

## Research Article

# A Computer-Aided Diagnosis System for Breast Cancer Using Independent Component Analysis and Fuzzy Classifier

**Ikhlas Abdel-Qader and Fadi Abu-Amara**

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

Correspondence should be addressed to Ikhlas Abdel-Qader, [abdelqader@wmich.edu](mailto:abdelqader@wmich.edu)

Received 27 August 2007; Revised 10 January 2008; Accepted 11 March 2008

Recommended by Ewa Pietka

Screening mammograms is a repetitive task that causes fatigue and eye strain since for every thousand cases analyzed by a radiologist, only 3–4 are cancerous and thus an abnormality may be overlooked. Computer-aided detection (CAD) algorithms were developed to assist radiologists in detecting mammographic lesions. In this paper, a computer-aided detection and diagnosis (CADD) system for breast cancer is developed. The framework is based on combining principal component analysis (PCA), independent component analysis (ICA), and a fuzzy classifier to identify and label suspicious regions. This is a novel approach since it uses a fuzzy classifier integrated into the ICA model. Implemented and tested using MIAS database. This algorithm results in the classification of a mammogram as either normal or abnormal. Furthermore, if abnormal, it differentiates it into a benign or a malignant tissue. Results show that this system has 84.03% accuracy in detecting all kinds of abnormalities and 78% diagnosis accuracy.

Copyright © 2008 I. Abdel-Qader and F. Abu-Amara. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1. INTRODUCTION

Breast cancer is considered one of the most common and fatal cancers among women in the USA [1]. According to National Cancer Institute, 40 480 women died due to this disease and on average every three minutes one woman is diagnosed with this cancer. Right now there are over two and a half million women in the US who have been treated from it [1]. Radiologists visually examine mammograms to search for signs of abnormal regions. They usually look for clusters of microcalcifications, architectural distortions, or masses.

Early detection of breast cancer via mammography improves treatment chances and survival rates [2]. Unfortunately, mammography is not perfect. False positive (FP) rates are 15–30% due to the overlap in the appearance of malignant and benign abnormalities while false negative (FN) rates are 10–30%. A result of FP is defined to be when a radiologist reports a suspicious change in the breast but no cancer is found after further examinations. Therefore, it leads to unnecessary biopsies and anxiety. A result of FN means failure to detect or correctly characterize breast cancer in a case of which later tests conclude that cancer is present. Nonetheless, mammography has an overall accuracy rate of 90% [3].

CAD algorithms have been developed to assist radiologists in detecting mammographic lesions. These systems are regarded as a second reader, and the final decision is left to the radiologist. CAD algorithms have improved total radiologist accuracy of detection of cancerous tissues [4]. CADD algorithms are considered as an extremely challenging task for various reasons. First, the imaging system may have serious imperfections. Second, the image analysis task is compounded by the large variability in the appearance of abnormal regions. Finally, abnormal regions are often hidden in dense breast tissue. The goal of the detection stage is to assist radiologists in locating abnormal tissues.

Many methods have been proposed in the literature for mammography detection and diagnosis utilizing a wide variety of algorithms. Chang et al. [5] developed a 3D snake algorithm that finds the tumor's contour after reducing the noise levels and followed by an edge enhancement process. Finally, the tumor's contour is estimated by using the gradient vector flow snake. Kobatake et al. [6] proposed the iris filter to detect lesions as suspicious regions with a low contrast compared to their background. The proposed filter has the features' extraction ability of malignant tissues. Bocchi et al. [7] developed an algorithm for microcalcification

detection and classification by which the existing tumors are detected using a region growing method combined with a neural network-based classifier. Then, microcalcification clusters are detected and classified by using a second fractal model. Also, Li et al. [8] developed a method for detecting tumors using a segmentation process, adaptive thresholding, and modified Markov random fields, followed by a classification step based on a fuzzy binary decision tree. Bruce and Adhami [9] used the modulus-maxima technique of discrete wavelet transform as a feature extraction technique combined with a Euclidean distance classifier. A radial distance measure of mass boundaries is used to extract multiresolution shape features. Finally, the leave-one-out and apparent methods are used to test their proposed technique. Peña-Reyes and Sipper [10] applied a combined fuzzy-genetic approach with new methods as a computer-aided diagnosis system. Zheng and Chan [11] combined artificial intelligent methods with the discrete wavelet transform to build an algorithm for mass detection. Hassanien and Ali [12] proposed an enhanced rough set technique for feature reduction and classification. Swiniarski and Lim [13] integrated ICA with rough set model for breast-cancer detection. First, features are reduced and extracted using ICA. Then, extracted features are selected using a rough set model. Finally, a rough set-based method is used for rule-based classifier design.

This work is based on integrating PCA, ICA, and fuzzy classifier to identify and label suspicious regions from digitized mammograms. The rest of this paper is organized as follows: Section 2 presents PCA and ICA algorithms and covers fuzzy logic adaptation as a classifier. The proposed integrated approach is presented in Section 3. Section 4 presents the experimental results followed by the conclusions in Section 5.

