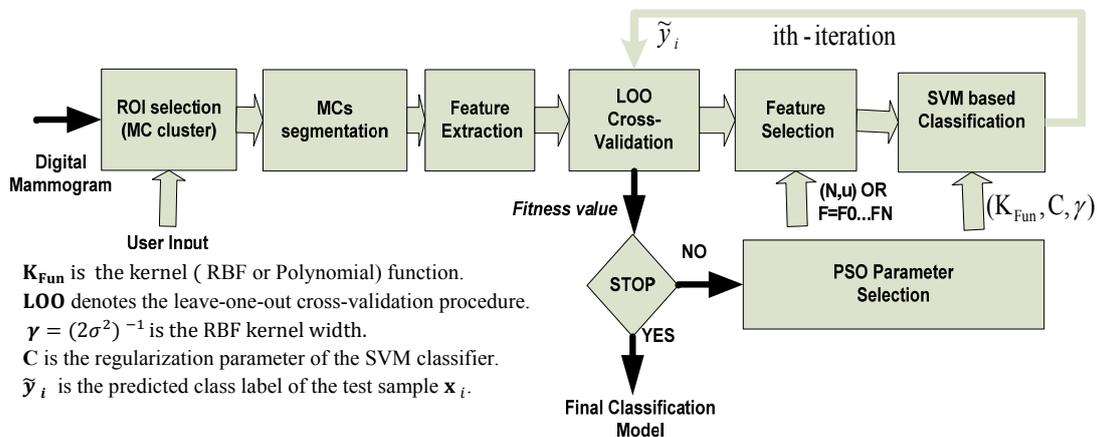

Fig. 1. Diagnosis of MCs using the shape based CADx. The user input represents the ground truth of each microcalcification cluster

Morphological image processing is based mainly on dilation and erosion operations [33]. Several studies have reported on the effectiveness of mathematical morphology for MC segmentation and detection [32], [34] and [35]. In these approaches, segmentation of MCs was accomplished by combining top-hat transform with tools such as Sobel and Canny edge detectors [32], difference of Gaussian filter [34] and watershed transform [35]. Mathematical morphology is effective because it can detect and segment bright objects and preserve their shape even when the gray-level of the surrounding region is inhomogeneous. This is precisely what makes morphological algorithms, such as watershed and top-hat transforms, excellent candidate algorithms for segmenting MCs and implementing shape based CADx.

1) *Segmentation of MCs using morphological filtering:* since it is difficult and impractical to conduct a subjective evaluation of the segmentation outcome, we used the overall performance of the classification scheme and the discriminative power of the extracted shape-descriptors to design and evaluate the proposed segmentation scheme [4]. After several experiments and performance evaluations of the extracted shape descriptors, we have proposed a new segmentation method illustrated by Fig. 2. This proposed scheme accomplishes MCs segmentation as logical combinations of the binary output of dual modified top-hat transform. A threshold computed using low order statistics (first and second moments) of the filtered region is applied to the output of each morphological filter bank to produce a binary image representing the segmented MCs.

As demonstrated in Fig. 2, the basic difference between the conventional top-hat transform and the proposed one is that the later applies additional morphological closing operations that smoothe the background image prior to its subtraction from the original image.

a)   b)
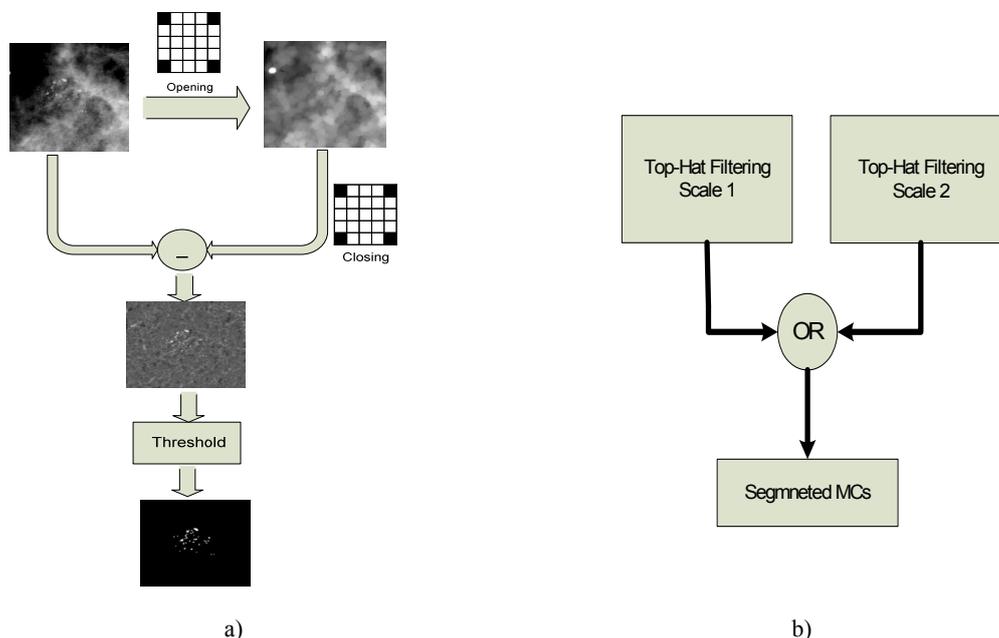
Fig. 2. The MC segmentation stage [28] a) illustration of a single scale modified top-hat morphological filtering stage, which smoothes an opened image via a closing operation using the same structure element, b) segmentation of MCs using a dual top-hat filtering scheme. The final segmentation result is obtained by applying logical sum (OR) operation to the results of each signal scale

Although one can employ more than two scales, our experimental results indicate that the two structure elements of size 5×5 and 7×7 can be effectively employed. An important step for an efficient supervised learning is the purity of the training examples which represent each class and require an efficient MC segmentation and post-processing step to reduce the number of false detected signals. In this study, we used a ground truth file accompanied with each mammogram that included the location and size of the region that best fits MC cluster to generate a binary mask and eliminate all detected signals located outside the rectangular region enclosed. A sample of the results of this process is shown in Fig. 3.

