Hindawi Publishing Corporation
International Journal of Biomedical Imaging
Volume 2009, Article ID 680508, 12 pages
doi:10.1155/2009/680508

Research Article

Hybrid Mammogram Classification Using Rough Set and Fuzzy Classifier

Fadi Abu-Amara and Ikhlas Abdel-Qader

Department of Electrical and Computer Engineering, Western Michigan University, MI 49008, USA

Correspondence should be addressed to Ikhlas Abdel-Qader, ikhlas.abdelqader@wmich.edu

Received 29 December 2008; Revised 18 June 2009; Accepted 25 July 2009

Recommended by Jayanta Mukherjee

We propose a computer aided detection (CAD) system for the detection and classification of suspicious regions in mammographic images. This system combines a dimensionality reduction module (using principal component analysis), a feature extraction module (using independent component analysis), and a feature subset selection module (using rough set model). Rough set model is used to reduce the effect of data inconsistency while a fuzzy classifier is integrated into the system to label subimages into normal or abnormal regions. The experimental results show that this system has an accuracy of 84.03% and a recall percentage of 87.28%.

1. Introduction

Breast cancer is the most common cancer among women worldwide. National cancer institute [1] estimates that 192 370-female and 1910-male new cases of breast cancer will appear in the United States in 2009. Also, it is estimated that 40 170 females and 440 males will die of this cancer. Early detection of this disease remains the best known method for reducing its mortality. Also, mammography remains one of the best modalities used by radiologists for early detection of cancerous tumors before clinical symptoms appear. Unfortunately, the growing demand for mammograms is limited by insufficient number of radiologists [2]. A CAD system can be used to assist radiologists in differentiating between normal and suspicious regions, and thus reducing number of unnecessary biopsies and false-positive rates (FP) by the radiologist, FP is an erroneous positive diagnosis when the breast is normal.

Several rough set-based and fuzzy-based methods have been proposed in literature for breast cancer detection. Hassanien and Ali [3] proposed a rough set technique for feature reduction and classification-rule generation from mammographic images. Hu et al. [4] proposed a rough set model (RSM) based on relational algebra that replaces the traditional rough set models. Their proposed algorithm is very efficient in large data sets and may be adaptable for real-time applications. Şahan et al. [5] proposed a hybrid machine learning algorithm by hybridizing k-nearest neighbor algorithm with a fuzzy-artificial immune method where a 10-fold cross validation criterion was used to compute algorithm's accuracy. Hassanien [6] proposed a hybrid method that first uses fuzzy logic to enhance image contrast, extracts region of interest, and enhances its edges. Then, the gray-level cooccurrence matrix is used as a feature extraction method. RSM is used for further subset selection and rule generation and classification. RSM can also be used as a feature selection algorithm [7–10] while fuzzy logic as a classifier [11–13].

In [14], an algorithm was proposed that combined PCA, ICA, and fuzzy classifier for breast cancer detection. Membership functions of fuzzy sets were generated from the product space of the selected features. Also, the selected features from PCA-ICA phase suffered from data inconsistency which degraded the fuzzy classifier performance. In this work, an integration of PCA, ICA, Rough Set, and fuzzy classifier to identify and label suspicious regions from digitized mammograms is developed. Results of this system showed a higher efficiency in detecting suspicious regions and reducing false-negative (FN) rates in comparison with the results of [14] where FN is an erroneous negative diagnosis but the breast tissue has cancer. This work presents a new approach since the mapping range is integrated into the rough set model as opposed to being part of a fuzzy classifier as was the case with [14]. The RSM is integrated into the proposed system as a feature subset selection method in order to reduce the impact of data inconsistency. Finally, the membership functions of the fuzzy sets are based on the mean and standard deviation of the testing data.

In [15] an algorithm was proposed that combined ICA with RSM for breast cancer detection where ICA was used for feature extraction and reduction while in this work PCA is used for feature reduction since PCA is superior to ICA in dimensionality reduction which will enhance the ICA performance, and since it is recommended to preprocess the data through whitening prior to ICA as a tool to reduce the complexity of the problem [16], PCA was a natural choice since whitening is an intrinsic step in PCA.

The novelty of this work is the integration of RSM for feature selection with a fuzzy classifier as well as generating the framework for the integration of the PCA, ICA, RSM, and fuzzy classifier for breast cancer detection. The rest of this paper is organized as follows. Section 2 presents a brief introduction to PCA, ICA, and RSM. Section 3 presents fuzzy logic adaptation while the proposed approach is presented in Section 4. Experimental results are presented in Section 5 followed by Conclusions in Section 6.

2. Background

2.1. PCA. PCA is an orthogonal transform and a decorrelation technique that captures maximum variance. The correlation between components of a vector is used to measure data redundancy. This means that most of the information contained in the original vector can be represented by a much smaller vector after the PCA stage. In this paper, PCA is used as a dimensionality and noise reduction module. This step ensures that the source components of a vector are uncorrelated.

2.2. ICA. ICA is a statistical technique that can be used to extract hidden features within a set of data.

A mammographic image X can be expressed as a linear mixture of a set of features or basis functions a_i as shown in (1):

$$x = \sum_i a_i s_i \tag{1}$$

where s_i are stochastic coefficients that are data dependent. Other transforms such as Wavelets and Gabor assume basis vectors that are independent from data while ICA assumes basis vectors that costumed to the data under consideration. Using matrix notations, (1) can be expressed as shown in (2):

$$\mathbf{X} = \mathbf{AS} \tag{2}$$

where S is a matrix contains the source components and A is the mixing matrix. This means that a mammographic image consists of a mixture of source components S. Their combination can be described using the coefficients of the mixing matrix A which can be used as extracted features that describe efficiently any normal and suspicious region.

The ICA algorithm estimates the separating matrix W (inverse of A) that makes the source components S as statistically independent as possible with non-Gaussian (super- or sub-Gaussian) distribution which results in obtaining independent components as shown in (3). This means that A should be a square matrix which can be achieved by preprocessing of PCA:

$$\mathbf{S} = \mathbf{WX} \tag{3}$$

The ICA algorithm can be presented as an optimization process of which an objective function is modeled to minimize statistical dependency between the source components. The statistical estimation of the **W** and **S** matrices is a result of this optimization process. The dependency between the source components can be minimized using several suggested methods such as minimizing the mutual information of the components representation [17], maximizing their likelihood [18], or maximizing their non-gaussianity [19, 20].

2.3. RSM. Rough set theory can be used as a feature subset selection algorithm. RSM determines and removes the dispensable attributes representing the redundant information within the data while it aims to keep the core attributes representing the minimum essential information.

By relaxing the core algorithm, more attributes can be selected which are called *Reduct*. In this paper, *Reduct* attributes are considered as the minimum selected features. The selected *Reduct* should have the same discernibility and representation power as the original data.