## 2. BACKGROUND

### 2.1. PCA

PCA is a decorrelation-based technique that finds the basis vectors for a subspace in order to select the most important information. PCA consists of two phases. The first phase finds  $\nu$  uncorrelated and orthogonal vectors; and the second phase projects the testing data into a subspace spanned by these  $\nu$  vectors [14]. PCA algorithm can be presented as follows:

- (i) construct  $R_{\text{train}}$  matrix with dimension  $N \times M$ , where  $N$  is the total number of training subimages and  $M$  is the size of each square subimage; then, generate its normalized matrix  $P_{M \times N}$ ;
- (ii) covariance matrix is constructed using

$$C_{N \times N} = P_{N \times M}^T P_{M \times N}; \quad (1)$$

- (iii) let  $\lambda_i$  and  $E_i$ ,  $i = 1, 2, \dots, M$ , be its eigenvalues and eigenvectors that satisfy the equation  $C E_i = \lambda_i E_i$ , where  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M \geq 0$ ; discard of all eigenvalues less than  $T$  (a predetermined threshold) and retain

the rest (the principal components) to produce the reduced matrix  $R_{M \times \nu}^R$ .  $T$  is calculated using

$$T = \frac{\sum_{k=1}^{\nu} \lambda_k}{\sum_{q=1}^M \lambda_q}. \quad (2)$$

The given testing data  $R_{\text{test}}$  is projected into the space spanned by the reduced training matrix  $R_{M \times \nu}^R$  using

$$W_{\nu \times N} = (R_{M \times \nu}^R)^T R_{\text{test}, M \times N}. \quad (3)$$

### 2.2. ICA

Higher-order statistics, such as ICA techniques, are used to compensate for PCA shortcomings. ICA is based on the use of moments and cumulants up to fourth-order to describe any distribution of a random variable.

In general, ICA is a relatively new technique developed to find a linear representation of nongaussian data so that the data components are statistically as independent as possible. ICA has the ability to describe localized shape variations and it does not require a Gaussian distribution of the data as in PCA. However, the resulting vectors are not ordered; and, therefore, ICA requires a method for ordering the resulting vectors.

The statistical latent variables model is used to define ICA. Assuming that we have  $n$  linear mixtures  $r_1 \dots r_n$  of  $n$  independent components  $s_1 \dots s_n$  according to

$$r = \sum_{i=1}^n a_i s_i \quad \text{or} \quad R = AS. \quad (4)$$

The digital mammographic image  $R$  is considered as a mixture of linear combination of statistically independent source regions  $S$  where  $A$ , the mixing matrix, and its coefficients describe uniquely the mixed source regions and can be used as extracted features. After estimating the matrix  $A$  and its inverse  $W$  (the separating matrix), the independent components can be estimated using

$$S = WR. \quad (5)$$

### 2.3. Fuzzy classifier

Fuzzy logic can be interpreted as the emulation of human reasoning on computers [15]. Fuzzy rules are more comprehensible than crisp rules since they can be expressed in terms of linguistic concepts. The value of the linguistic variable is not a number but a word. For example, the linguistic variable "size" might have the values "small," "medium," and "large." Each one of these values is called a fuzzy set when implemented using fuzzy logic and thus fuzzy sets can be used to model linguistic variables.

Fuzzy classifier is ideally suited to the labeled observed data to provide interpretable solutions. It handles imprecise data and the resulting fuzzy rules are interpretable, that is, fuzzy classifier structure can be analyzed through its semantic structure. There are two different methods for development of fuzzy classifiers; approximate and descriptive fuzzy rule base.

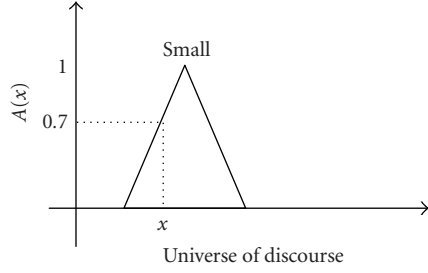


FIGURE 1: A triangular membership function of the fuzzy set “Small.”

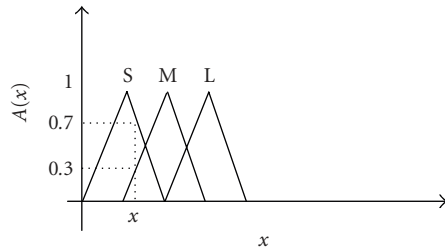


FIGURE 2: Fuzzy space of the object of Figure 1 that consists of three fuzzy sets: “Small,” “Medium,” and “Large.”

Each fuzzy rule is defined using membership function of fuzzy sets in an approximate fuzzy rule base which is implemented in this work. Values of the linguistic variable can be described in terms of numerals using membership functions. The object membership degree to a fuzzy set defines a membership function. Its domain is the universe of discourse (all values an object may take) and its range of the interval  $[0.0, 1.0]$ . A commonly used membership function is the triangular function. Figure 1 shows a triangular membership function of a fuzzy set “Small.”

In Figure 1, an object  $x$  has a membership degree of 0.7 to the fuzzy set “Small.” A fuzzy space is defined to be the set of fuzzy sets that define fuzzy classes for a particular object as shown in Figure 2.

Fuzzy space allows the object to partially belong to different classes simultaneously. This idea is very useful in cases where the difference between classes is not well defined. For example, the object  $x$  has a membership degree of 0.7 to the fuzzy set “Small” and 0.3 to the fuzzy set “Medium.” Similarly, in mammographic images, the difference between benign/malignant and normal/abnormal subimages is not well defined. For example, an abnormal subimage may be classified as benign rather than malignant which can be described in terms of numerals using membership functions as it has a membership degree of 0.7 to the fuzzy set “benign” and 0.3 to the fuzzy set “malignant.” Fuzzy membership functions are easy to implement and their fuzzy inference engines are fast.