2) *Segmentation of MC clusters:* previous studies have demonstrated that analyzing the shape of the entire MC cluster can also be beneficial for distinguishing a malignant from benign cluster, [2] and [7]. In this work, a binary region representing an entire MC cluster is produced using successive applications of six morphological dilation operations to merge binary regions of the individual MCs into one region, which is adapted from [7]. Utilizing a prior knowledge such as the size of the ROI encloses each MC cluster and allows for an

accurate automated delineation of the cluster area. False detected MCs, located outside the actual cluster, might change the regularity of the shape of the cluster area and so alter the computation of related shape descriptors.
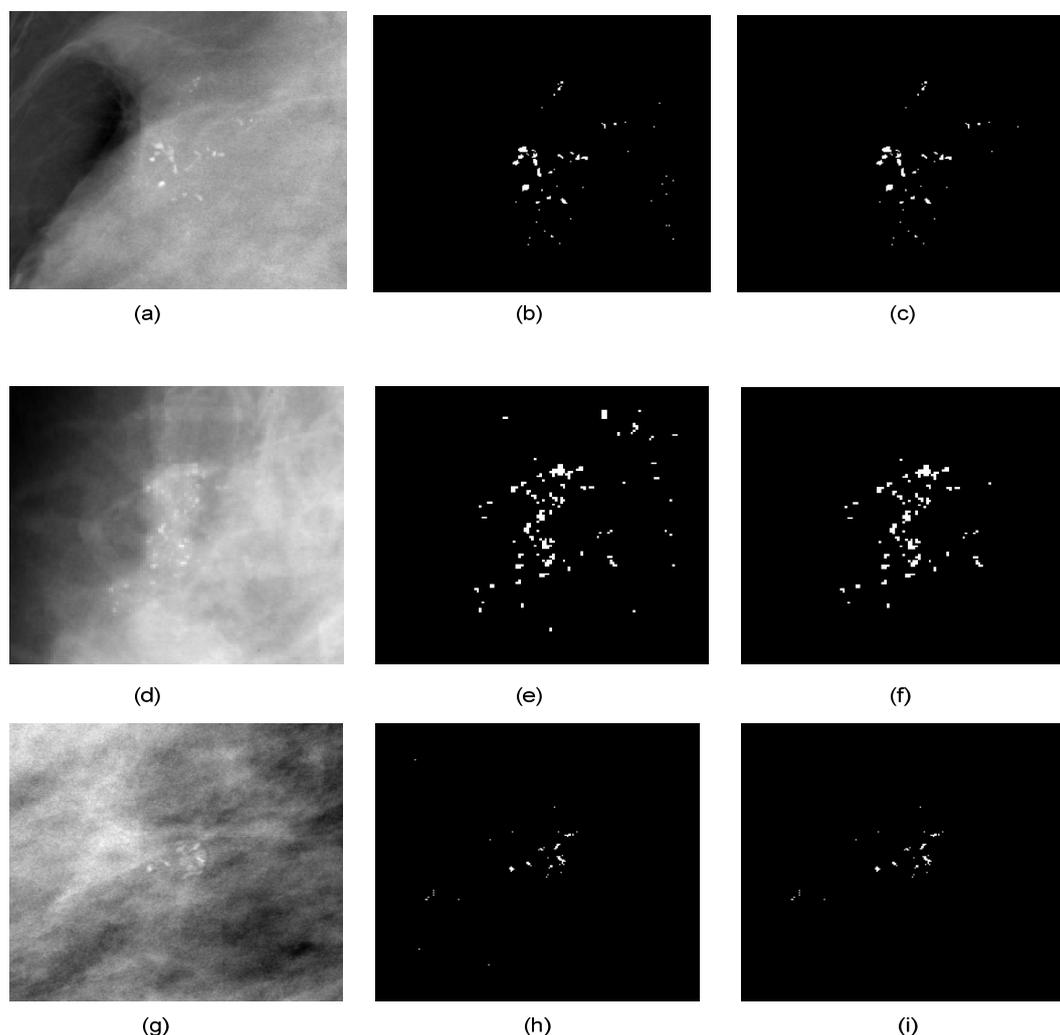

Fig. 3. Segmentation of microcalcification clusters. Original mammogram regions are shown in a, d and g. The segmentation results, before applying the binary mask, are in b, e and h. The improved segmentation, utilizing a binary mask corresponds to the microcalcification cluster best fitting region are shown in c, f and i

B. Shape Based Feature Extraction

Several shape features have been used for analyzing the malignancy of MC clusters. A summary of various shape descriptors and their importance can be found in [4]. This work introduces the criteria that have affected the development of the feature extraction stage and the selection of the shape features. Feature extraction is planned to model the radiologists' approach by using the morphology and distribution of MCs as the primary domain for the characterization and diagnosis of clustered microcalcifications [5], including various shape descriptors from the literature [4]-[12]. This study also avoided eliminating some descriptors based on their single variable evaluation that may help some shape descriptors to perform better when combined with others.

TABLE 1
EXTRACTED SHAPE FEATURES

No.	Feature Name	No.	Feature Name
1	Number of MCs in a cluster	23	SD- Compactness **
2	Number of single- pixel MCs	24	Mean- Perimeter of MCs**
3	Sum- Area	25	SD- Perimeter of MCs**
4	Mean- Area	26	Mean- Distances from a cluster's centroid
5	SD- Area	27	SD- Distances from a cluster's centroid
6	Mean- Equivalent diameter	28	Mean- F2'
7	SD- Equivalent diameter	29	SD- F2'
8	Mean- Solidity	30	Mean- F4'
9	SD- Solidity	31	SD- F4'
10	Mean- Eccentricity	32	Max- F4'
11	SD- Eccentricity	33	Mean- FF
12	Mean- Extent	34	SD- FF
13	SD- Extent	35	Area (MC)
14	Mean- Minor Axis length	36	Convex Area (MC)
15	SD- Minor Axis length	37	Eccentricity (MC)
16	Mean- Major Axis length	38	Circularity (MC)
17	SD- Major Axis length	39	Major Axis (MC)
18	Mean- Convex Area	40	Minor Axis (MC)
19	SD- Convex Area	41	Axis ratio (MC)
20	Mean- Orientation*	42	F2' (MC)
21	SD- Orientation*	43	F3'- F1' (MC)
22	Mean- Compactness **	44	FF- (MC)