Cardinality is used to replace traditional rough set theory operations. Therefore, algorithm efficiency will be improved with reduced complexity. The cardinality of a set is defined as the number of elements in the set. For example, Table 1 shows three selected features for 8 images (symbols are used instead of pixel values for simplicity). The decision is either normal (N_m) or suspicious (S) image. The cardinality of Table 1 is

$$|\mathrm{I}| = 6 \text{ objects}, \tag{4}$$

where I = {Feature 1, Feature 2, Feature 3, Decision}. Core attributes should be in every Reduct to ensure correct classification. Therefore, removing any core attribute affects the classifier accuracy. Hu et al. [4] defined the core attributes by (5) as

$$\frac{\left|I - C_j\right|}{\left|I - C_j - D\right|} \neq 1, \tag{5}$$

where I is the decision matrix $I = [C\vdots D]$, C is the condition attributes (selected features), D is the decision attribute (normal or suspicious image), and C_j is the current attribute to be classified as a core or not. The merit value of an attribute or the significance of the attribute is calculated

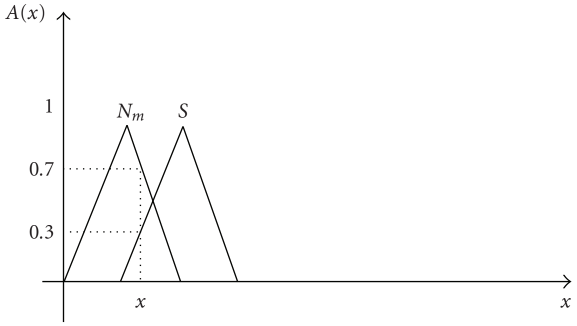


Figure 1: Fuzzy space for an object x consisting of two fuzzy sets: "Normal" and "Suspicious".

Table 1: Selected features for eight images.

Image	Feature 1	Feature 2	Feature 3	Decision
A_1	A	F	C	N_m
A_2	A	F	D	S
A_3	E	E	C	N_m
A_4	B	E	C	S
A_5	B	E	C	S
A_6	A	E	D	N_m
A_7	A	E	D	S
A_8	A	F	D	S

using (6) which is a measure of the degree of dependency for an attribute on the condition and decision attributes:

$$S\left(C_j\right) = \frac{|I| - \left|I - C_j\right|}{|I|} = 1 - \frac{\left|I - C_j\right|}{|I|}. \tag{6}$$

Two objects are considered consistent if they have the same condition and decision values. For example, in Table 1, the 2nd and the 8th objects are said to be consistent. On the other hand, the 6th and the 7th objects are inconsistent. Inconsistent objects are conflicting objects since they have same selected features but belong to different classes. Rough set model is used in this work to reduce number of inconsistent objects.

3. Fuzzy Logic

Human reasoning can be emulated using fuzzy logic. Fuzzy logic is proved to be a powerful tool to handle and process noisy and vague data. Fuzzy rules are more flexible than crisp rules for many reasons. They allow partial set membership and overlapping between fuzzy set definitions which should simplify the classification phase as opposed to crisp rules that are restricted to either a membership or nonmembership to the set. Also, they can be expressed in terms of linguistic statements based on expert knowledge. Finally, the interpretability of the results can be improved by fitting fuzzy rules to the labeled observed data.

Fuzzy membership functions are easy to implement and they improve speed of inference engines. The difference between normal and suspicious mammographic images may not be well defined. Figure 1 shows, for example, that the object x has a membership degree of 0.7 to the fuzzy set "normal" and 0.3 to the fuzzy set "suspicious".

Several approaches have been developed for automatic derivation of fuzzy rules from the labeled observed data such as genetic algorithm [21], Neuro-fuzzy [22], and fuzzy clustering [23]. In all, the derived fuzzy rules should be accurate, compact, and linguistically interpretable.

Fuzzy if-then rules are used to implement membership function of fuzzy sets as shown in (7):

$$\text{IF antecedent THEN consequent } [\text{weight}]. \tag{7}$$

The weight is a number in the interval [0.0, 1.0] that can be evaluated based on the antecedent numbers. For example, a tested subimage has a membership degree of 0.7 to the fuzzy set "normal" and 0.3 to the fuzzy set "suspicious". In this case, a single fuzzy if-then rule can be used which produces a classifier output of normal for the tested subimage as shown by the following.

$$\begin{gathered}\text{If x is normal } [0.7]\text{and x is suspicious } [0.3], \text{then} \\ \text{Y} = \text{normal}.\end{gathered} \tag{8}$$

Equation (8) is evaluated in two steps. First, a fuzzy operator is applied in order to fuzzify the antecedent numbers. For example, the union fuzzy operator can be applied using (9):

$$u(a, b) = \max(a, b), \tag{9}$$

where a and b are the membership degrees for the membership functions. Applying (9) to the antecedent of (8) will result in selecting the normal fuzzy set from the antecedent with membership degree of 0.7 as follows:

$$u(a, b) = \max(0.7, 0.3) = 0.7. \tag{10}$$

The antecedent results are applied then to the consequent, which is known as the inference step. In this case, the classifier will label the tested subimage as normal.

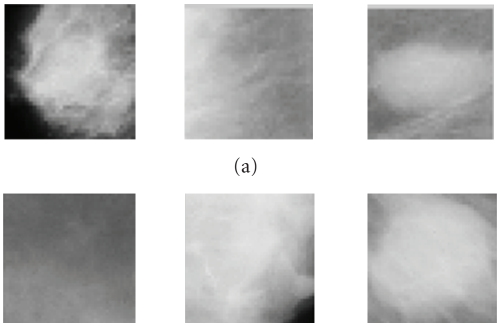
4. Proposed CAD Algorithm

This paper integrates four techniques, namely, PCA, ICA, Rough Set, and Fuzzy classifier to build a CAD system. PCA algorithm is used as a dimensionality and noise reduction tool (prewhitening), and ICA algorithm is used as a feature extraction module while RSM is used as a feature subset selection module followed by a fuzzy classifier.

4.1. Data Preprocessing. 119 regions of suspicion (ROS) are manually extracted from MIAS database [24] based on center of each abnormality of which 51 are malignant and 68 are benign regions. Two sets are formed where the first set is with subimages of size 45 × 45 while the second set of size 35 × 35 pixels.

Table 2: Data sets that used in the evaluation of the proposed algorithm performance.

Set no.	Training set			Testing set			
	ROS	Normal	Total	ROS	Normal	Total	Size-pixels
1	60	59	119	59	60	119	35 × 35
2	60	59	119	59	60	119	45 × 45
3	60	59	119	59	60	119	45 × 45
4	60	59	119	59	60	119	45 × 45


(a)

(b)

Figure 2: (a) Benign, normal, and malignant subimages of size 35 × 35 pixels and (b) benign, normal, and malignant subimages of size 45 × 45 pixels.