In descriptive fuzzy rule base, linguistic variables are commonly defined by fuzzy if-then rules where labels  $A_{ij}$  are used to represent a discrete set of linguistic fuzzy sets. For example, fuzzy classification rules that describe each class of

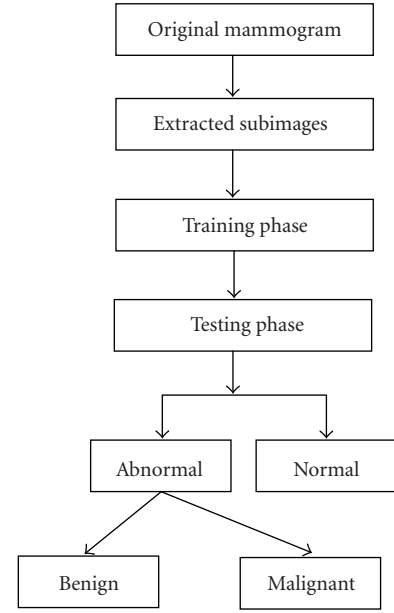


FIGURE 3: Block diagram of the proposed CADD system.

subimages may be developed to represent each class. Fuzzy rules have the form

$$\text{IF antecedent THEN consequent [weight].} \quad (6)$$

Fuzzy rules can also be expressed as

$$R_i : \text{If } x_1 \text{ is } A_{i1} \text{ and } \dots x_t \text{ is } A_{it} \text{ then } Y = \text{Class}_i \text{ [weight],} \quad (7)$$

where  $Y$  represents the decision class (i.e., normal, abnormal, benign, or malignant) and  $A_{ij}$  represents a fuzzy set for  $j: 1, \dots, t$ th selected feature.

### 3. PROPOSED CADD ALGORITHM

In this section, a computer-aided detection and diagnosis algorithm of suspicious regions in mammograms is developed. PCA algorithm is used as a dimensionality reduction module followed by ICA as a feature extraction module. Finally, a fuzzy classifier is used to classify testing subimages into normal/abnormal and at a later stage to classify the abnormal subimages into malignant/benign as a diagnosis system. Figure 3 presents the general framework for this system.

#### 3.1. Subimages generation

MIAS database has a total of 119 regions of suspicion (ROS) divided into 51 malignant and 68 benign. Two different sets of abnormal subimages, each set consists of 119 ROS, are cropped and scaled into  $35 \times 35$  and  $45 \times 45$  pixels based on the center of each abnormality.

Then, five different sets of normal subimages, each set consists of 119 subimages, are cropped and scaled randomly from normal MIAS mammograms where two sets of size  $35 \times 35$  and three sets of size  $45 \times 45$  pixels.

Each set of abnormal subimages is mixed with one set of normal subimages every time and then divided into two groups; one for training phase and the other group for testing phase as shown in Table 1.

Each training set is used to create the matrix  $R_{\text{train}}$  with dimension  $N \times M$  where each row contains a subimage. The training matrix dimensionality is reduced by using PCA algorithm to generate  $R^R$ . Then, the covariance matrix is estimated by using

$$C_{N \times v} = R_{\text{train}_{N \times M}} R_{M \times v}^R. \quad (8)$$

### 3.2. Unsupervised learning

Estimation of the separating matrix,  $W$ , and the independent source regions,  $S$ , is done in an unsupervised manner. The independent source regions are estimated by using (9), where  $(R^R)^T$  is the transpose of the reduced matrix  $R^R$ . The separating matrix,  $W$ , is initialized to the identity matrix yielding

$$S = W(R^R)^T. \quad (9)$$

To reach the maximum statistical independence of  $S$ , the nonlinear function  $\Phi(S)$  is used to estimate the marginal probability density function of  $S$  using its central moments and cumulants. Minimum mutual information algorithm [16] is used to estimate  $\Phi(S)$  as shown in (10)–(14). Equations (10) and (11) are used to estimate the  $i$ th central moments and cumulants where  $E$  is the expected value and  $\mu$  is the mean of the current feature  $r$ . Equations (12)–(14) are used to estimate  $\Phi(S)$  ( $\circ$  indicates the Hadamard product of two matrices)

$$m_i = E(r - \mu)^i, \quad (10)$$

$$k_3 = m_3, \quad k_4 = m_4 - 3, \quad (11)$$

$$\Phi(S) = f_1(k_3, k_4) \circ S^2 + f_2(k_3, k_4) \circ S^3, \quad (12)$$

$$f_1(k_3, k_4) = 0.5 k_3 (4.5 k_4 - 1), \quad (13)$$

$$f_2(k_3, k_4) = 1.5(k_3)^2 + \frac{1}{6} k_4 (4.5 k_4 - 1). \quad (14)$$

Natural gradient descent method [16] is used to estimate the change of  $W$  according to  $dW/dt = \eta[I - \Phi(S)S^T]W$ , where  $\eta(t)$  is the learning rate and  $I$  is the identity matrix. If  $dW/dt$  is not close to zero,  $W$  is updated using

$$W_i(t+1) = W_i(t) + \frac{dW}{dt}. \quad (15)$$

Finally, selected features resulting from the training process are estimated using minimum square error method (MSE) [17, 18].