* Single pixel MCs is excluded, ** Prior region up-scale is applied

Shape features, obtained from regions representing individual MCs as well as an entire cluster, are used to describe each MC cluster that is examined in this study. Extracted shape features are listed in Table 1 are mainly grouped into three subsets: region descriptors subset [33], [6], [10], [11], such as area extent [11], [15], compactness or circularity [6], [7], [9], [11] eccentricity [4], [11], solidity [11], [15], and equivalent diameter [11], [15]. The second subset contains the boundary descriptors that include normalized shape moments $F_1^{'}, F_2^{'}, F_3^{'}$, and $F_3^{'} - F_1^{'}$ or $F_4^{'}$ [4], [6], [10], [15], and normalized Fourier descriptors (FF) [4], [6], [10]. The third subset includes features describing the distribution of MCs in a cluster such as orientation [11], and spreading of MCs in the cluster [4], [11], [15], which is extracted from the binary region of each MCs and the whole cluster. Other features used in this work are neither region nor boundary descriptors such as the number of MCs [7], [10], [11] as well as the number of MCs represented by one pixel [10], [15]. Each MC cluster is modeled using 44 shape features as listed in Table 1. This feature set consists of 34 features obtained from individual MCs in each cluster; and the remaining 10 features describe the region and boundary of the entire MC cluster.

C. PSO-SVM Model Selection

The model selection stage presented in this paper is an algorithm that integrates feature and classifier's model selection processes. This integrated algorithm is more efficient in optimizing the performance of the given classification scheme than having to independently conduct the two selection processes. This effectiveness becomes obvious when feature selection and SVM classifier's model selection are accomplished using computationally expensive search techniques such as GA or exhaustive search [22]. Hence, we developed a heuristic based embedded feature selection scheme under a PSO-SVM framework that allows

for features search and selection during SVM learning process. In this work, we constructed the candidate set to the feature selection task using two methods, namely the nested subset and heuristic search methods. In addition, we used a leave-one-out (LOO) training and testing method to minimize the risk of data over-fitting and ensure the availability of unseen test patterns that have not been previously used in any training or feature selection stages.

Because the main purpose of the PSO-SVM model selection framework is to optimize classification performance and generalization capacity of the SVM classifier, each candidate solution for the model selection problem is composed of some parameters assigned into two subsets. While the first subset is used to search for the best subset of features, the second subset of parameters is intended to select the SVM learning model that leads to the optimal generalization ability. The dimensionality of the candidate solution, which is a PSO particle, is determined by the feature search method. For the outweighed univariate based nested subsets method, each candidate is represented by 5 coordinates, which include two parameters for feature search (an index feature subset N and average cross-correlation based penalization u) and a set of three parameters for SVM model selection. The SVM model selection parameters include the classifier hyper-parameter, kernel function (K_{Fun}), kernel's control parameter γ, and a classifier's regularization constant C. In addition to the three parameters used for the SVM model selection, a binary PSO feature search requires N parameters that are converted into a binary string of 44 bits to represent a potential feature subset. We define the fitness function as the leave-one-out generalization error of the classifier.

IV. EXPERIMENTAL RESULTS

A. Mammogram Datasets

The proposed four-stage CADx scheme, including shape feature extraction and model selection methods, has been tested using two mammogram datasets.

1) *mini-MIAS dataset:* This is a screen film mammography dataset provided by the Mammographic Image Analysis Society (MIAS) [36]. Each mammogram in the mini-MIAS dataset is of 1024 × 1024 pixels with 200μm pixel's size and 8-bit depth. The dataset contains 20 mammograms with 25 MC clusters (13 benign and 12 malignant). In addition, each mammogram has its ground truth file that specifies the size and coordinates of the centroid of the abnormal region (microcalcification cluster). We used this ground truth file as a radiologist's input to extract 128×128 regions centered at each cluster's centroid.

2) *Bronson Methodist Hospital (BMH) dataset:* This dataset consists of 30 digital mammograms of 100-μm pixel's size and 16-bit depth. These digital mammograms contain 32 MC clusters of which 17 are benign and 15 are malignant cases. We also used the ground truth file to determine the size of the region that best fits each MC cluster.

B. Computational Complexity and Experimental Setup

All segmentation, shape feature extraction, PSO heuristic parameter search, embedded features selection, and kernel based SVM classification methods are implemented using MATLAB version 7.9.0 (R2009b). Table 2 summarizes the execution times for MC segmentation, shape feature extraction, and embedded feature selection using PSO-SVM framework. The time for MC segmentation represented the average time required for segmenting MCs in each mammographic region. The cross-validation time is the average time needed for accomplishing LOO training and testing of each learning model. We mainly

focused on investigating the impact of shape feature extraction and feature selection. We did not attempt to optimize the computational complexity or speed up the execution time.

During the search process, the PSO algorithm constructs two sets of parameters. The first set of parameters is selected during the initialization stage and kept fixed for the whole search process. This set includes the size of the swarm (the number of particles), boundaries of the search space, and maximum and minimum velocities for each dimension, and the termination criterion, which can be selected as the number of iteration or desired fitness level (the average generalization error). The second set of parameters controls the movement of the particles and the PSO search process and includes $c_1, c_2, w, r_1,$ and r_2 [24].

All experimental results presented in this paper are obtained by using PSO heuristic search with a swarm of size 100 particles and termination criteria of either maximum iterations of 50 or a zero generalization error. Additionally, we choose search space limits to be individually selected for each coordinate. For example, the classifier's regularization constant C is a real-valued number between 1 and 10^5, while the kernel parameter σ is real-valued between 0.5 and 35 for RBF kernels and an integer P between 1 and 5 for the polynomial kernel. We also used the classifier's generalization error and the ratio of the number of falsely classified test patterns to the overall number of test patterns as a primary criterion for model selection and feature selection processes. The corresponding area under ROC or Az index was used as a secondary performance metric to evaluate the obtained models.

