SUBCLASS FUZZY-SVM CLASSIFIER AS AN EFFICIENT METHOD TO ENHANCE THE MASS DETECTION IN MAMMOGRAMS

F. MOAYEDI, R. BOOSTANI, A. R. KAZEMI, S. KATEBI AND E. DASHTI

Abstract. This paper is concerned with the development of a novel classifier for automatic mass detection of mammograms, based on contourlet feature extraction in conjunction with statistical and fuzzy classifiers. In this method, mammograms are segmented into regions of interest (ROI) in order to extract features including geometrical and contourlet coefficients. The extracted features benefit from the superiority of the contourlet method to the state of the art multi-scale techniques. A genetic algorithm is applied for feature weighting with the objective of increasing classification accuracy. Although fuzzy classifiers are interpretable, the majority are order sensitive and suffer from the lack of generalization. In this study, a kernel SVM is integrated with a nero-fuzzy rule-based classifier to form a support vector based fuzzy neural network (SVFNN). This classifier benefits from the superior classification power of SVM in high dimensional data spaces and also from the efficient human-like reasoning of fuzzy and neural networks in handling uncertainty information. We use the Mammographic Image Analysis Society (MIAS) standard data set and the features extracted of the digital mammograms are applied to the fuzzy-SVM classifiers to assess the performance. Our experiments resulted in 95.6%,91.52%,89.02%, 85.31% classification accuracy for the subclass FSVM, SVFNN, fuzzy rule based and kernel SVM classifiers respectively and we conclude that the subclass fuzzy-SVM is superior to the other classifiers.

1. Introduction

In the last two decades, breast cancer has been the second leading cause of cancer deaths in women, following lung cancer [20] and early detection of this disease is the key to improving prognosis. Mammography is a reliable method for early detection of breast carcinomas; also, the process is easy and has a few side effects. However, because of the limitations of the human visual system, it is difficult for radiologists to provide both accurate and uniform evaluation for the enormous number of mammograms generated in widespread screening. There has been a great body of research in the subject of classification of mass and non-mass mammograms [7, 19, 2]. Most of these studies consist of the three parts: image segmentation and enhancement, feature extraction (applying linear or nonlinear transforms

Received: September 2008; Revised: July 2009; Accepted: November 2009

Key words and phrases: Mammography, Support vector based fuzzy neural network, Fuzzy support vector machine, Contourlet.

on the images), and classification. Here the techniques used in feature extraction and classification are modified with the aim of analyzing the digitized mammograms by applying computer image processing techniques to enhance x-ray images and then extracting salient features from suspicious regions, thus characterizing the underlying texture of the breast regions. Finally, several powerful classifiers are applied to the feature vectors extracted for classifying different regions of interest and to test if they are masses or non-masses. Mass detection requires a preprocessing step that segments the input image into different areas, such as the breast region, background, and redundant labels and text. Many techniques have already been devoted to mammograms segmentation; e,g, point operations, adaptive thresholding, gradient-based approaches, region growing, polynomial modeling, active contours, and classifier-based techniques [24, 11]. The second step is feature extraction. Texture is a commonly used feature in the analysis and interpretation of images. Malagelda [11] distinguishes textures employed in mammography based on three main extraction methods; statistical methods, model-based methods, and signal processing methods. The features of statistical approach include those obtained from co-occurrence matrices [11], from surface variation measurements (smoothness, coarseness and regularity) [11], and run-length statistics [18]. The analysis of texture in model-based methods is based on prior models such as Markov random fields [19], auto-regressive models and fractals [11]. The features extracted by signal processing techniques are obtained as either pixel characteristics or image frequency spectra including Laws energy filtering [11], Gabor filtering [11], wavelets [22, 21]. The performance of the process is very much dependent on the capability of the classifier. Some classifiers that have been applied for mass classification include a statistical Bayesian model [7], a three layered multi-layer perceptron (MLP) [19, 2], Adaptive Nero Fuzzy Inference System (ANFIS) classifier [16], Gaussian kernel based Radial Basis Function network [2], K-Nearest Neighbors (KNN) algorithm, support vector machine (SVM) [20, 9] and Decision Tree classifiers [8]. A detection scheme using successive enhancement learning SVM was proposed in [6]. In this paper SVFNN, FSVM and fuzzy rule based methods are applied to the features extracted from mammograms.