(i) From (8), the training matrix is reconstructed as

$$R_{\text{train}} \approx C(R^R)^T. \quad (16)$$

(ii) Substitute (9) into (16):

$$R_{\text{train}} \approx C_{N \times v}(R^R)^T = C_{N \times v}AS. \quad (17)$$

(iii) There, the reduced dimensionality selected features from the training set are estimated by

$$Q_{\text{train}} = CA. \quad (18)$$

Same procedure followed for training data is used for testing; and  $R_{\text{test}}$  is projected into the reduced matrix ( $R^R$ ) from the training procedure. The reduced dimensionality extracted features from the testing procedure are estimated by using

$$Q_{\text{test}} = R_{\text{test}}R^RA. \quad (19)$$

### 3.3. Fuzzy classifier modeling

The matrices  $Q_{\text{train}}$  and  $Q_{\text{test}}$  contain the reduced dimensionality extracted features from subimages where each one of size  $N$  by  $v$ . Each class of subimages (normal, abnormal, benign, and malignant) is represented by a single fuzzy rule by aggregating the membership functions of each antecedent fuzzy set using the information about selected feature values of training subimages.

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

- (1) Four activation functions  $\mu_{\text{bs}}$ ,  $\mu_{\text{ms}}$ ,  $\mu_{\text{as}}$ ,  $\mu_{\text{ns}}$ , with each one is of size  $N$  by 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. Each one represents the degree of activation of the selected feature values and so these parameters are defined as
  - (i)  $\mu_{\text{bs}}$ : represents the degree of activation for the benign testing subimages,
  - (ii)  $\mu_{\text{ms}}$ : represents the degree of activation for the malignant testing subimages,
  - (iii)  $\mu_{\text{as}}$ : represents the degree of activation for the abnormal testing subimages, and
  - (iv)  $\mu_{\text{ns}}$ : represents the degree of activation for the normal testing subimages.
- (2) Since subimages have different intensities and the goal is to reduce the variation and the computational complexity, the selected features of  $Q_{\text{train}}$  and  $Q_{\text{test}}$  are mapped into a limited range of  $[r1, r2]$  using

$$q(x, y) = r_1 + \frac{(q(x, y) - \min(q))(r_2 - r_1)}{\max(q) - \min(q)}. \quad (20)$$

- (3) Using (21), membership functions of fuzzy sets of the testing subimages are obtained from the product space of the selected features from the training phase:

$$A_{ij}(x_j) = \frac{s_i(x_j)}{s(x_j)}, \quad i = 1, \dots, v; j = 1, \dots, N, \quad (21)$$

where  $s_i(x_j)$  represents number of samples of the current feature  $x_j$ ,  $s(x_j)$  represents the total number of all samples in the current feature  $x_j$ , that is, the product space of the current feature. Also, the subscript ( $j$ )

TABLE 1: Different sets used to evaluate the detection algorithm performance.

#	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	35 × 35
3	60	59	119	59	60	119	45 × 45
4	60	59	119	59	60	119	45 × 45
5	60	59	119	59	60	119	45 × 45

is the index for the selected feature for each training subimage, and  $(i)$  is the index for the current processed sample of the current feature.

- (4) The membership functions are normalized by using

$$A_{ij}(x_j) = \frac{A_{ij}(x_j)}{\max_{ij}(A_j(x_j))}. \quad (22)$$

- (5) The degree of activation of the developed membership functions is computed for the testing subimages for  $\mu_{as}$ ,  $\mu_{ns}$  in the detection phase and for  $\mu_{bs}$ ,  $\mu_{ms}$  in the diagnosis phase by aggregating estimated membership functions:

$$\mu_i(x) = \sum_{j=1}^N A_{ij}(x_j). \quad (23)$$

- (6) There are many methods used in the literature to determine to which class a subimage belongs (i.e., normal/abnormal or benign/malignant). An efficient one is the maximum algorithm. It classifies the testing subimage into the class that has the maximum degree of activation according to (24) where  $C_1$  is used as an index of a testing subimage being identified as normal or abnormal and  $C_2$  for being identified as benign or malignant:

$$\begin{aligned} C_1 &= \max(\mu_{as}(x), \mu_{ns}(x)), \\ C_2 &= \max(\mu_{ms}(x), \mu_{bs}(x)). \end{aligned} \quad (24)$$

#### 4. EXPERIMENTAL RESULTS

Table 2 shows results of the proposed CADD algorithm against PCA and ICA algorithms for the same testing data using fuzzy classifier. Algorithm accuracy is defined as the ratio between number of correctly classified testing subimages and total number of testing subimages. Results demonstrate that combining ICA and PCA algorithms improves the total algorithm performance in all testing sets over usage of PCA algorithm only. PCA algorithm has a best result of 80.67% while 84.03% for the proposed CADD algorithm as shown in Table 2. The proposed algorithm improved PCA algorithm accuracy with an average of 8.56% for all tests.

Table 2 also shows the simulation results of ICA algorithm versus the proposed CADD algorithm. ICA algorithm has an accuracy of 49.58% in all testing sets. In contrast, the best result of applying the proposed CADD algorithm is 84.03%. These results indicate that using PCA algorithm for

dimensionality reduction before ICA algorithm improves the ICA algorithm accuracy with an average of 50.51%. Results from ICA algorithm show that fuzzy classifier performance is degraded when no dimensionality reduction module is implemented. A fuzzy classifier requires features reduction method in order to minimize total number of membership functions and improves its accuracy. As for ICA algorithm alone, each subimage has larger number of selected features and therefore fuzzy classifier performance is degraded in all testing subimages.

The experimental results of the proposed CADD algorithm as a computer-aided diagnosis system are shown in Table 3. The best result is 78% where 15 malignant subimages out of 25 are correctly classified and 31 benign subimages out of 34 are correctly classified.

This system uses several parameters that impact the performance and accuracy of results such as the number of selected principal components, learning rate, and mapping range.