TABLE 2
MATLAB EXECUTION TIME, IN SECONDS, FOR VARIOUS STAGES OF CADX

Dataset*	MC Segmentation	Feature Extraction	LOO Cross-Validation
MIAS	0.07	15.50	0.40
BMH	0.17	58.87	0.51

*Datasets are different with respect to the number of cases and region size

C. Results on Mini-MIAS Dataset

1) *Impact of MCs segmentation:* In this study, MCs segmentation with reduced false detected signals and discriminative shape features has been accomplished by considering several design factors. These factors included designing the filtering scheme, utilizing ground truth data (location and size of MC cluster provided by MIAS) to improve the segmentation process, and employing different shape descriptors to characterize the region and boundary of the individual MCs, and the entire cluster and the distribution of MCs in the cluster. As for the filtering method, we used a modified top-hat transform that applies additional morphological closing (or smoothing) to the background before performing an image subtraction. Experimental results, as illustrated in Fig. 4a, indicated that the proposed morphological filtering achieved a better classification performance than segmentation using a standard top-hat transform. The impact of the selection of the threshold level which is used to produce a binary representation of MCs is presented in Fig. 4b. We have tested several threshold levels and used classification performance to select a threshold level that led to shape features with best discrimination between malignant and benign classes. Aiming to eliminate a false calcification outside the actual cluster region, a radiologist's input is used to construct a binary mask from the mammogram annotations and specify the size and centroid of the region that best fits each MC cluster. We also found this mask to be useful for segmenting the margin of the entire MC cluster.
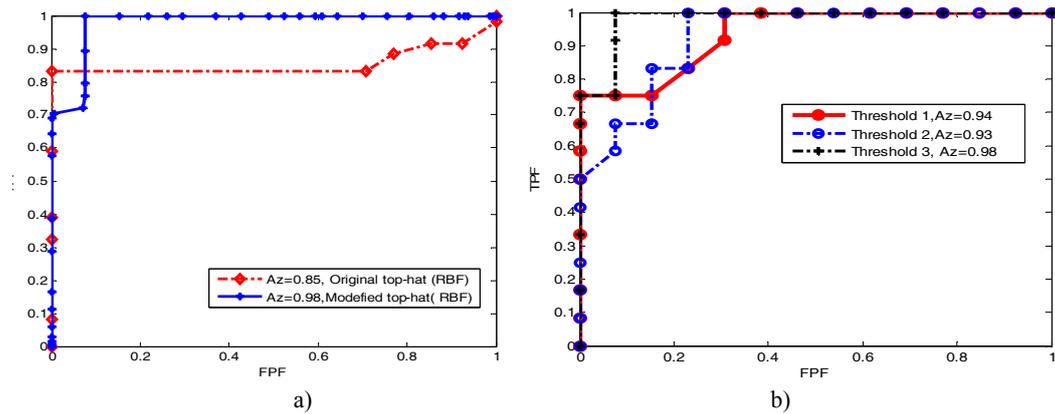
Fig. 4. Impact of the segmentation of the MCs on the classification performance, a) comparison of the performance of MC segmentation using the original top-hat and the modified top-hat transforms, b) the impact of different threshold's levels on the performance

2) *Results on feature selection:* The embedded feature selection, which integrates feature selection task with parameter's adjustment of SVM classifier, used two feature search strategies: outweighed univariate based nested subsets method and heuristic search using binary PSO technique. These methods are mainly different with respect to the complexity of the search process and the size of search space. Embedded feature selection using conventional univariate nested subset method (search is guided by a univariate ranking only) requires N evaluation of the classification performance, which is more computationally efficient than other sequential feature selection (SFS) and heuristic search methods. However, the simplicity of the search space using conventional univariate based nested subset technique will mostly miss an optimal feature subset and lead to a sub-optimal feature selection process. Using cross-correlation based outweighing scheme as an additional criterion can improve such a method. Experimental results, presented in Fig. 5, indicate that feature search using outweighed univariate-nested subsets method ($u = 0.33, 0.66$ and 1.0) generates more predictive feature subsets than the univariate ranking-nested subsets method ($u = 0$). The correlation level becomes more influential, when the feature subset has a small size (n is less than 20). Since no prior knowledge of the best size of feature subset and correlation level is available, it was essential to optimize this process as part of the model selection process.
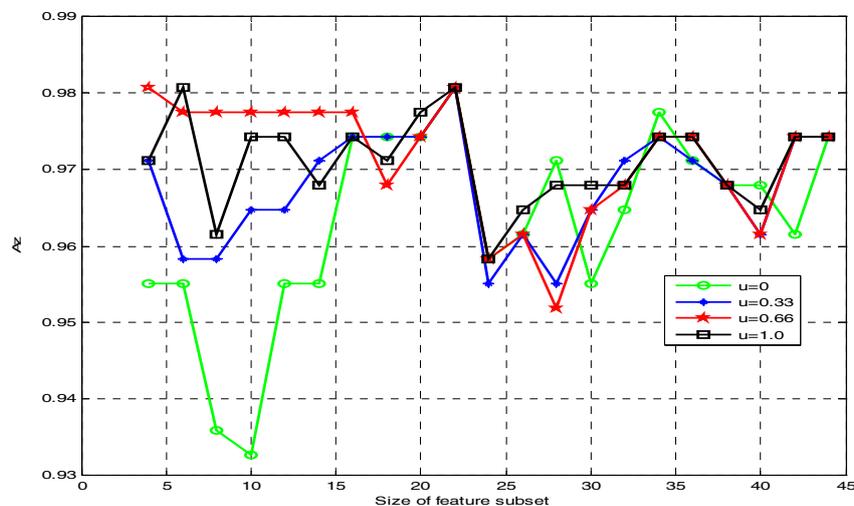

Fig. 5. Classification performance of feature subsets constructed using conventional (u=0) and outweighed ($u \geq 0$) nested subsets

When using the PSO-SVM with an outweighed nested subsets approach as a feature selection method, several learning models (feature subsets, kernel and regularization parameters) have achieved the best classification performance. An example of this is presented in Table 3, which indicates feature subsets (N=4, 10, 14, and 17) and produces a similar classification accuracy of 96% and approximately Az of 0.98.