Four other sets of normal subimages are randomly and automatically extracted such that the first set is of size 35 × 35 and the other sets are of size 45 × 45 pixels from the normal MIAS mammograms. Each set has 119 subimages. Each set of ROS is mixed with one set of normal subimages and then divided into two groups: one for the training phase and the other is for the testing phase as shown in Table 2. Figure 2 shows a sample of the extracted subimages.

4.2. Training Phase Using PCA-ICA. A training matrix $R_{\text{train}_{N\times M}}$ is constructed by placing training subimages as rows in the matrix where N represents number of training subimages (119) and M represents size of each square subimages. PCA algorithm is used to reduce its dimensionality according to the following equation where v represents number of selected principal components and $R_{M\times v}$ represents a matrix with the principal components in its columns sorted by descending order according to their variances

$$R^R_{N\times v} = R_{\text{train}_{N\times M}} R_{M\times v}. \tag{11}$$

In this paper, ICA scheme is based on minimizing the mutual information of the source components which can be achieved using cumulants. This is proposed (a modified version of [17]) in order to estimate the separating matrix $W$and the independent source region matrix S in an unsupervised mode as follows.

(i) W is initialized to the identity matrix. Then, S is calculated using the following equation. This means that ICA is performed on a set of v linear combinations of the original subimages instead of performing it on all N subimages. This should reduce its computational complexity and hence increase its speed:

$$S_{v\times M} = W_{v\times v} \left(R_{M\times v}\right)^T. \tag{12}$$

(ii) The change in W is calculated using the natural gradient [25], that is,

$$\Delta W = \eta\left[I - G(S)S^T\right] W, \tag{13}$$

where η is the learning rate (step size), I is the identity matrix, and $G(s)$ must be a nonlinear and nonfast growing function. This function is used to measure the statistical dependence between the source components. In this paper, $G(s)$ [26] is used as follows:

$$G(S) = f_1(k_3, k_4) \circ S^2 + f_2(k_3, k_4) \circ S^3, \tag{14}$$

where k_3 and k_4 are the 3rd and 4th cumulants and $(\circ)$ indicates Hadamard product of two matrices and

$$\begin{aligned} f_1(k_3, k_4) &= 0.5\ k_3(4.5k_4 - 1), \\ f_2(k_3, k_4) &= 1.5(k_3)^2 + \frac{1}{6}k_4(4.5k_4 - 1), \end{aligned} \tag{15}$$

as were defined in [26].

(iii) The momentum method is used to boost the convergence speed of (13) using

$$\Delta W(t+1) = \Delta W(t) + \alpha\Delta W(t-1), \tag{16}$$

where α is in the range [0, 1]. In this paper, alpha is chosen to be 0.5.

(iv) The separating matrix is updated and then normalized:

$$\begin{aligned} W(t+1) &= W(t) + \Delta W(t), \\ W(t+1) &= \frac{W(t)}{\| W(t)\|}. \end{aligned} \tag{17}$$

(v) Stop the algorithm when W converges.

Finally, the reduced dimensionality selected features can be estimated as follow.

A minimum square error approximation of the training matrix $R_{\text{train}_{N\times M}}$ can be found using the following equation [27] based on (11):

$$\begin{aligned} X_{\text{rec}} &= R^R_{N\times v}\ R^T_{M\times v} \\ &= R_{\text{train}_{N\times M}} R_{M\times v}\ R^T_{M\times v} \approx R_{\text{train}_{N\times M}}. \end{aligned} \tag{18}$$

From (10),

$$(R_{M\times v})^T = W_{v\times v}^{-1} S_{v\times M}. \tag{19}$$

And substitution of (19) into (18) yields

$$X_{\text{rec}} = R_{N\times v}^R R_{M\times v}^T = R_{N\times v}^R W_{v\times v}^{-1} S_{v\times M}. \tag{20}$$

Since X_{rec} is an approximation of R_{train} and by comparing (20) with (2), the extracted features from the corresponding training set are estimated using (21):

$$Q_{\text{train}_{N\times v}} = R_{N\times v}^R W_{v\times v}^{-1}. \tag{21}$$

4.3. Testing Phase Using PCA-ICA. First, a testing matrix $R_{\text{test}_{N\times M}}$ is constructed, where each testing subimage forms a row in the matrix. Second, its rows are normalized by their mean. Third, The regions in $R_{\text{test}_{N\times M}}$ are projected on the reduced data from the training procedure using (22):

$$Q_{t_{N\times v}} = R_{\text{test}_{N\times M}}\ R_{M\times v}. \tag{22}$$

The reduced dimensionality extracted features from the corresponding testing set are estimated using (23) which is the same principal as (21):

$$Q_{\text{test}_{N\times v}} = Q_{t_{N\times v}} W_{v\times v}^{-1}. \tag{23}$$

4.4. Mapping into a Limited Range. The estimated matrices $Q_{\text{train}_{N\times v}}$ and $Q_{\text{test}_{N\times v}}$ contain N rows where each row contains v selected features from the corresponding subimage. A linear stretching method is used to map them into a limited range of [0, r] using (24):

$$q(x,y) = \frac{(q(x,y) - \min(q))(r)}{\max(q) - \min(q)}. \tag{24}$$

4.5. Rough Set Model. There are some inconsistent elements (subimages) in the estimated matrices $Q_{\text{train}_{N\times v}}$ and $Q_{\text{test}_{N\times v}}$. These elements have same selected features but belong to different classes. Rough Set Reduction is used as a subset selection in order to remove features that cause inconsistency and thus improve classification results.

4.5.1. Training Phase. The proposed training framework can be summarized as follows.

(1) The consistent elements from the training matrix are removed. The resulting matrix is $Q_{\text{train}_{NN\times v}}$, where NN < N.

(2) Construct the decision matrix, $I_{NN\times(v+1)} = [Q_{\text{train}_{NN\times v}} \vdots D_{NN\times 1}]$, where Q contains the condition attributes (selected features from PCA-ICA phase) and D is the decision attribute (1: abnormal, 0: normal).

(3) Find the Core attributes using the following procedure.

(i) Initialize Core vector into $\varnothing$.

(ii) Check the cardinality for each attribute $C_j \in C$; if it satisfies $|I - C_j|/|I - C_j - D| \neq 1$, then update core vector as Core = $[\text{Core} \vdots C_j]$.

(4) Find *Reduct* attributes using the following procedure which is a modified version of [4].

(i) Initialize *Reduct* vector: Reduct = Core.

(ii) Set *Rest* = *I* − *Reduct* and compute the significance of its attributes using:

$$S(C_j) = \frac{|I| - \left|I - C_j\right|}{|I|} = 1 - \frac{\left|I - C_j\right|}{|I|}. \tag{25}$$

(iii) Let $C_{\max}$ be the attribute with the largest significance value, update *Reduct* as: *Reduct* = $[Reduct \vdots C_{\max}]$

(iv) Update Rest = I − Reduct.