##### 4.1. Number of selected PC

Using PCA algorithm to reduce data dimensionality as a pre-processing step for ICA algorithm affects the total algorithm accuracy. In Table 4, simulation results on test sets 1–5 (PC indicates the number of selected principal components) are shown. These results indicate that selecting less than 11 principal components achieves acceptable results in all simulations. This means that less than 0.81% of principal components are selected for subimages of size 35 × 35 pixels and less than 0.5% of principal components are selected for subimages of size 45 × 45 pixels. This is harmony with all literature that used PCA algorithm for dimensionality reduction.

##### 4.2. Learning rate

The learning rate for computing the change in  $W$  for ICA algorithm determines the speed of convergence for  $dW/dt$  and it impacts the total algorithm accuracy. Figures 4–8 show learning rate impact on test sets 1–5. It can be concluded that choosing a learning rate close to 0.0045 produce acceptable results for all sets.

##### 4.3. Mapping range

Figures 9–13 show the accuracy of the results versus the mapping range values for all test sets 1–5 and it can be concluded

TABLE 2:  $F_P$  and  $F_N$ ; and total PCA, ICA, and PCA-ICA algorithms accuracy.

Set	PCA				ICA			CADD			
	PC	$F_P$	$F_N$	Accuracy	$F_P$	$F_N$	Accuracy	PC	$F_P$	$F_N$	Accuracy
1	19	17.65%	25.21%	57.14%	10.08%	40.34%	49.58%	25	15.13%	18.48%	66.39%
2	20	26.05%	10.08%	63.87%	10.08%	40.34%	49.58%	10	12.61%	18.48%	68.91%
3	5	10.08%	14.29%	75.63%	10.08%	40.34%	49.58%	5	9.24%	6.73%	84.03%
4	6	12.61%	21%	66.39%	10.08%	40.34%	49.58%	5	20.17%	10.08%	69.75%
5	5	11.75%	7.58%	80.67%	10.08%	40.34%	49.58%	6	7.56%	8.41%	84.03%

TABLE 3: Computer-aided diagnosis using CADD algorithm.

Set	Training set			Testing set			Size-pixels	PC	CADD Algorithm		
	Benign	Malignant	Total	Benign	Malignant	Total			FP	FN	Accuracy
1	34	26	60	34	25	59	$35 \times 35$	36	5.1%	16.9%	78%
2	34	26	60	34	25	59	$45 \times 45$	6	8.48%	13.55%	77.97%

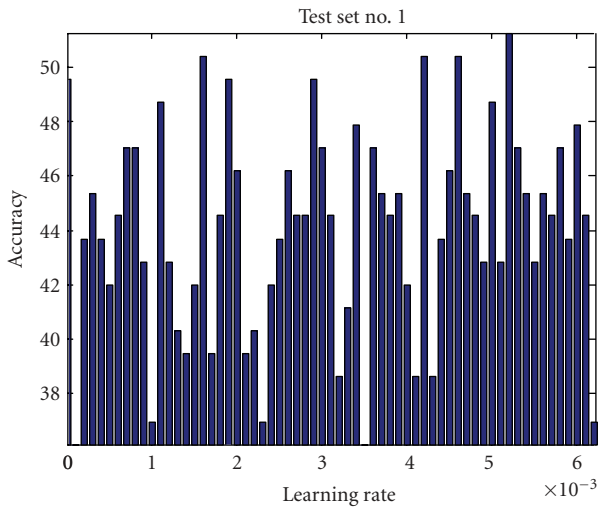


FIGURE 4: Learning rate impact on algorithm accuracy for test set no. 1 where other parameters are kept constant.

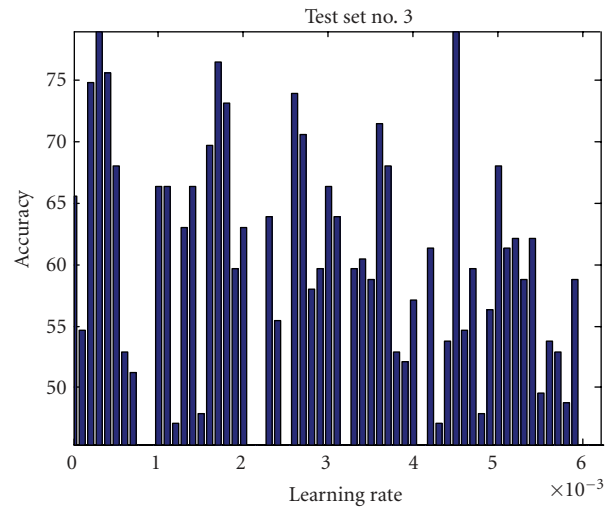


FIGURE 6: Learning rate impact on algorithm accuracy for test set no. 3 where other parameters are kept constant.

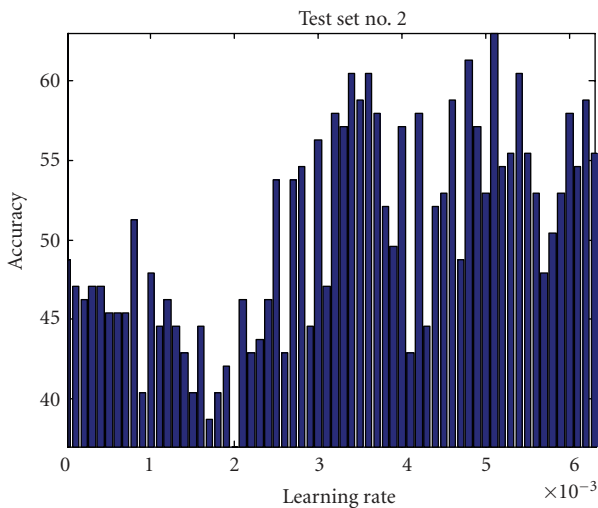


FIGURE 5: Learning rate impact on algorithm accuracy for test set no. 2 where other parameters are kept constant.