Our objective is to develop an automated imaging system for mass detection of digital mammograms. The aim is to design a subclass fuzzy SVM classifier and exploit the superiority of contourlets [14] in representing line singularities over wavelets to achieve a more efficient mammogram mass detection. The rest of this paper is organized as follows. In section 2, the methodology, including preprocessing, segmentation of region of interests, extracting statistical, geometrical and contourlet coefficients, feature weighting by genetic algorithm along with the description of SVFNN, FSVM, and fuzzy rule-based classifier is explained. Section 3 contains results and a discussion and finally, section 4 concludes the paper.

2. **Methodology**

The proposed system consists of three stages:

(1) Segmentation

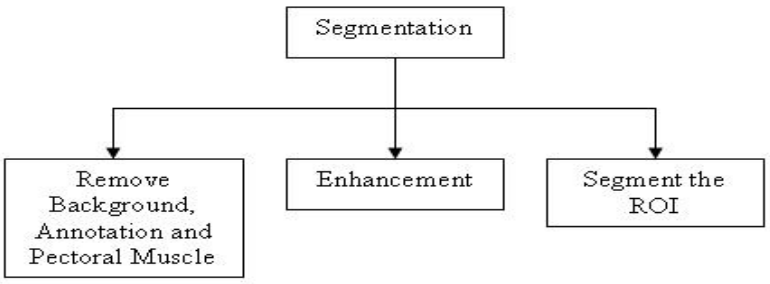


FIGURE 1. The segmentation stage.

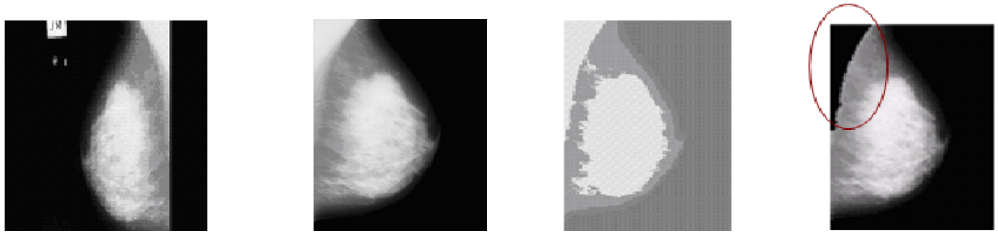
(2) Feature extraction and selection
(3) Classification process

2.1. **Segmentation.** Mammograms are difficult to interpret and a preprocessing phase of image is necessary to improve the quality of images and make the feature extraction phase more reliable. Figure 1 diagrams the segmentation phase.

2.1.1. **Remove Pectoral Muscle.** In an MIAS database, most of images consist of a black background with significant noise. Cropping, removes the part of image not necessary for analysis. The cropping operation is applied automatically by sweeping through the image and cutting horizontally and vertically parts that have mean and variance less than a certain threshold. Breast segmentation from the background is performed based on local statistical properties such as variance of the pixel gray level intensity. The gray level of the pectoral muscle is similar to that of the mass; we remove this muscle to avoid an increase in false detection rate at a later stage. This operation is carried out in the MLO[1] view, where the pectoral muscle in the mammogram is slightly brighter compared with the rest of the breast tissue. Segmentation of the breast, using the logarithm of the pixel energy and a region growing algorithm based on vertical and horizontal gradient and morphological operation, are employed in this research. Sometimes the location of mass is close to the pectoral muscle; hence, to promote mass detection and avoid mass loss, a non-rigid threshold is selected in this stage. This processes is illustrated in Figure 2.

2.1.2. **Contrast Enhancement.** Many medical images such as mammograms look blurred and fuzzy, and suffer from low contrast. Therefore, contrast enhancement is necessary before any further processing or analysis. A new intensity is assigned to each pixel according to an adaptive transfer function that is designed on the basis of local minimum(min)/maximum(max)/average(avg) intensity. This method follows the idea of adaptive histogram equalization and retinex model proposed by Yu and Bajaj [26].

[1]Media Lateral-Oblique



(a) (b) (c) (d)

FIGURE 2. First stage of preprocessing: (a) original image, (b) image cropped, background and annotation removed, (c) logarithm of pixels energy, (d) pectoral muscle removed.(for more visualization, these images have been shown in identical size.)