(v) If K ≥ T or the significance values of the remaining attributes are zeros, stop the procedure. Equation (26) means that *Reduct* has inconsistent elements (with ratio of T) greater than or equal to that of the decision matrix:

$$K = \frac{\text{number of inconsistent rows for [Reduct} \vdots \text{D]}}{\text{number of inconsistent rows for I}}. \tag{26}$$

(vi) Else, go to step (II).

4.5.2. Testing Phase. In this step, features are selected from the matrix $Q_{\text{test}_{N\times v}}$ in the same order they were selected from $Q_{\text{train}_{N\times v}}$ during the training phase.

Finally, $Q_{\text{train}_{N\times vv}}$ and $Q_{\text{test}_{N\times vv}}$ are reconstructed with selected Reduct features while dispensable features are thrown away.

4.6. Fuzzy Classifier. Two single fuzzy if-then rules are used to represent the normal and abnormal fuzzy sets. The membership functions of each antecedent fuzzy set are aggregated using the information about the selected feature values of the training subimages.

The proposed fuzzy-based classification algorithm can be summarized as follows:

(1) Two activation functions $\mu_{\text{as}_{N\times 1}}$ and $\mu_{\text{ns}_{N\times 1}}$ are initialized to 0 where each element of them represents the aggregated membership functions of the selected feature values for the corresponding testing subimage. These parameters are defined as.

(i) $\mu_{\text{as}_{k\times 1}}$ represents the membership degree of the kth testing subimage to the fuzzy set abnormal.

(ii) $\mu_{\text{ns}_{k\times 1}}$ represents the membership degree of the kth testing subimage to the fuzzy set normal where $1 \leq k \leq N$.

(2) Using (27), membership functions of fuzzy sets of the testing subimages are obtained from the mean and standard deviation of their selected features based on the information from the selected feature values of the training subimages:

$$A_{ij}\left(x_j\right) = \exp\left(-\frac{\left(x_j - \mu_j\right)^2}{2\left(\sigma_j\right)^2}\right), \tag{27}$$

where μ_j represents mean of all samples of the current selected feature x_j, σ_j represents their standard deviation, and i is an index for the selected features from the training phase.

(3) The membership functions are normalized using

$$A_{ij}\left(x_j\right) = \frac{A_{ij}\left(x_j\right)}{\max_{ij}\left(A_j\left(x_j\right)\right)}. \tag{28}$$

(4) The membership functions are aggregated using (29) in order to find the degree of activation of each fuzzy set where i is an index for the selected features from the testing phase:

$$\mu_i(x) = \sum_{j=1}^{N} A_{ij}\left(x_j\right). \tag{29}$$

(5) By assigning the corresponding testing subimage into the fuzzy set with the maximum degree of activation, a crisp decision is made, that is, normal or abnormal. Equation (30) is used for this purpose where C is used as an index of a testing subimage being identified as normal or abnormal:

$$C = \max\left(\mu_{\text{as}}(x), \mu_{\text{ns}}(x)\right). \tag{30}$$

5. Experimnetal Results

Table 3 presents results of using PCA-ICA-Rough-Fuzzy (PIRF), PCA-ICA-Fuzzy (PIF), PCA-Fuzzy (PF), PCA-Rough-Fuzzy (PRF), ICA-Fuzzy (IF), and ICA-Rough-Fuzzy (IRF) in terms of accuracy, recall, precision, FN rates, and FP rates as computer-aided detection systems. Algorithm accuracy is defined as the ratio between the total number of correctly classified subimages to the total number of testing subimages.

Table 4 compares the performance of these CAD systems. Our proposed PIRF CAD system shows a robust performance in comparison with the other algorithms. For example, PIRF achieved an average accuracy of 77.73%, PIF of 75.21%, IRF of 74.16%, PRF of 71.85%, PF of 71.64%, and IF of 49.58%. As Table 3 shows, PIRF has the highest recall percentage among all the other algorithms while it has an average precision of 73.33%. PIF and IRF have average precision of 75.83% each.

As the results show, fuzzy classifier cannot be implemented with ICA model alone without a dimensionality reduction since, without it, a large number of membership functions will be generated. Also, without a feature subset selection module, the classifier task complexity is increased and performance is degraded. Furthermore, results indicate that integrating ICA model with PF generated better results than integrating RSM with PF. The average accuracy was improved by 4.68% and false negative rates were improved by 4.76% if a PCA model was used with the ICA model while following it with RSM improved its average accuracy by 0.29% and its FN rates by 6.33%. Integrating RSM improved total PF algorithm performance by 0.29% but degraded its FN rates by 6.34%. Results also indicate that RSM and PIF integration improves accuracy with an average of 3.35%.

Comparing the results using FN rates, we find that PIRF has an FN of 8.82%, PIF of 12.61%, IRF of 13.66%, PF of 13.24%, PRF of 14.08%, and IF of 40.34%. Results indicate that using PCA as a dimensionality reduction module reduces FN rates in PIRF and PF at the expense of a little increase in the FP rates. Also, average FN rates are very close to average FP rates in PIF and PRF algorithms. On the other hand, average FN rates are increased in IRF and IF algorithms when no dimensionality reduction was integrated. Finally, integrating RSM into PIF and PF algorithms reduces the number of principal components required to obtain Reduct. The previous discussion shows that each one of the integrated techniques (PCA, ICA, RSM, and Fuzzy Classifier) is necessary and should be implemented in the proposed sequence in order to achieve the highest accuracy rates.

An implementation of the PIF proposed in [14] reports, Table 3, a lower accuracy than our proposed PIRF system in two testing sets while they had same accuracy in the other two testing sets. The average accuracy of the PIF in all test sets is 75.21% while 77.73% for PIRF. FN rates improved in three testing sets for the PIRF in comparison with the PIF. The average FP and FN rates of the PIF are 12.19% and 12.61%, respectively, while 13.45% and 8.82% for PIRF. These observations are summarized in Table 5.

The average accuracy for PIF improved by 3.35% with PIRF system and its average FN rate improved 30.01%. Also, the average selected number of principal components in PIRF algorithm which is 7.75 is less than that of PIF algorithm which is 9.75. In other classification methods such as in [15], three sets of sizes 20 × 20, 40 × 40, and 60 × 60 pixels were extracted from MIAS mammographic images where each set consists of 330 subimages. Their results were 65.71%, 59.36%, and 82.22% for the three sets using ICA-Rough algorithm and 81.9%, 88.57%, and 69.27% using PCA-Rough algorithm.

The proposed CAD system uses several parameters that impact performance accuracy such as number of the principal components in the PCA algorithm, learning rate and alpha in the ICA algorithm, threshold in the Reduct process, and mapping range.

Number of PCs Selected. Reducing data dimensionality using PCA module affects PIRF algorithm accuracy. When large number of principal components is selected, extracted

Table 3: Results of PCA-ICA-Fuzzy, PCA-ICA-Rough-Fuzzy, PCA-Fuzzy, PCA-Rough-Fuzzy, ICA-Fuzzy, and ICA-Rough-Fuzzy algorithms. NA: not applicable.