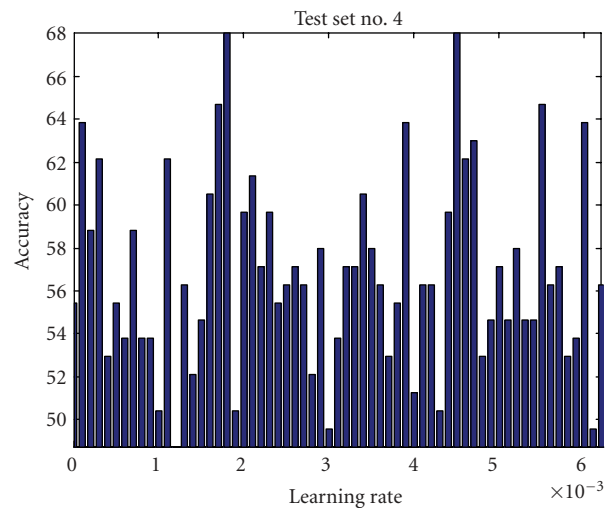


FIGURE 7: Learning rate impact on algorithm accuracy for test set no. 4 where other parameters are kept constant.

TABLE 4: Number of selected principal components impact on algorithm accuracy where learning rate and mapping range of each set are kept fixed.

PC	Set no. 1	Set no. 2	Set no. 3	Set no. 4	Set no. 5
5	61.35%	65.55%	<b>84%</b>	<b>69.75%</b>	81.52%
6	55.46%	64.71%	79.83%	68.07%	<b>84.03%</b>
7	62.19%	66.39%	78.99%	68.91%	82.35%
8	<b>66.39%</b>	66.39%	78.99%	65.55%	78.15%
9	59.66%	66.39%	70.59%	67.23%	76.47%
10	58.82%	<b>68.91%</b>	80.67%	66.39%	74.79%
11	63.03%	65.55%	69.75%	63.87%	75.63%
12	58.82%	60.5%	72.27%	63.03%	78.15%
13	62.19%	63.87%	70.59%	63.87%	74.79%
14	63.87%	62.19%	73.95%	62.19%	77.31%
15	57.98%	63.03%	69.75%	63.03%	73.95%
16	62.19%	59.66%	68.07%	63.03%	76.47%
17	62.19%	67.23%	72.29%	63.87%	77.31%
18	63.03%	60.5%	71.43%	62.19%	76.47%
19	64.71%	67.23%	72.29%	64.71%	79.83%
20	62.19%	64.71%	79.83%	62.19%	73.95%
21	60.5%	66.39%	74.79%	61.35%	80.67%
22	63.03%	66.39%	78.15%	62.19%	71.43%
23	63.03%	63.87%	80.67%	63.87%	74.79%
24	58.82%	60.5%	73.95%	62.19%	80.67%
25	59.66%	60.5%	68.91%	63.03%	79.83%

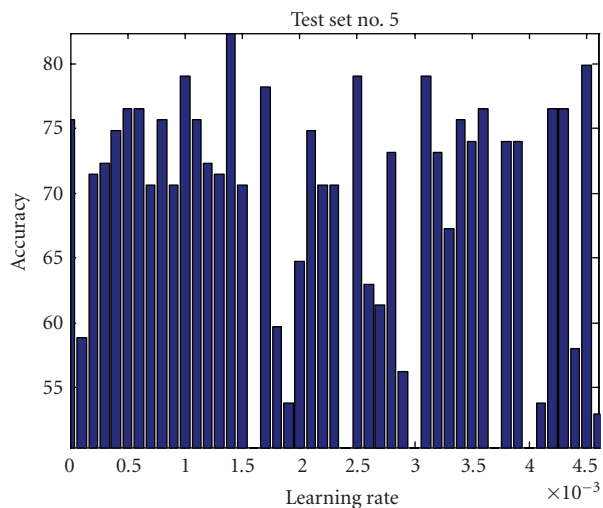


FIGURE 8: Learning rate impact on algorithm accuracy for test set no. 5 where other parameters are kept constant.

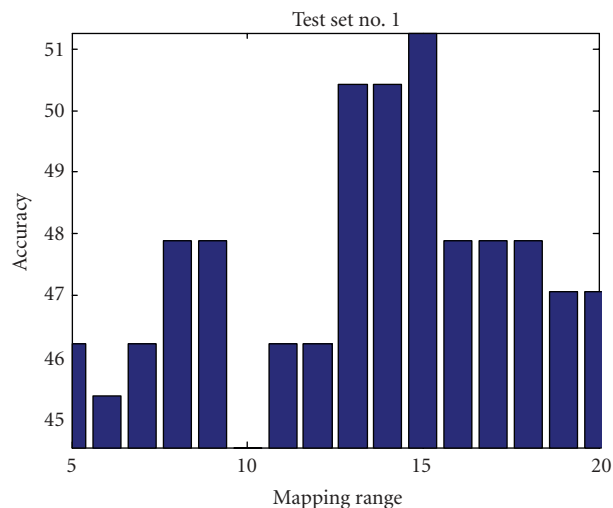


FIGURE 9: Mapping range impact on algorithm accuracy for test set no. 1.

that choosing a mapping range equal to  $[0, 9]$  or  $[0, 15]$  is acceptable for all testing sets.