The local min/max/avg of a pixel are defined as the minimal, maximal and averaging intensities within a fixed size local window. Assume that $lmin$, $lmax$ and $lavg$ stand for the local min/max/avg maps, respectively, and are initialized with the image intensity values. The conditional propagation scheme from image indices $(m-1,n)$ to (m,n) is defined as follows:

$$\begin{cases} lavg_{m,n} = (1-C) * lavg_{m,n} + C * lavg_{m-1,n} \\ if(lmin_{m-1,n} < lmin_{m,n}) \\ \quad lmin_{m,n} = (1-C) * lmin_{m,n} + C * lmax_{m-1,n} \\ if(lmax_{m-1,n} > lmax_{m,n}) \\ \quad lmax_{m,n} = (1-C) * lmax_{m,n} + C * lmax_{m-1,n} \end{cases} \tag{1}$$

where C; $0 \leq C \leq 1$ is the conductivity factor. Once we obtain the local statistics (local min/max/avg) for every pixel, we need to design a transfer function from pixel to pixel. The original range at a pixel is given by the absolute difference $l_{max} - l_{min}$ between the obtained local minimum and maximum intensities at that pixel. After this mapping we obtain a broader range (denoted by w) such that $lmin$ and $lmax$ are mapped to 0 and w, respectively. Meanwhile, the original image intensity I_{old} and average image A_{old}, which satisfy $l_{min} \leq I_{old}$; $A_{old} \leq l_{max}$, are linearly stretched to their new values I_{new}, A_{new} as follows:

$$I_{new} = w * \frac{I_{old} - l_{min}}{l_{max} - l_{min}}, \tag{2}$$

and

$$A_{new} = w * \frac{A_{old} - l_{min}}{l_{max} - l_{min}} \tag{3}$$

2.1.3. **Segment Region of Interest.** We must find the region of interest for each image and restrict the process to it. The ROI are extracted to reduce the complexity of the system. Masses usually hide themselves in denser tissue with high intensity. Therefore, pixels with very low intensity should not be considered for processing. Only regions where the intensity of the pixels is close to that of the mass must be retained and processed.

Masses in mammographic images have approximately uniform textures across their interiors. Thus the variance is low, the high frequency in the mass is close to zero and by exiting from the mass the intensity will be changed.

The suspicious area is decomposed by a complete wavelet transform. The transformed images are divided into four subimages (LL, LH, HL and HH). If a mass lies in the suspicious area, the high frequency components (in LH, HL and HH) at the corresponding location must be close to zero. All locations where the high frequency components are close to zero should be registered into a binary map, which supposedly points to the mass. This region must be completely separated from the original image to be able to extract its features. On the other hand, each region in an image is considered as an input sample. Hence every region must have a label.

A refinement step is necessary in order to select only segmented regions whose abnormality likelihood is relatively high. Therefore, region labeling is performed such that those pixels that are connected to each other (which are called Connected Components (CC)) based on some morphological operations, are similarly labeled. According to the mammogram pixel resolution, we empirically observe that CCs of the 25% of the largest areas are plausible candidates for suspicion; i.e., ROI is defined as:

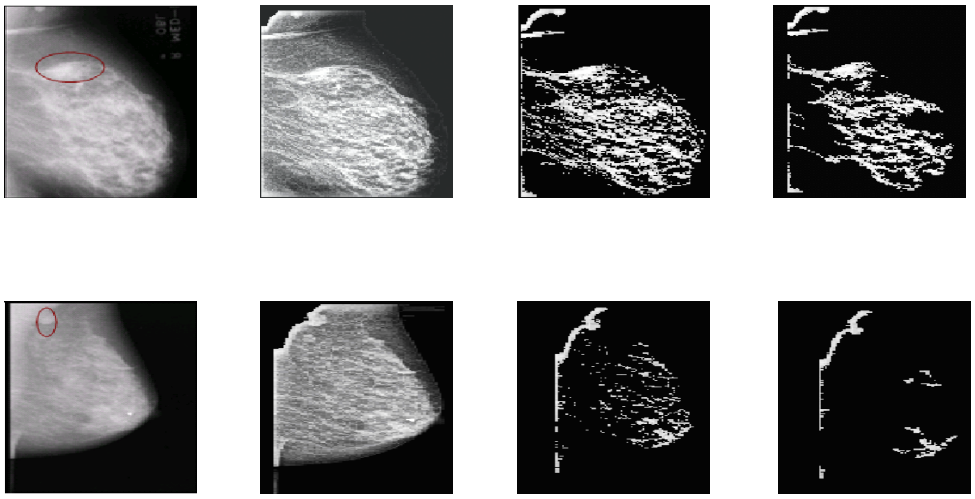
$$label(CC_i) = \begin{cases} 1 & \text{if } area_i \in 25\% \text{ of CC set, largest areas} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

Every CC_i with label 1 is assumed to be an ROI, *i.e.*, a candidate for feature extraction. These stages are illustrated in Figure 3. As is shown in second row of Figure 3, the location of mass is close to the pectoral muscle and the proposed method detects this mass simply because of closing the non-rigid threshold during the stage of removing the pectoral muscle.