Algorithm	Set no.	PC	FP	FN	Accuracy	Precision	Recall
PCA-ICA-Rough-Fuzzy	1	8	21.85%	9.24%	68.91%	56.66%	75.56%
	2	7	9.24%	7.56%	83.19%	**81.67**%	84.49%
	3	8	12.61%	12.61%	74.79%	74.99%	74.99%
	4	8	10.08%	5.88%	**84.03**%	80.01%	**87.28**%
PCA-ICA-Fuzzy	1	8	16.81%	14.28%	68.91%	66.66%	70.18%
	2	20	10.92%	5.89%	**83.19**%	78.34%	**87.02**%
	3	6	12.61%	21%	66.39%	74.99%	64.29%
	4	5	8.4%	9.25%	82.35%	**83.34**%	81.96%
PCA-Fuzzy	1	20	26.05%	10.08%	63.87%	48.33%	70.74%
	2	5	10.08%	14.29%	75.63%	**80.01**%	73.84%
	3	6	12.61%	21%	66.39%	74.99%	64.29%
	4	5	11.75%	7.58%	**80.67**%	76.7%	**83.61**%
PCA-Rough-Fuzzy	1	16	20.17%	18.49%	61.35%	60%	62.06%
	2	5	9.25%	10.08%	**80.67**%	**81.65**%	**80.33**%
	3	6	16.81%	17.65%	65.55%	66.66%	65.57%
	4	8	10.08%	10.08%	79.83%	80.01%	80.01%
ICA-Fuzzy	1	NA	10.08%	40.34%	49.58%	80.01%	50%
	2	NA	10.08%	40.34%	49.58%	80.01%	50%
	3	NA	10.08%	40.34%	49.58%	80.01%	50%
	4	NA	10.08%	40.34%	49.58%	80.01%	50%
ICA-Rough-Fuzzy	1	NA	15.97%	15.13%	68.91%	68.33%	69.48%
	2	NA	10.08%	9.24%	**80.67**%	80.01%	**81.36**%
	3	NA	14.29%	18.49%	67.22%	71.66%	66.15%
	4	NA	8.4%	11.77%	79.83%	**83.34**%	78.12%

Table 4: A Comparison of the different computer-aided detection system results.

Algorithm	Best accuracy	Average accuracy	Average FN	Average FP
PIRF	84.03%	77.73%	8.82%	13.45%
PIF	83.19%	75.21%	12.61%	12.19%
IRF	80.67%	74.16%	13.66%	12.19%
PRF	80.67%	71.85%	14.08%	14.08%
PF	80.67%	71.64%	13.24%	15.12%
IF	49.58%	49.58%	40.34%	10.08%

Table 5: Observations of the different developed algorithms.

Algorithm	Observations
PIRF	It has the highest accuracy and recall percentage but not the highest precision
PIF	Needs a feature subset selection module and it has the highest average precision
IRF	Needs a dimensionality and noise reduction module and it has the highest average precision
PRF	Needs a feature extraction module
PF	Needs a feature extraction module and a feature subset selection module
IF	Needs a dimensionality and noise reduction module and a feature subset selection module and it has the lowest accuracy and recall percentage

Table 6: The influence of the number of PC on accuracy of the results while learning rate, mapping range, and threshold parameters are kept constants.

PC	Set no. 1	Set no. 2	Set no. 3	Set no. 4
5	62.19%	78.99%	69.75%	80.67%
6	63.03%	81.52%	69.75%	80.67%
7	62.19%	**83.19**%	69.75%	81.52%
8	**68.91**%	78.15%	**74.79**%	**84.03**%
9	68.07%	74.79%	73.95%	78.99%
10	67.23%	75.63%	71.43%	73.95%

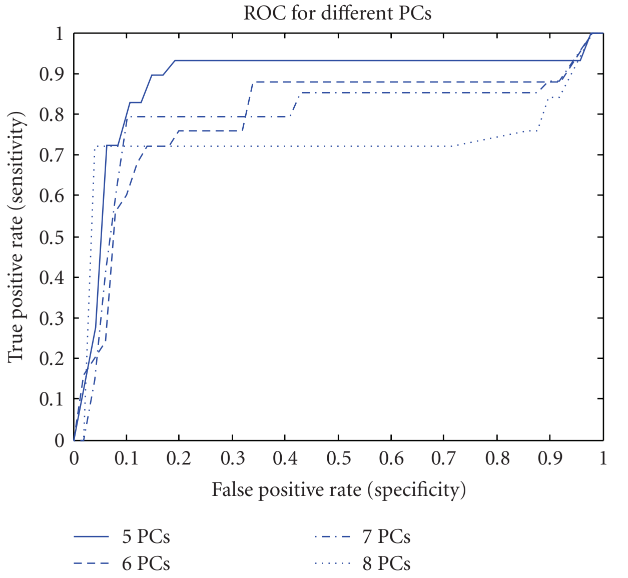


Figure 3: ROC plot for different values of selected principal components for testing set number 4.

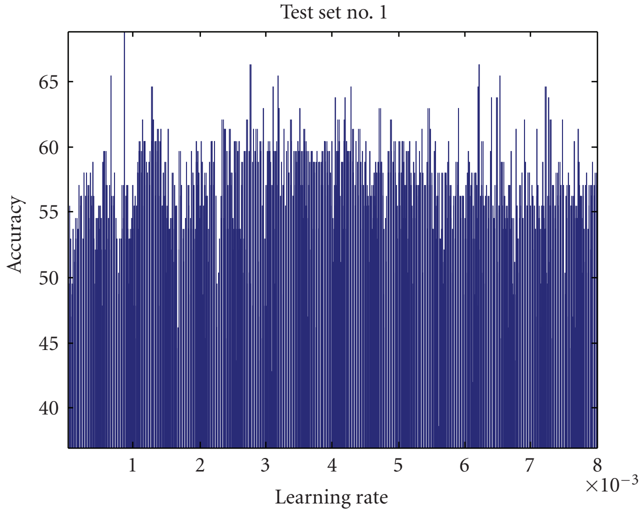


Figure 4: Learning rate impact on accuracy for test set number 1 (all other parameters were kept constant).

features will have redundant information and therefore will degrade the performance accuracy. However, if a small number is selected, extracted features cannot be estimated precisely and the fuzzy classifier performance will also be degraded.

Table 6 shows the highest accuracy for four testing sets using different numbers of the selected principal components while the other parameters are kept constant. Results also show that selecting less than 9 principal components achieves best results in all cases which means that less than 0.65% of the image features are selected for the 35 × 35 subimages and less than 0.4% of the image features are selected for the 45 × 45 subimages. This is in agreement with all reported literature that used PCA algorithm for dimensionality reduction [14, 15].