The proposed system performance is a parameter-dependent and an investigation of this dependency is outside this presentation but rather is left for future investigations. Efforts developed earlier such as in [19, 20] can be investigated. Estimating the parameters will continue to be one of the main disadvantages of algorithms such ICA where human intervention is needed.

In other classification methods such as in fractal models, [7], a set of 30 mammograms are used that contains single and clustered microcalcifications. 50 subimages are extracted and divided into 30 subimages for the training phase and 20 subimages for the testing phase. Results of using two different multilayer subnetworks in neural network-based classifier indicate that the proposed system has a classification accuracy of 90%. Also, in discrete wavelet transform method [9], a set of 60 mammograms are used. Masses are segmented

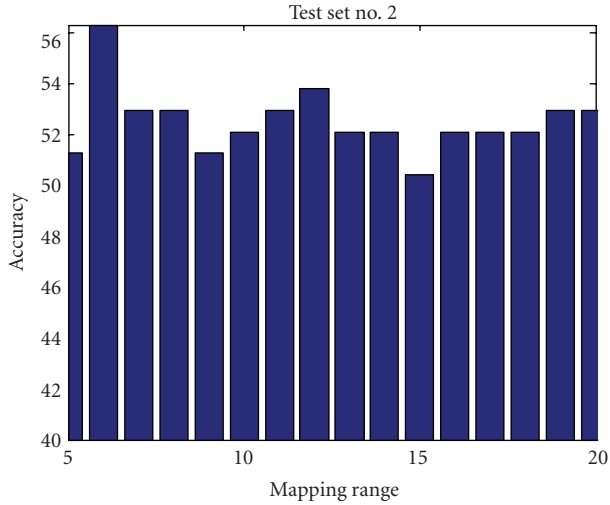


FIGURE 10: Mapping range impact on algorithm accuracy for test set no. 2.

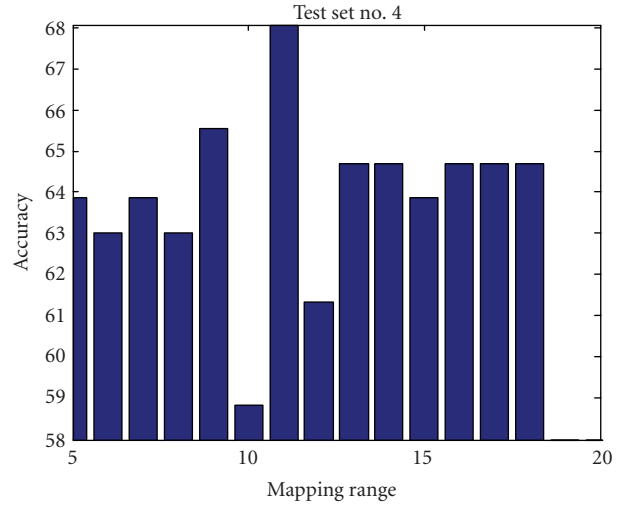


FIGURE 12: Mapping range impact on algorithm accuracy for test set no. 4.

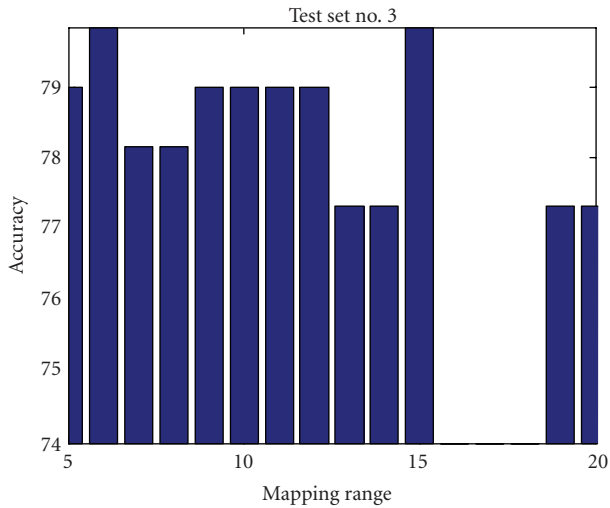


FIGURE 11: Mapping range impact on algorithm accuracy for test set no. 3.

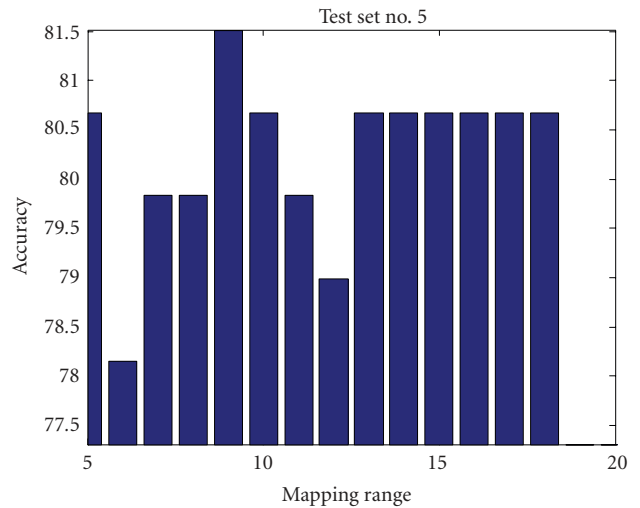


FIGURE 13: Mapping range impact on algorithm accuracy for test set no. 5.

manually as a preprocessing step for the classification system. The proposed system classifies masses into round, nodular, or stellate. Results indicate a classification accuracy of 83%. In [13], 330 subimages are cropped and scaled into sizes of  $20 \times 20$ ,  $40 \times 40$ , and  $60 \times 60$  pixels from all MIAS mammograms as one subimage from each mammogram. Results using ICA-Rough indicate a classification accuracy of 82.22% for subimages of size  $60 \times 60$  pixels and for PCA-Rough of 88.57% for subimages of size  $40 \times 40$  pixels.