3. Feature Extraction and Selection

3.1. **Feature Extraction.** Feature extractions are obtained by a technique based on contourlets and using co-occurrence matrices and geometrical process.

- *Contourlet features:* with multi-resolution analysis, an image can be represented at different levels,. Wavelet transforms have been used as a technique for 1-D piecewise smooth signal. Medical images are not as simple as 1-D piecewise smooth lines. Donoho [3] believes that wavelets perform very well for objects with point singularities in 1 and 2 dimensions. Wavelet transforms decompose an image into a series of low and high pass channels, which



(a) (b) (c) (d)

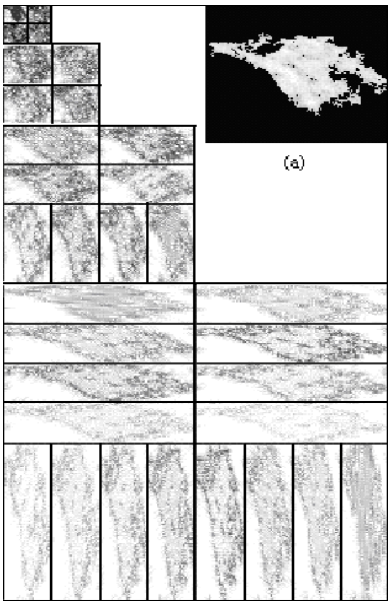
FIGURE 3. Two mammograms segmented according to the proposed approach: (a) original image (red curve shows the mass), (b) removed pectoral muscle, (c) binary map of the connected components and (d) ROIs, extracted based on selection schema given in Eq. 2.1.3.

capture only horizontal, vertical and diagonal directional details. These directions may not preserve enough directional information in medical images. Ridgelet analysis is suitable for catching radial directional details in the frequency domain. This analysis is most effective for detecting linear radial structures, which are not dominant in medical images [23].

A recent extension of ridgelet analysis introduced by Donoho and Candes [3], is the curvelet transform. It is proved that curvelets are more effective at detecting image activity along curves than radial directions [17]. Curvelets also catch structural information along multi-scales, locations, and orientations, with aspect ratio. Multi-scale decompositions capture point discontinuities and directional decomposition that link point discontinuities into linear structures.

Curvelets can represent a smooth contour with fewer coefficients compared with wavelets. Wavelet basis functions are isotropic, thus they cannot adapt to geometrical structures. Curvelet transforms have the multi-scale and time frequency localization properties of wavelets, but offer a high degree of directionality (i.e. are locally adaptive) and anisotropy [3].

This transform is indexed by three parameters: a scale parameter $a, 0 < a < 1$; an orientation parameter $\theta, \theta \in [-\pi/2, \pi/2)$, and a location parameter $b, b \in \mathbf{R}^2$. At scale a, the family of curvelets is generated by translation and rotation of a basic element φ_a:



(b)

FIGURE 4. Contourlet transform of an ROI: (a) ROI sample, (b) Four-level pyramidal decomposition. Dark gray value shows large valued coefficients.

$$\varphi_{a,b,\theta}(x) = \varphi_a(R_\theta(x-b)) \tag{5}$$

where R_θ is a rotation by θ radians and φ_a is a type of directional wavelet with spatial width a, spatial length $\sqrt{a}$ and minor axis pointing towards the horizontal direction:

$$\varphi_a(x) \approx \varphi(D_a x), \quad D_a = \begin{bmatrix} 1/a & 0 \\ 0 & 1/\sqrt{a} \end{bmatrix} \tag{6}$$

where D_a is a parabolic scaling matrix.
Do and Vetterli [5] construct a discrete domain multi resolution and multi direction expansion using non-separable filter banks. This construction is called the contourlet transform. The features extracted are a combination of the maximum, mean, standard deviation, energy, entropy, skewness from four level contourlet coefficients. An example of a contourlet transform for one ROI is shown in Figure 4. The image was decomposed into four pyramidal levels, which were then further decomposed into four, eight and sixteen directional sub-bands.

- *Co-occurrence matrix features:* Co-occurrence matrices are essentially 2-D histograms of the occurrence of grey-level pairs for a given displacement

vector. Formally, the co-occurrence P of two intensity values can be specified as a matrix of relative gray levels i and j. Co-occurrence matrices are not generally used as features; rather, various statistical features are derived from the matrix entities $P_{i,j}(d, \theta)$, in which two pixels are separated by a predefined distance d and an angle θ [11]. We have used four common directions: 0°, 45°, 90° and 135° and two distances: $d = 1$ and $d = 2$ pixels. For each co-occurrence matrix we determine energy, correlation, inertia, entropy, inverse difference moment, sum average, sum variance, sum entropy, difference average, difference variance, difference entropy, and an information measure of correlation.