On the other hand, Figure 3 shows the Receiver operating characteristic (ROC) plot for a different number of selected principal components for testing set number 4. This figure is generated by plotting true positive rates against false positive rates. As the figure indicates, selecting five principal components produces the largest area under the curve which means that it produces the highest average accuracy.

Learning Rate. The estimation of the matrices W and S is affected by the learning rate, which determines the speed and accuracy of convergence to the optimal value. Since optimal values of W and S are unknown and they are data dependent, optimal value of η cannot be estimated adaptively. Also, since η represents the step size for ΔW, choosing a small value of it ensures accuracy but reduces the speed of convergence. Learning rate impact on four testing sets is shown in Figures 4, 5, 6, and 7 where all parameters were kept fixed except for the learning rate. Figure 8 shows the ROC plot for different values of the learning rate for testing set number 4. As the figure indicates, the smallest value of η (0.001) produces the largest area under the curve which means that it produces the highest average accuracy.

Momentum Method Constant. This constant determines the ratio of the previous ΔW that should be added to the current ΔW to increase the convergence speed of W. Since ΔW utilizes the natural gradient to find direction of W toward a minimum point, adding its previous value to its current

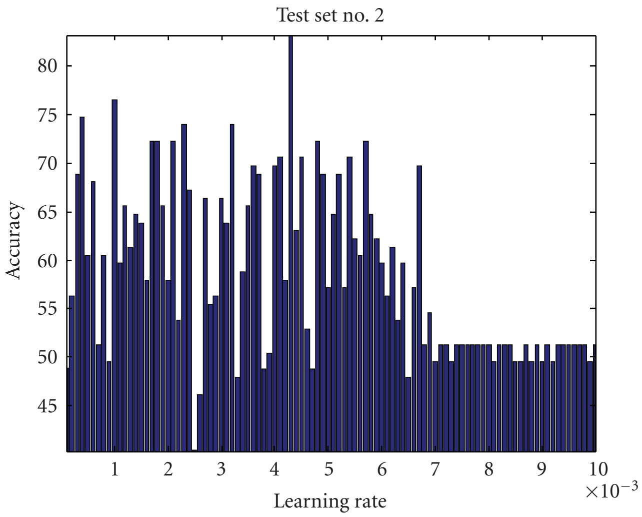


Figure 5: Learning rate impact on accuracy for test set number 2 (all other parameters were kept constant).

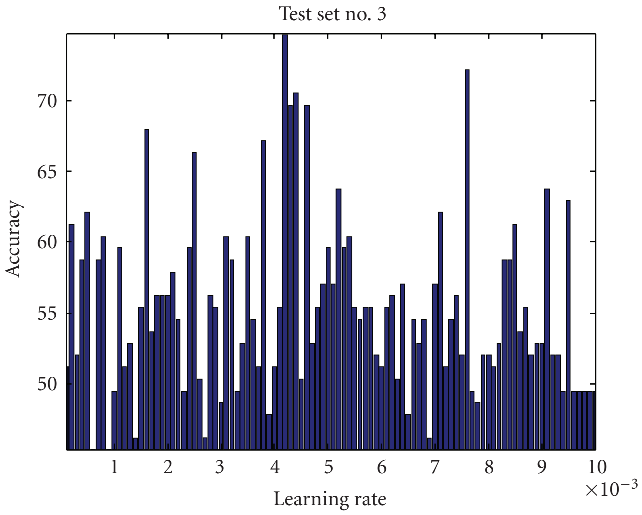


Figure 6: Learning rate impact on accuracy for test set number 3 (all other parameters were kept constant).

Table 7: Threshold T impact on accuracy for test set number 1 while learning rate, mapping range, and PC are kept constants.

T	Set no. 1
1	69%
0.75	65.55%
0.5	63.87%
0.25	63.87%

value pushes it toward the minimum point faster but does not change its direction.

Mapping Range. In investigating the mapping range values' effect on the accuracy of the results, we found that mapping

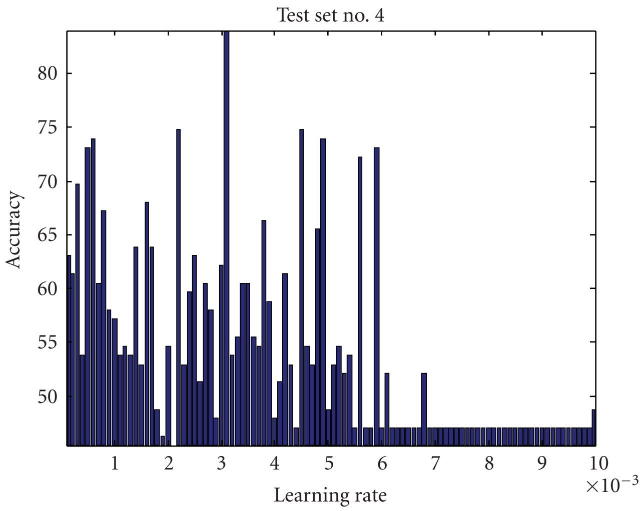


Figure 7: Learning rate impact on accuracy for test set number 4 (all other parameters were kept at constant values).

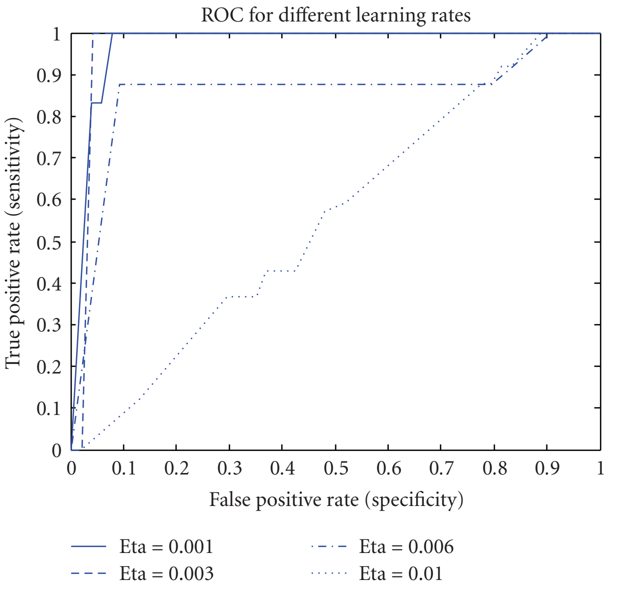


Figure 8: ROC plot for different values of the learning rate for testing set number 4.

the data into a limited range results in accuracy loss but simplifies computational complexity and processing time. Figures 9, 10, 11, and 12 show accuracy results versus mapping range for four testing sets while other parameters are kept constant. Figure 13 shows the ROC plot for different values of the mapping range for testing set number 4. As the figure indicates, choosing a mapping range in the interval $[0, 7]$ produces the largest area under the curve which correlates with the highest average accuracy.