Clearly, the binary PSO algorithm is relatively more complex than nested subset methods because the former requires N-dimensions of PSO parameter space to accomplish feature selection task; only 2 dimensions are needed for the nested subsets method. This relatively complex feature search using binary PSO provides a larger search space with a higher possibility of finding an optimal feature subset. The results presented in Table 4 demonstrate the superiority of feature selection using several learning models and binary PSO method, which achieved an optimal classification performance and 100% classification accuracy. We found that the higher the classification accuracy was the relatively higher complexity in the search method and the size of the best feature subset. The smallest size of the optimal feature subset was 9 compared to a subset of size 4 from the nested subset method.

The results presented in Table 3 and Table 4 indicated that the best classification performance was achieved using different learning models. Hence, one may ask which feature subset one should select as the final classification model. A very well acceptable answer can be formulated using Occam's razor principle that suggests the following: a simple solution is a correct one. In other words, selecting a solution or model with a lower number of features mostly leads to a classifier with better generalization abilities.

TABLE 3
RESULTS OF THE MODEL SELECTION USING THE OUTWEIGHED UNIVARIATE-BASED NESTED SUBSETS METHOD

σ	C	u	N	Sensitivity/FN	Specificity/FP	Accuracy	Az
18.3	439	0.18	17	1.0/0	0.92/1	0.96	0.98
5.5	55	0.71	14	1.0/0	0.92/1	0.96	0.98
6.9	55	0.88	10	1.0/0	0.92/1	0.96	0.98
11.3	75	0.84	4	0.91/1	1.0/0	0.96	0.98

TABLE 4
RESULTS OF THE MODEL SELECTION USING THE BINARY PSO METHOD

C	σ	N	Members of the feature subset
10.74	2.62	9	$F_{13}F_{16}F_{23}F_{26}F_{30}F_{31}F_{33}F_{35}F_{37}$
378.0	11.85	13	$F_7 F_{13}F_{18} F_{22}F_{25}F_{27}F_{30}F_{31}F_{34}F_{37}F_{38}F_{40}F_{42}$
235.0	1.28	14	$F_9F_{11}F_{13}F_{15}F_{16}F_{17}F_{18}F_{20}F_{22}F_{31}F_{32}F_{34}F_{38}F_{40}$
167.0	8.20	16	$F_5F_{10}F_{12}F_{16}F_{18}F_{23}F_{24}F_{25}F_{29}F_{30}F_{31}F_{35}F_{37}F_{40}F_{44}$
384.0	9.87	18	$F_3F_9F_{11}F_{13} F_{17}F_{18}F_{21}F_{22} F_{30}F_{31}F_{32} F_{35}F_{36}F_{37}F_{38}F_{39}F_{40}F_{44}$
102.0	6.77	21	$F_4F_7 F_8F_{12}F_{11}F_{13} F_{17}F_{19}F_{22}F_{25} F_{29}F_{30}F_{31}F_{32}F_{34}F_{37}F_{39}$
315.0	2.88	24	$F_1F_6F_8 F_9F_{10}F_{11}F_{14}F_{15} F_{17}F_{18}F_{19}F_{20}F_{23}F_{24} F_{26}F_{27}F_{30}F_{31}F_{32}$

* Indices Fi's represent shape features presented in Table 1

Considering various models obtained from outweighed univariate nested subsets method, presented in Table 3, we observe that a model with a feature subset of size 4 is a possible candidate. However, this subset produced 0 FP and 1 FN results while all other subsets produced 1 FP and 1 FN. Since FN result, missing a cancer diagnosis, has relatively a higher risk than that of FP, false breast diagnosis, a radiologist should avoid selecting subset 4. Following Occam's razor again, the candidate feature subset of size 10 is expected to be more suitable for the final model and for classifying new test patterns. This process of determining

best feature subsets was also applied to optimal feature subsets (using binary PSO). The results presented in Table 4 allowed for the selection of a feature subset of size 9 for the final classification model.

Even though a model selection using a heuristic principle might provide a general guideline, the empirical evidence is still necessary to validate any selection. Hence, in the next subsection, we examine the impact of the feature selection process on the robustness of the SVM classifier and variations of the regularization and kernel's parameters.

3) *Results of classifier's model selection*: Although it has been demonstrated that the generalization performance of SVM classifier is sensitive to the model selection process, only few studies have examined the robustness of their proposed SVM-based classification schemes to their parameter values [2] and [37]. In this study, not only did we examine the robustness of the SVM classifier to the selection of the kernel function and regularization constant but we also investigated how the feature selection process influences performance.
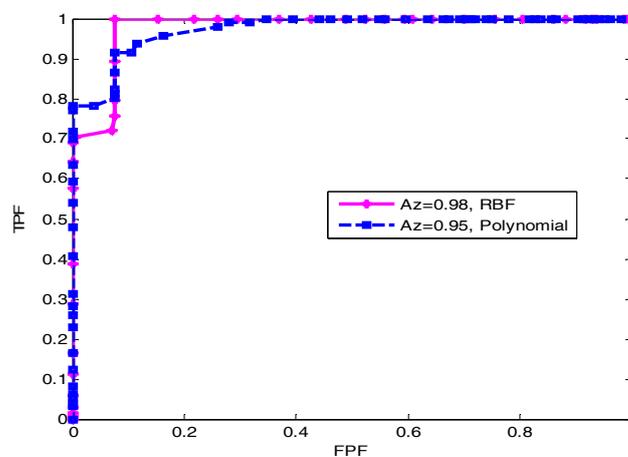

Fig. 6. Effect of the selected kernel functions on the classification performance using SVM

For instance, the RBF kernel outperformed polynomial kernel in all experiments of the classifier's hyper-parameter selection. The results obtained using univariate based feature selection method indicate a higher classification accuracy of 96 % (0 FN and 1 FP) and Az of 0.98 from RBF kernel compared to a 92% classification accuracy (1 FN and 1 FP) and Az of 0.95 using a polynomial kernel, as shown in Fig. 6. While both kernel functions produced perfect classification using binary PSO feature selection, the RBF kernel is more effective since it used a feature subset of size 9 compared to 16.