- *Geometrical features:* These include area, center of mass, orientation

3.2. **Feature Selection.** Because of the high dimentionality of input features, a genetic algorithm(GA) is deployed to reduce the redundancy and improve classification accuracy. So the features are expressed in terms of choromosomes to be manipulated by genetic operators. We generate an initial pool of candidate feature subsets on which we first apply genetic operators such as crossover and mutation and then evaluate their goodness using a fitness function. The feature subsets are further processed by a selection strategy to choose individuals with highest fitness. A stop criterion is examined after each iteration. The fitness of each chromosome is determined by the inverse value of the classification error rate. Thus features are weighted by the genetic algorithm according to their discriminating information. Finally, features with weights lower than a certain threshold are eliminated.

4. Classifiers

The extracted ROIs are labeled as either containing a mass or a normal tissue. Each ROI, denoted by $\underline{x}_i$, is assumed as belonging to the "mass present" class ($y_i = +1$) or the "mass absent" class ($y_i = -1$). The labeling y_i is done based on Euclidean distance between the mass center and that of each ROI, *i.e.*, $y_i = 1$ if this distance is less than the mass radius[2], $y_i = 0$ otherwise.

The neural network is a popular generation of information processing systems that demonstrate the ability to learn from training data. Many researchers [16, 6] have studied fuzzy neural networks (FNN), which combine the capability of fuzzy reasoning in handling uncertain information with the capability of neural networks in learning from processes.

Such learning algorithms cannot usually minimize the empirical risk (training error) and expected risk (testing error) simultaneously, and thus a good classification performance in the testing phase is not achievable. The SVM performs structural risk minimization and creates a classifier with minimum Vapnik Chervonenkis (VC) dimension. combination of SVM and fuzzy neural network has been proposed in [10].

[2]This parameter is available in MIAS dataset.

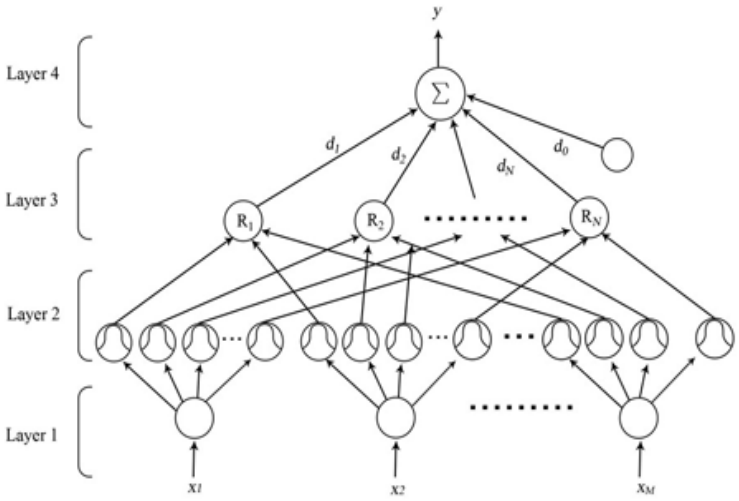


FIGURE 5. Structure of a four-layered fuzzy neural network.

4.1. **Phases of Support-vector-based Fuzzy Neural Network Implementation.** The proposed learning algorithm of SVFNN consists of three phases [10]. **Learning Phase 1:** In the first phase, the initial fuzzy rule (cluster) and membership of network structure are automatically established based on the fuzzy clustering method. A four layer FNN is shown in Figure 5.

- Layer 1: Input layer
- Layer 2: Membership function layer. With the use of Gaussian membership function.
- Layer 3: Rule Layer

$$O^{(3)} = \Pi_{i=1}^{M} a_i^{(3)} = e^{-[D_j(x-m_i)]^T[D_i(x-m_i)]} \tag{7}$$

where

$$D_j = diag[1/\sigma_{1j}, ..., 1/\sigma_{Mj}],\ m_i = [m_{1j}, m_{2j}, ..., m_{Mj}]^T \tag{8}$$

and $\underline{x} = [x_1, x_2, ..., x_M]^T$ is the FNN input vector. The output of a layer-3 node represents the firing strength of the corresponding fuzzy rule.

- Layer 4: output layer

$$O^{(4)} = \sum_{j=1}^{N} d_j * a_j^{(4)} + d_0 \tag{9}$$

where the connecting weight d_j is the output action strength of the layer-4 associated with the layer-3 rule and the scalar d_0 is a bias.