Threshold-T. A threshold value T is necessary, (26), as a criteria to stop the Reduct procedure. This determines the

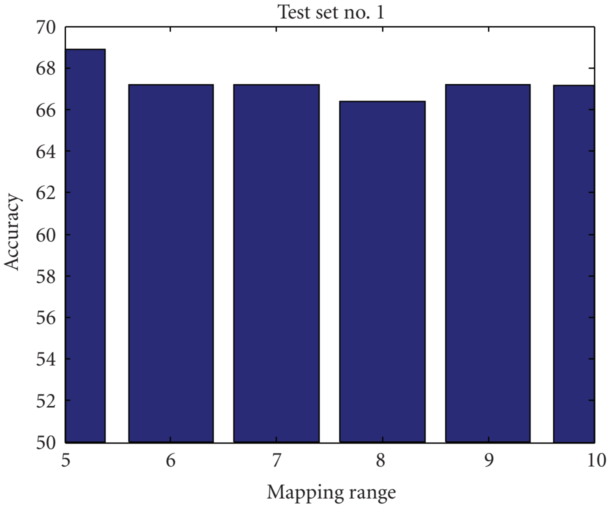


Figure 9: Mapping range impact on accuracy for test set number 1.

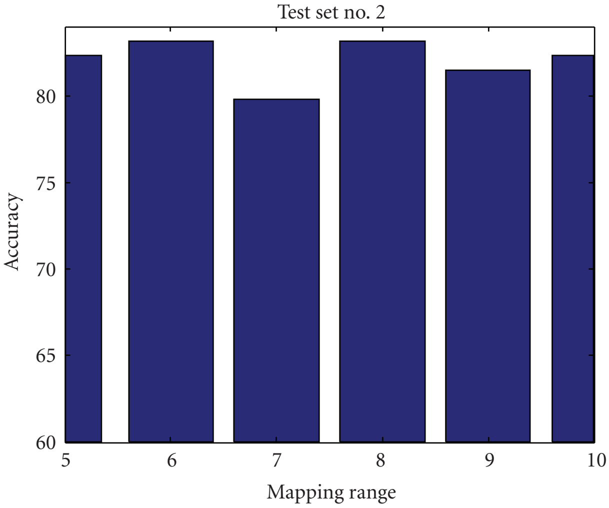


Figure 10: Mapping range impact on accuracy for test set number 2.

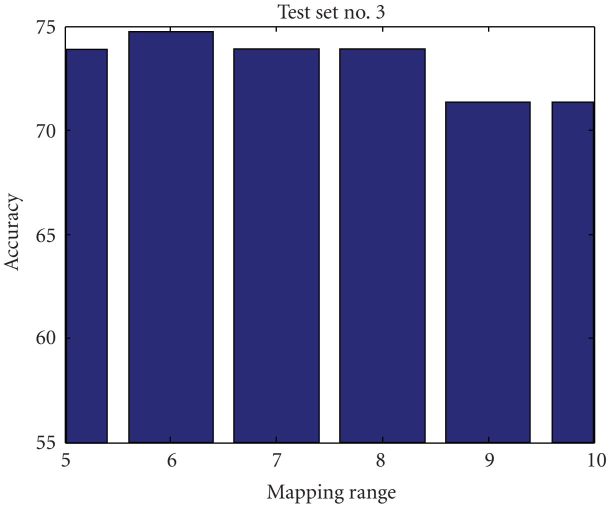


Figure 11: Mapping range impact on accuracy for test set number 3.

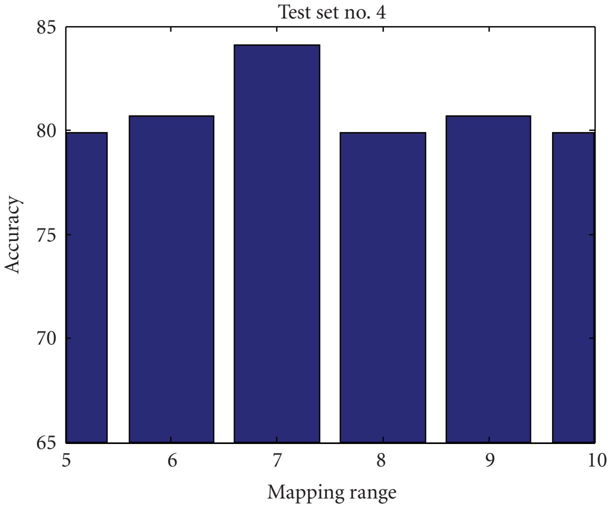


Figure 12: Mapping range impact on accuracy for test set number 4.

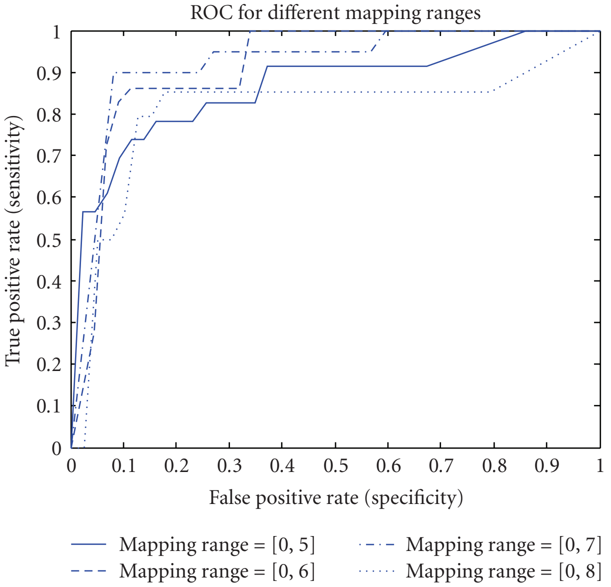


Figure 13: ROC plot for different values of the mapping range for testing set number 4.

number of selected features and consequently affects the classifier accuracy. Table 7 shows the impact of T on results of test set number 1. These results indicate that selecting a threshold equal to 1 achieves the highest performance. The optimum T value is the value, at which the Reduct attributes are complete, at which the number of inconsistent rows equals to that of the decision matrix. Furthermore, the cropped size impacts the accuracy of the results as shown in Table 3. As the table shows, the larger subimages (of size 45 × 45 pixels) resulted in the highest accuracy.

6. Conluding Remarks

A computer-aided detection system has been developed and implemented by integrating PCA, ICA, RSM, and a fuzzy classifier. Its performance is compared against the performance of PCA-ICA-Fuzzy, PCA-Fuzzy, PCA-Rough-Fuzzy, ICA-Fuzzy, and ICA-Rough-Fuzzy algorithms.

Results from Tables 3 and 4 indicate that PCA algorithm should be used in order to reduce FN rates at the expense of FP rates. It is shown that integrating RSM and PCA in one algorithm allows for a lower number of principal components to be selected while maintaining the performance accuracy as opposed to use PCA without RSM. Using ICA model and fuzzy classifier produced a CAD system with poor performance unless PCA is used for dimensionality reduction. RSM is used for further features reduction in order to reduce data inconsistency and consequently improve classifier performance. Results also indicate that PCA algorithm should be followed be ICA algorithm instead of RSM. Results of Table 3 indicate that the proposed PIRF algorithm is robust in comparison with the other algorithms. Finally, the proposed CAD algorithm reduces the FN rates considerably which is the main concern of CAD systems.