The classification performance of all learning models (Tables 3 and 4) is a function of both RBF kernel's parameter σ and regularization constant C. Investigating the effect of these parameters on the classifier's performance (i.e. generalization error) indicated the sensitivity of the classifier's performance to the values σ and *C.* We present our analysis of the models in Table 3, which is illustrated in Fig. 7, as follows:

- The generalization error of SVM descends as the C value deviates from its optimal value that is given in Table 4. As shown in Fig. 7a, learning models with feature subsets of size 4, 14, and 10 achieved best generalization error using C values of less than 60, while a learning model with a subset of size 17 required a C value of about 400.

- Using all learning models, results indicate that increasing the value of the regularization parameter C increases the generalization error. This may be due to the

fact that the SVM classifier tends to over-fit the training data when using larger values of the regularization parameter.

- Results of varying the RBF kernel parameter σ, as illustrated in Fig. 7b, indicated that the performance of the SVM classifier is more sensitive to small values of the kernel parameter σ. This can be justified by observing that a small value of σ leads to a highly nonlinear decision boundary that produces a poor generalization performance. As the σ value increases, the generalization ability of most learning models becomes more robust. This trend is mostly because large kernel's width tends to improve the linearity of the decision function and attain a better generalization performance. The feature selection process also indicated that there is a significant effect on the robustness of the SVM classifier to variations of σ and C and on the generalization performance.

- Using feature subsets of size 17, SVM shows superior robustness over a wide range of C values.

- For small values of the parameter C ($C < 40$), other feature subsets such as the subset of size 10 provide better robustness than a subset of size 17.

- Similarly, a feature subset of size 17 indicates better robustness over most of the range of σ. However, for small values of σ, small feature subsets such as a subset of size 4 provides a better performance.

- Using the same procedure, we also examined the learning models presented in Table 4. Results also pointed to the importance of selecting appropriate values of the parameters C and σ. For instance, a learning model with the feature subset of size 9 produced the best generalization error when the values of the parameters C and σ are chosen between 211 to 105, and 3 to 11, respectively. In addition, the learning model with a feature subset of size 18 outperformed all other models (excluding a model with a subset of size 9) when C values are set between 60 and 500. As for how robust is our feature selection especially in response to variations of C and σ, our results demonstrate that a feature subset of size 9 produced an average generalization error of less than 0.002, consistently provided the best robustness and outperformed all other models regardless of the value of the regularization constant C as well the parameters σ.

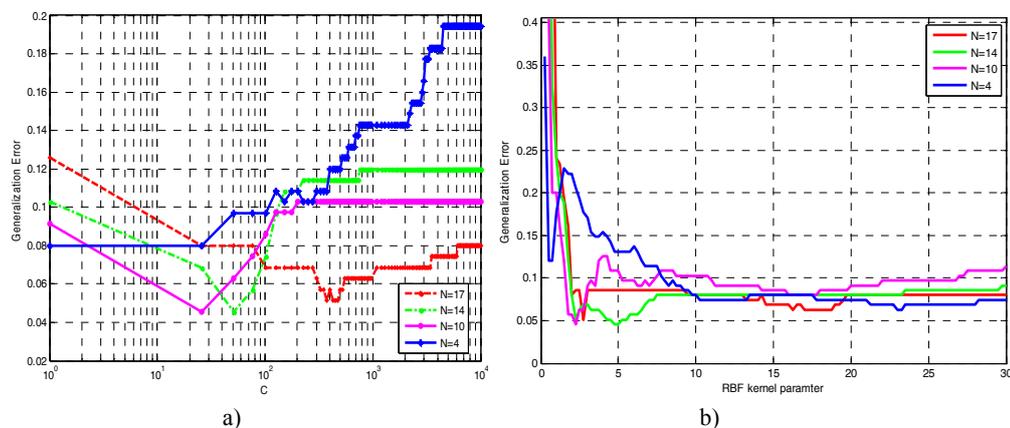a)   b)

Fig. 7. Impact of the model selection on the classifier generalization performance with feature subsets obtained using the modified nested subset method. Robustness of SVM classifier to variations of the a) regularization parameter, and b) the RBF kernel parameter σ. Generalization error in a and b was computed by averaging over several values of the parameter σ and C, respectively

D. Results on BMH Database

Using the same feature extraction and model selection methods, we have characterized the malignancy of 32 MC clusters from the BMH local database. Testing the proposed CADx and feature selection methods using binary PSO, results indicated that a classification performance of 100% accuracy can be achieved using a feature subset of 12 features only. Results also indicated that the feature selection using an outweighed nested subset scheme was less effective since it achieved a maximum classification accuracy of 88%.

E. Discriminative Power of the Shape Features

Using a single-variable evaluation of shape features based on the ROC analysis technique and an outweighed nested subsets method, results highlighted the importance of characterizing the shape and distribution of individual MCs as well as the shape of entire MC cluster in the discrimination process between malignant and benign MC clusters. For instance, shape features describing the region of MCs such as the standard deviation of the region's extent (F_{13}) and region's compactness (F_{23}) are the most discriminative features. Shape features such as the standard deviation of distances between individual MCs and their cluster's centroid (F_{27}), which model the distribution of MCs, produced high-ranking scores. Moreover, features such as the standard deviation of the normalized second order moments (F_{29}), fourth order moments (F_{31}), and normalized Fourier descriptors (F_{34}), which describe the shape boundary of MCs, strongly discriminate between malignant and benign classes. Shape features representing the entire MC cluster were also found effective. Examples of these features are the cluster area (F_{35}), cluster convex area (F_{36}), length of major axis (F_{39}) and minor axis (F_{40}), and normalized Fourier descriptors (F_{44}) of the cluster boundary. Considering the members of the best feature subsets obtained from outweighed nested subsets and binary PSO methods, one can find a good correlation between results (i.e. selected shape features) from both techniques of feature selection.