Furthermore, Table 2 shows that each test set has different algorithm accuracy so cropping size for example has an impact on the results.

## 5. CONCLUDING REMARKS

A CADD system has been developed and implemented. Its framework is based on integrating PCA, ICA, and fuzzy

logic. The performance of the proposed CADD is compared against PCA and ICA performance individually. Extensive simulations using 833 subimages are performed. These results indicate that combining ICA and PCA algorithms improves PCA algorithm accuracy about 8.56% for all test sets and ICA algorithm accuracy about 50.51%. The best results are obtained with subimage sizes of  $45 \times 45$  pixels over the  $35 \times 35$  size. Using ICA algorithm for feature extraction without using a preprocessing module of PCA degraded fuzzy classifier performance. ICA takes advantage of the reduction of dimensionality and noise to produce more accurate and robust results. Parameter values play a vital role in the system's performance and their selection should be investigated to improve system's robustness. Other membership functions can be modeled based on mean and standard deviation of selected feature values.



## ACKNOWLEDGMENTS

Partial support of this work was provided by the National Science Foundation Grant (MRI-0215356) and by Western Michigan University FRACASF Award (WMU: 2005–2007). The authors would like also to acknowledge Western Michigan University for its support and contributions to the Information Technology and Image Analysis (ITIA) Center.

## REFERENCES

- [1] National Cancer Institute, “U.S. National Institute of Health,” 2008, <http://www.cancer.gov/cancertopics/types/breast/>.
- [2] C. H. Lee, “Screening mammography: proven benefit, continued controversy,” *Radiologic Clinics of North America*, vol. 40, no. 3, pp. 395–407, 2002.
- [3] C. Smart, R. Hendrick, J. Rutledge, and R. Smith, “Benefit of mammography screening in women ages 40 to 49 years: current evidence from randomized controlled trials,” *Cancer*, vol. 75, no. 7, pp. 1619–1626, 1995.
- [4] M. L. Giger, N. Karssemeijer, and S. G. Armato III, “Computer-aided diagnosis in medical imaging,” *IEEE Transactions on Medical Imaging*, vol. 20, no. 12, pp. 1205–1208, 2001.
- [5] R.-F. Chang, W.-J. Wu, C.-C. Tseng, D.-R. Chen, and W. K. Moon, “3-D snake for US in margin evaluation for malignant breast tumor excision using mammatome,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 7, no. 3, pp. 197–201, 2003.
- [6] H. Kobatake, M. Murakami, H. Takeo, and S. Nawano, “Computerized detection of malignant tumors on digital mammograms,” *IEEE Transactions on Medical Imaging*, vol. 18, no. 5, pp. 369–378, 1999.
- [7] L. Bocchi, G. Coppini, J. Nori, and G. Valli, “Detection of single and clustered microcalcifications in mammograms using fractals models and neural networks,” *Medical Engineering & Physics*, vol. 26, no. 4, pp. 303–312, 2004.
- [8] H. D. Li, M. Kallergi, L. P. Clarke, V. K. Jain, and R. A. Clark, “Markov random field for tumor detection in digital mammography,” *IEEE Transactions on Medical Imaging*, vol. 14, no. 3, pp. 565–576, 1995.
- [9] L. M. Bruce and R. R. Adhami, “Classifying mammographic mass shapes using the wavelet transform modulus-maxima method,” *IEEE Transactions on Medical Imaging*, vol. 18, no. 12, pp. 1170–1177, 1999.
- [10] C. A. Peña-Reyes and M. Sipper, “A fuzzy-genetic approach to breast cancer diagnosis,” *Artificial Intelligence in Medicine*, vol. 17, no. 2, pp. 131–155, 1999.
- [11] L. Zheng and A. K. Chan, “An artificial intelligent algorithm for tumor detection in screening mammogram,” *IEEE Transactions on Medical Imaging*, vol. 20, no. 7, pp. 559–567, 2001.
- [12] 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.
- [13] R. Swiniarski and H. K. Lim, “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 (ICCV ’06)*, Las Vegas, Nev, USA, June 2006.
- [14] I. Abdel-Qader, L. Shen, C. Jacobs, F. Abu Amara, and S. Pashaie-Rad, “Unsupervised detection of suspicious tissue using data modeling and PCA,” *International Journal of Biomedical Imaging*, vol. 2006, Article ID 57850, 11 pages, 2006.
- [15] L. A. Zadeh, “Fuzzy logic,” *Computer*, vol. 21, no. 4, pp. 83–93, 1988.
- [16] 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.
- [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] M. Bartlett, M. Lades, and T. Sejnowski, “Independent component representation for face recognition,” in *Human Vision and Electronic Imaging III*, vol. 3299 of *Proceedings of SPIE*, pp. 528–539, San Jose, Calif, USA, January 1998.
- [19] S. Amari, A. Cichocki, and H. Yang, “A new learning algorithm for blind source separation,” in *Advances in Neural Information Processing*, vol. 8, pp. 757–763, MIT Press, Cambridge, Mass, USA, 1996.
- [20] R. Martín-Clemente and J. I. Acha, “New equations and iterative algorithm for blind separation of sources,” *Signal Processing*, vol. 82, no. 6, pp. 861–873, 2002.



**Hindawi**

Submit your manuscripts at  
<http://www.hindawi.com>