Parameter values as well as block size play a vital role in the system's performance and an investigation of this relation and perhaps automation of their selection is needed to further improve system's robustness. Although cumulants offer simple computations, they are sensitive to outliers (large values within the set). Therefore, an alternative route that may be worthwhile to investigate is to use a learning rule of the ICA algorithm that is based on negentropy instead of cumulants.

Acknowlegments

The authors would like to acknowledge Western Michigan University for its support and contributions to the Information Technology and Image Analysis (ITIA) Center, funded by the National Science Foundation Grant (MRI-0215356).

References

[1] National Cancer Institute, U.S. National Institute of Health, May 2009, http://www.cancer.gov/cancertopics/types/breast.

[2] J. Billingsley, "Radiologists' mammogram accuracy varies widely," June 2005, http://news.healingwell.com/index.php?p=news1&id=526229.

[3] A. E. Hassanien and J. M. H. Ali, "Enhanced rough sets rule reduction algorithm for classification digital mammography," *Journal of Intelligent Systems*, vol. 13, no. 2, pp. 151–171, 2004.

[4] X. Hu, T. Y. Lin, and J. Han, "A new rough sets model based on database systems," *Fundamenta Informaticae*, vol. 59, no. 2-3, pp. 135–152, 2004.

[5] S. Şahan, K. Polat, H. Kodaz, and S. Güneş, "A new hybrid method based on fuzzy-artificial immune system and *k*-nn algorithm for breast cancer diagnosis," *Computers in Biology and Medicine*, vol. 37, no. 3, pp. 415–423, 2007.

[6] A. Hassanien, "Fuzzy rough sets hybrid scheme for breast cancer detection," *Image and Vision Computing*, vol. 25, no. 2, pp. 172–183, 2007.

[7] Q. Hu, D. Yu, J. Liu, and C. Wu, "Neighborhood rough set based heterogeneous feature subset selection," *Information Sciences*, vol. 178, no. 18, pp. 3577–3594, 2008.

[8] C. Cornelis and R. Jensen, "A noise-tolerant approach to fuzzy-rough feature selection," in *Proceedings of the 17th IEEE International Conference on Fuzzy Systems*, pp. 1598–1605, 2008.

[9] N. Mac Parthaláin and Q. Shen, "Exploring the boundary region of tolerance rough sets for feature selection," *Pattern Recognition*, vol. 42, no. 5, pp. 655–667, 2009.

[10] Y. Caballero, D. Alvarez, R. Bello, and M. M. Garcia, "Feature selection algorithms using rough set theory," in *Proceedings of the 7th International Conference on Intelligent Systems Design and Applications (ISDA '07)*, pp. 407–411, 2007.

[11] A.-M. Yang, Y.-X. Yang, and S.-Y. Jiang, "Approaches of individual classifier generation and classifier set selection for fuzzy classifier ensemble," in *Proceedings of the 5th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD '08)*, vol. 1, pp. 519–524, 2008.

[12] A. Çelikyilmaz, I. B. Türksen, R. Aktas, M. Mete Doganay, and N. Basak Ceylan, "A new classifier design with fuzzy functions," in *Proceedings of the 11th International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing*, vol. 4482 of *Lecture Notes in Computer Science*, pp. 136–143, Toronto, Canada, May 2007.

[13] A. Bonarini, M. Matteucci, and M. Restelli, "Learning fuzzy classifier systems: architecture and exploration issues," *International Journal on Artificial Intelligence Tools*, vol. 16, no. 2, pp. 269–289, 2007.

[14] I. Abdel-Qader and F. Abu-Amara, "A computer-aided diagnosis system for breast cancer using independent component analysis and fuzzy classifier," *Modelling and Simulation in Engineering*, vol. 2008, Article ID 238305, 9 pages, 2008.

[15] R. Swiniarski, H. K. Lim, J. H. Shin, and A. Skowron, "Independent component analysis, principal component analysis and rough sets in hybrid mammogram classification," in *Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV '06)*, vol. 2, pp. 640–645, Las Vegas, Nev, USA, June 2006.

[16] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, New York, NY, USA, 2001.

[17] I. Christoyianni, A. Koutras, E. Dermatas, and G. Kokkinakis, "Computer aided diagnosis of breast cancer in digitized mammograms," *Computerized Medical Imaging and Graphics*, vol. 26, no. 5, pp. 309–319, 2002.

[18] G.-J. Jang and T.-W. Lee, "A maximum likelihood approach to single-channel source separation," *Journal of Machine Learning Research*, vol. 4, no. 7-8, pp. 1365–1392, 2003.

[19] M. Girolami and C. Fyfe, "Negentropy and kurtosis as projection pursuit indices provide generalized ICA algorithms," in *Advances in Neural Information Processing Systems, Blind Signal Separation Workshop*, pp. 249–266, Aspen, Colo, USA, December 1996.

[20] F. Abujarad and A. Omar, "Comparison of independent-component-analysis (ICA) algorithms for GPR detection of non-metallic land mines," in *Image and Signal Processing for Remote Sensing XII*, vol. 6365 of *Proceedings of SPIE*, Stockholm, Sweden, September 2006.

[21] M. J. Del Jesus, P. Gonzalez, and F. Herrera, "Multiobjective genetic algorithm for extracting subgroup discovery fuzzy rules," in *Proceedings of the IEEE Symposium on Computational Intelligence in Multicriteria Decision Making (MCDM '07)*, pp. 50–57, 2007.

[22] A. Riid and E. Rustern, "Neuro-fuzzy extraction of interpretable fuzzy rules from data," in *Proceedings of the IEEE*

International Conference on Systems, Man and Cybernetics (SMC '04), vol. 3, pp. 2266–2271, The Hague, The Netherlands, October 2004.

[23] H.-E. Lee, K.-H. Park, and Z. Z. Bien, "Iterative fuzzy clustering algorithm with supervision to construct probabilistic fuzzy rule base from numerical data," *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 1, pp. 263–277, 2008.

[24] http://peipa.essex.ac.uk/ipa/pix/mias.

[25] S.-I. Amari, "Natural gradient works efficiently in learning," *Neural Computation*, vol. 10, no. 2, pp. 251–276, 1998.

[26] H. H. Yang, S.-I. Amari, and A. Cichocki, "Information-theoretic approach to blind separation of sources in non-linear mixture," *Signal Processing*, vol. 64, no. 3, pp. 291–300, 1998.

[27] M. Bartlett, M. Lades, and T. Sejnowski, "Independent component representation for face recognition," in *Human Vision and Electronic Imaging III*, Proceedings of SPIE, pp. 528–539, San Jose, Calif, USA, January 1998.