F. Comparison with other CADx

Several studies have evaluated their CADx schemes using 40 mammograms form the Nijmegen database [10], [11], [38] and [39]. Using multiwavelet features, Zadeh et al. [10] achieved the best classification performance of Az of 0.89 which produced a sensitivity of 0.85 and sensitivity of 0.9. Using the Nijmegen dataset, Verma and Zakos [38] achieved a classification accuracy of 88.9% using the neural network and fuzzy-based feature extraction model. Using texture features obtained from wavelet transforms and Haralick measures, Kramer and Aghdasi [39] produced a classification accuracy of 100%. Using 14 shape features and a neural network classifier, Kallergi et al. [4] classified a set of 100 mammograms from a local database and achieved a classification performance of Az of 0.98 that corresponds to a sensitivity of 100% and specificity of 85%. Using a neural network classifier and 10 Haralick features, Dhwan et al. [13] classified 85 mammograms from a local database with an accuracy of 74%. Also, Cordella et al. [40] presented a classification performance of Az 0.74 by using multiple expert systems to classify 40 images form DDSM dataset. Using 100 MC clusters form DDSM database, Karahaliou et al. [14] analyzed the texture of tissue surrounding MCs using different techniques and produced a classification accuracy of 89%.

The fact that there is no one common dataset used by different CADx approaches makes any direct comparison difficult. However, comparing the results of CADx scheme proposed in this paper with other CADx algorithms that used MIAS dataset indicated that our scheme achieved better classification results than others such as those in [11] and [16]. Papadopoulos et al. [11] used SVM and ANN to obtain Az of 0.81 and 0.78. In addition, Wang et al. [15] used a mixed texture, shape features and GA for SVM model selection and dimensionality reduction of the feature space and achieved Az of 0.86.

V.   CONCLUSIONS

In this paper, we characterized the malignancy of MC clusters using a morphology-based CADx scheme with a heuristic PSO-SVM embedded feature selection method. The necessary parameters for the feature selection and kernel-based SVM classification were successfully selected using a hybrid PSO algorithm. In this work, we also examined the impact of the feature extraction and SVM parameter selection on the classification performance.

Experimental results demonstrated that several design factors have a significant effect on automating the diagnosis of MCs. Thus, we must be cautious in selecting them. Factors may include the choice of the segmentation technique, the feature selection method and the level of redundancy within selected features, and the elected learning model for kernel-based SVM classifier. We, also, would like to emphasize the importance of appropriate feature selection methods and their effect on the generalization capability of the SVM classifiers, and their robustness to the variations of kernel's and regularization parameters.

We also compared the performance of embedded feature selection with the feature search using outweighed nested subsets and heuristic binary PSO algorithms. Using MC clusters from mini-MIAS and BMH datasets, the feature search using outweighed-nested subsets methods achieved classification performance of accuracy 96% and 88%, respectively. On the other hand, features selected using a heuristic binary PSO feature search method indicated the superior performance of this method and achieved a classification accuracy of 100% for both datasets. The inferior performance of feature selection using outweighed nested subsets is mostly due to the limitations of its search space. This becomes evident when weak shape patterns of MC clusters are present. The results of this study are promising and demonstrate the potential of the proposed CADx to improve the PPV for malignancy analysis of MC clusters. Our future work will include larger datasets of mammograms from the local hospital and public data sources such as DDSM dataset, after we revise the algorithm to address the limitations in our current pilot implementation.

REFERENCES

[1] I. Zyout, J. Czajkowska, M. Crzegorzek, "Multi-scale textural feature extraction and particle swarm optimization based model selection for false positive reduction in mammography," *Computerized Medical Imaging and Graphics*, vol. 64, no. 2, pp. 95-107, 2015.

[2] L. Wei, Y. Yang, R. M. Nishikawa and Y. Jiang, "A study on several machine-learning methods for classification of malignant and benign clustered microcalcifications," *IEEE Transactions on Medical Imaging*, vol. 24, no. 3, pp.1278-1285, 2005.

[3] M. Elter and A. Horsch, "CADx of mammographic mass and clustered microcalcifications: a review," *Medical Physics*, vol. 36, no. 6, pp. 2052-2068, 2009.

[4] M. Kallergi, "Computer-aided diagnosis of mammographic microcalcification clusters," *Medical Physics*, vol. 31, no. 2, pp. 314-326, 2004.

[5] C. J. D'Orsi and D.B Kopans, "Mammographic feature analysis," *Seminars in Roentgenology*, vol. 28, no. 3, pp. 204-230, 1993.

[6] L. Shen, R. M. Ranayyan and J. E. L. Desautels, "Application of shape analysis to mammographic calcifications," *IEEE Transactions Medical Imaging*, vol. 13, no. 2, pp. 263-274, 1994.

[7] Y. Jiang, R. M. Nishikawa, D. E. Wolverton, C. E. Metz, M. L. Giger, R. A. Schmidt, C. J. Vyborny and K. Doi, "Malignant and benign clustered microcalcifications: automated feature analysis and classification," *Radiology*, vol. 198, no. 3, pp. 671-678, 1996.

[8] H. P. Chan, B. Sahiner, K. L. Lam, N. Petrick, M.A. Helvie, M. M. Goodsitt and D. D. Adler. "Computerized analysis of mammographic microcalcifications in morphological and texture feature spaces," *Medical Physics*, vol. 25, no. 10, pp. 2007-2019, 1998.

[9] H. S. Zadeh, S. P. Nezhad and F. R. Rad, "Shape based and texture-based feature extraction for classification of microcalcifications in mammograms," *Proceedings of the SPIE Medical Imaging Conference*, 2001.

[10] S. H. Zadeh, F. R. Rad and P. S. Nezhad, "Comparison of multiwavelet, wavelet, Haralick, and shape features for microcalcification classification in mammograms," *Pattern Recognition*, vol. 37, no. 10, pp.1973-1986, 2004.

[11] A. Papadopoulos, D. I. Fotiadis and A. Likas, "Characterization of clustered microcalcifications in digitized mammograms using neural networks and support vector machines," *Artificial Intelligence in Medicine*, vol. 34, no. 2, pp.141-150, 2005.

[12] R. R. Hern´andez-Cisneros and H. Terashima-Mar´ın, "Evolutionary neural networks applied to the classification of microcalcification clusters in digital mammograms," *Proceedings of IEEE Congress on Evolutionary Computation*, pp. 2459-2466, 2006.