Learning Phase 2: The partitioning feature space determines the initial fuzzy rules used to determine the fuzzy kernels. In the second phase, the means of

membership functions and the connecting weights between layers 3 and 4 of SVFNN are optimized using the result of SVM learning with fuzzy kernels.(see Figure 5)

$$K(\underline{x},\underline{z}) = \begin{cases} \Pi_{i=1}^{M} u_j(x_i).u_j(z_i) & \text{if } \underline{x} \text{ and } \underline{z} \text{ are both in the } j\text{th cluster} \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

where $\underline{x} = [x_1, x_2, ..., x_M] \in R^M$ and $\underline{z} = [z_1, z_2, ..., z_M] \in R^M$ are any two training samples, and $u_j(x_i)$ is the membership function of the jth cluster. The dual quadratic optimization of SVM is solved to obtain an optimal hyper-plane in the dimension space which optimizes the following relation:

$$L(\underline{\alpha}) = \sum_{i=1}^{\nu} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{\nu} y_i y_j \alpha_i \alpha_j K(x_i, x_j), \quad 0 \le \alpha_i \le C \tag{11}$$

and

$$\sum_{i=1}^{\nu} y_i \alpha_i = 0 \tag{12}$$

where $K(x_i, x_j)$ is the kernel of SVM, ν is the number of total samples, and C is a user-specified positive parameter to control tradeoff between the SVM complexity and the number of non-separable points. This quadratic optimization problem is solved and a solution to $\underline{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_{nsv})$ is obtained, where each α_i is a Lagrange coefficient, and nsv denotes number of support vectors.

Learning Phase 3: Unnecessary fuzzy rules are recognized and eliminated by the least squares method. A set of independent fuzzy basis functions that minimize the residual error in a least squares sense is selected.

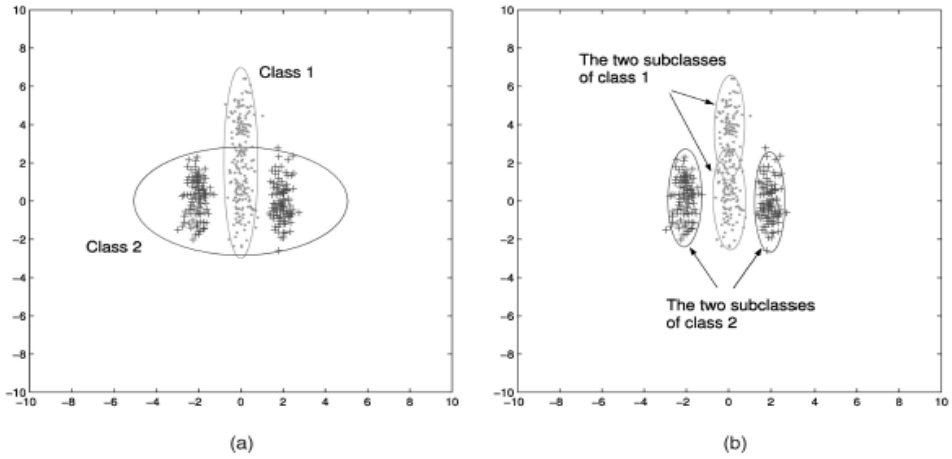
In this classifier, the leave-one-out validation procedure is applied to improve reliability.

4.2. **Subclass Fuzzy Support Vector Machine.** Xiaodan *et al.* [25] have proposed a fuzzy SVM (FSVM) based on a one versus one(1-vs-1) SVM. This procedure uses a fuzzification procedure on scores of component SVMs using a fuzzy reasoning instead of majority voting used in standard 1-vs-1 SVM to improve the voting and so resolve non-classifiable regions between hyper plans and improve classification rate. Here, for a C class problem, $C * (C-1)/2$ binary SVMs are trained. To classify an input data $\underline{x}$, suppose $g_{ij}(\underline{x})$ be the distance of $\underline{x}$ to the $Plain(i,j)$ (the hyper plain between class i, j). In standard SVM, the distance (SVM score) is compared to a threshold to find the class label

$$f_{ij}(\underline{x}) = sgn(g_{ij}(\underline{x}) - threshold) \tag{13}$$

If $f_{ij}(\underline{x}) = 1$ then $\underline{x}$ belongs to class i and belongs to class j otherwise. FSVM does not use the $sgn(.)$ function at the output of SVM(instead, $G_{ij}(\underline{x})$, finds its membership degree in the membership function $MF(\underline{x})$. Applying the input data $\underline{x}$ to all SVMs we find the matrix of scores G:

$$G_{ij} = -G_{ji} = g_{ij}(\underline{x}) \tag{14}$$



(b)

FIGURE 6. (a) Shown here are the two example of original Gaussian distributions as given by the mean and covariance matrix of the samples in each of the two classes. (b) Each class is now described using two Gaussian distributions. These have been estimated using the subclass approach.