[13] A. P. Dhawan, Y. Chitre, C. Bonasso and K. Wheeler, "Radial-basis-function- based classification of mammographic microcalcifications using texture features," *Proceedings of Annual International Conference, IEEE Engineering in Medicine and Biology Society*, pp. 535-536, 1995.

[14] A. Karahaliou, I. Boniatis, P. Sakellaropoulos, S. Skiadopoulos, G. Panayiotakis and L. Costaridou, "Can texture of tissue surrounding microcalcifications in mammoramphy be used for breast cancer diagnosis?," *Nuclear Instruments and Methods in Physics Research*, vol. 580, no. 2, pp. 1071-1074, 2007.

[15] C. Wang, W. Jiang, and X. Dong, "Characterization of clustered microcalcifications in mammogram based on support vector machine with Genetic Algorithms," *IEEE International Conference Biophotonics, Nanophotonics and Metamaterials*, pp.114-117, 2006.

[16] T. Yang, S. Guo and X. Wu, "Approach based on immune algorithm and SVM for detection and classification of microcalcifications," *Proceedings of International Conference on Bioinformatics and Biomedical Engineering,* pp. 588-591, 2007.

[17] I. Zyout and I. Abdel-Qader, "An improvement of texture-based classification of microcalcification clusters in mammography using PSO-SVM approach," *Proceedings of IEEE International Conference on Communications, Computers and Applications*, pp. 7-11, 2012.

[18] I. Zyout, "Classification of clustered microcalcifications in mammograms using particle swarm optimization and least-squares support vector machine," *International Journal of Computer Applications*, vol. 59, no.17, pp. 23-28, 2012.

[19] J. Kennedy and R. Eberhart, "Particle swarm optimization," *Proceedings of IEEE International Conference on Neural Networks*, vol. 4, pp. 1942-1948, 1995.

[20] I. Guyon, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol.3, pp. 1157-1182, 2003.

[21] I. Guyon, S. Gunn, M. Nikravesh and L. Zadeh, *Feature extraction-foundations and applications*, Springer, 2006.

[22] X. C. Guo, J. H. Yang, G. C. Wu, C. Y. Wang and Y. C. Liang, "A novel LS-SVMs hyper-parameter selection based on particle swarm optimization," *Neurocomputing*, vol. 71, no. 16-18, pp. 3211-3215, 2008.

[23] C. J. Tu, L. Y. Chuang, J. Y. Chang and C.-H. Yang, "Feature selection using PSO-SVM," *IAENG International Journal of Computer Science*, vol. 33, no. 1, 2007.

[24] H. J. Escalante, M. Montes and L. E. Sucar, "Particle swarm model selection," *Journal of Machine Learning Research*, vol. 10, pp. 405-440, 2009.

[25] C. L. Huang and J. F. Dun, "A distributed PSO–SVM hybrid system with feature selection and parameter optimization," *Applied Soft Computing*, vol. 8, no. 4, pp. 1381-1391, 2008.

[26] J. Kennedy and R. Eberhart, "A discrete binary version of the particle swarm algorithm," *Proceedings of International Conference on Systems, Man, and Cybernetics*, pp. 4104-4109, 1997.

[27] W. Siedlecki and J. Sklansky, "A note on genetic algorithm for large scale feature selection," *Pattern Recognition Letter*, vol. 10, no. 5, pp. 335-347, 1989.

[28] I. Zyout, *Toward Automated Detection and Diagnosis of Mammographic Microcalcifications*, Ph.D. Dissertation, Department of Electric and Computer Engineering, Western Michigan University, 2010.

[29] B. E. Boser, I. Guyon and V. Vapnik, "A training algorithm for optimal margin classifiers," *Proceedings of Annual Workshop on Computational Learning Theory*, pp.144-152, 1992.

[30] C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp.121-167, 1998.

[31] V. Vapnik, *The Nature of Statistical Learning Theory*. Springer-Verlag, 1995.

[32] J. Fu, S. Lee, S. Wong, N. Yeh, A. Wang and H.Wu, "Image segmentation feature selection and pattern classification for mammographic microcalcifications," *Computerized Medical Imaging and Graphics*, vol. 29, no. 6, pp.419-429, 2005

[33] M. Sonka, V. Hlavac and R. Boyle, *Image processing, Analysis, and Machine Vision*. Thomson Learning, 2008.

[34] J. Dengler, S. Behrens and T. F. Desaga, "Segmentation of microcalcifications in mammograms," *IEEE Transactions on Medical Imaging*, vol. 12, no. 4, pp.634-642, 1993.

[35] D. Betal, N. Robert and G. H. Whitehouse, "Segmentation of microcalcifications and numerical analysis on mammograms using mathmetical morphology," *British Journal of Radiology*, vol. 70, no. 837, pp. 903-917, 1997.

[36] J. Suckling, J. Parker, D. Dance, S. Astley, I. Hutt, C. Boggis, I. Ricketts, E. Stamatakis, N. Cerneaz, S. Kok, P. Taylor, D. Betal and J. Savage, "The mammographic image analysis society digital mammogram database," *Exerpta Medica*, vol. 11069, pp. 375-378, 1994.

[37] M. Roffilli, "Advanced machine learning techniques of digital mammography," *Technical Report* UBLSC-2006-12, 2006. Available: www.cs.unibo.it/pub/TR/UBLCS/2006/2006-12.pdf.

[38] B. Verma and J. Zakos, "A computer-aided diagnosis system for digital mammograms based on fuzzy-neural and feature extraction techniques," *IEEE Transactions on Information Technology in Biomedicine*, vol. 5, no. 1, pp. 46-54, 2001.

[39] D. Kramer and F. Aghdasi, "Classification of microcalcifications in digitized mammograms using multiscale statistical texture analysis," *Proceedings of the South African Symposium on Communications and Signal Processing*, pp. 121-126, 1998.

[40] L. P. Cordella, F. Tortorella and M. Vento, "Combing experts with different features for classifying clustered microcalcifications in mammograms," *Proceedings of International Conference on Patten Recognition*, pp. 324-327, 2000.