Now the membership matrix $M = MF(G)$ is determined. To select the label of $\underline{x}$, first we find the minimum of each column in M, to form a row vector, then set the index of the maximum of this vector as the label of $\underline{x}$.

$$MF(x) = \begin{cases} 1 & x \geq 0 \\ (1+x)/2 & -1 < x < 0 \\ 0 & x \leq -1 \end{cases} \tag{15}$$

This is in fact similar to fuzzy reasoning in which a fuzzy set is used on each variable and then the min, max operators are used as "and and "or, thus resulting in a min-max reasoning. Here we have two unbalanced classes, large number of normal data and some abnormal data. In the cases where the features of classes are scattered in different locations of feature spaced (multi modal) data, subclass techniques, which consider each class of data as a virtual class, have been developed to increase classification performance [27].(see Figure 6)

We propose a new classifier based on subclass techniques and fuzzy-SVM. To apply the subclass FSVM on this problem the normal data set is first clustered into multiple clusters using k-means clustering. Each cluster is now considered to be a subclass of normal data with a virtual labeland the capability of FSVM on multiclass problems is used to classify, train and test samples to their virtual class. After all the labels that belong to subclasses of normal class are found to be normal, the accuracy and confusion matrix is calculated based on leave-one-out cross validation. The number of subclasses of normal data can be determined using the gap statistic algorithm. The number of subclasses for each class is 3.

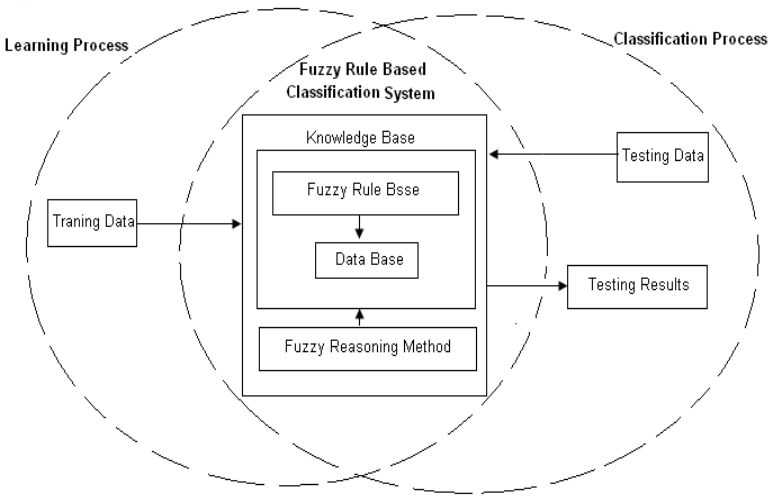


FIGURE 7. Fuzzy rule based classification system from training to testing

4.3. **Fuzzy Rule Based Classification System (FRBCS).** Many types of fuzzy classifer systems have been applied in different classification problems [16, 12, 13, 1]. In a supervised inductive learning process, a pattern classification system is developed by means of a set of classified examples used to establish a correspondence between the new pattern descriptions and the class labels. When a classification system is composed based on a set of fuzzy rules, it is called Fuzzy Rule-based Classification System (FRBCS). An FRBCS consists of two main components: 1) a Knowledge Base (KB), i.e., a Rule Base(RB) and Data Base(DB) for specific classification problem, and 2) a Fuzzy Reasoning Method(FRM), that classifies a new pattern using the information in the KB.

The most commonly used fuzzy inference method, maximum matching, classifies an example using the rule consequent with the highest association degree [4]. The block diagram of a FRBCS is depicted in Figure 7. As an example, a fuzzy classifier in which each rule represents a subclassifier for a c-class pattern classification problem with n attributes, can be written as:

$$\text{Rule } R_j : \text{If } x_1 \text{ is } A_{j1} \text{ and } ... \text{ and } x_n \text{ is } A_{jn} \text{ then Class } C_j \ , \ j = 1, 2, ..., N$$

where $\underline{x} = (x_1, ..., x_n)$ is an n-dimensional pattern vector, A_{ji} is an antecedent linguistic value, C_j is a consequent class, and N is the number of fuzzy rules. A grid-type fuzzy partitioning is used for generating fuzzy rules (each rule acts as a classifier). The triangular membership function is used for its simplicity and interpretiblity. For rule weighting and unnecessary rule pruning, the confidence estimate is computed. In the classification phase, a fuzzy reasoning method is applied based on the single winner rule [4].

5. **Result and Discussion**

The method was applied to a set of 60 (30 normal and 30 abnormal cases) MLO mammograms taken from the MIAS data set.

The extracted ROI's were marked as either containing mass or normal tissues of the breast. This label was computed based on Euclidean distance between mass center from annotations of database and ROI center with radius center. Each ROI, denoted by x_i , was then treated as an input pattern for the mass present class ($y_i = +1$). Mass absent samples were collected ($y_i = -1$) similarly.

The mammograms were digitized in 8 bits gray resolution and the size of each image was $1024 * 1024$ pixels. First, the images were segmented and the pectoral muscle removed with 0.97% accuracy. Then, the region of interest was extracted from each of these enhanced images. In the next stage, features of interest were extracted from each suspicious region. The features used in this research contain a combination of the following ten statistics: maximum, mean, standard deviation, energy, entropy, skewness in contourlet domain, co-occurrence matrix, area, position, and orientation in spatial domain. The dimensions of feature vectors weare reduced by a genetic algorithm where the fitness function was the inverse value of classification error. Finally the SVFNN, FSVM, kernel SVM and FRBCS classifiers were applied to the extracted features and the leave-one-out cross validation process was performed in the training phase to optimize the classifier parameters. Parameters of SVM such as C and gamma were calculated adaptively; 2^{11} for C and 0.0000001 for gamma were reported. To evaluate the performance of each classifier, specificity, sensitivity parameters, and accuracy rates were then calculated from each of the misclassification matrices.

Ideally, one wishes to achieve high specificity and sensitivity. Theoretically, however, these two measures are inversely proportional. Since accuracy is a function of sensitivity and specificity measures together, this descriptor was selected to determine the overall preciseness of the classifier. Receiver Operating Characteristics (ROC) curves were used to evaluate the class-discriminating capability of different classifiers. The results, ROC curve, accuracy and standard deviation of assessed classifiers in comparared with different features and classifiers are shown in Tables 1, 2, 3 and Figure 8, 9. It can be seen that the results of FSVM and SVFNN, which combine the positive points of SVM and fuzzy reasoning, are superior to the kernel SVM and FRBS. This shows that even a very well-known classifier like SVM just has one positive point (good generalization) and the FRBCS is just equipped with human like reasoning, and hence neither is a perfect classifier. Therefore, it is time to stop arguing about which classifier is the best and time to integrate the positive points of classifiers in an efficient manner and build up a new combinational structure based on them. Although FRBCS and SVM are efficient classifiers, our experiment show that the combinational construction based on them is more efficient. In order to assess the significance of the results, p-value and t-value are determined from t-test as 0.01 and 14.56 respectively.

Kernel(RBF) SVM			FRBS			SVFNN			subclass FSVM	
Number of Support Vector	Train Error Rate	Test Error Rate	Number of fuzzy Rules	Train Error Rate	Test Error Rate	Number of fuzzy Rules	Train Error Rate	Test Error Rate	Train Error Rate	Test Error Rate
19	12.52± 0.1	14.69± 0.1	26	7.75± 0.1	10.98± 0.1	28	7.92± 0.07	8.48± 0.12	4.2± 0.0001	4.4± 0.0404

TABLE 1. Experimental result and comparison between kernel SVM, FRBS, SVFNN and subclass FSVM.

Measures	SVFNN				subclass FSVM			
	co-occurence	wavelets	contourlets	contourlets & co-occurence & morphology	co-occurence	wavelets	contourlets	contourlets & co-occurence & morphology
Mean Sensitivity	0.600	0.750	0.812	0.867	0.880	0.900	0.900	**0.938**
Mean Specificity	0.950	0.960	0.960	0.967	0.958	0.953	0.967	**0.98**
Mean Accuracy	0.800 ±0.15	0.881 ±0.12	0.908 ±0.1	0.915 ±0.12	0.921 ±0.05	0.940 ±0.04	0.945 ±0.04	**0.956** ±**0.04**

TABLE 2. Performance comparison of SVFNN and subclass FSVM classifiers using different features in detecting normal and abnormal tissues.

Classifier	Mean Accuracy	Mean Sensitivity	Mean Specificity
RBF kernel based SVM	85.31 ± 0.10	55	91.3
Fuzzy Rule based	89.02 ± 0.10	60	95
SVFNN	91.52 ± 0.12	86.7	96.7
subclass FSVM	95.60 ± 0.04	93.8	98

TABLE 3. Sensitivity, Specificity and test Accuracy for different classifiers.

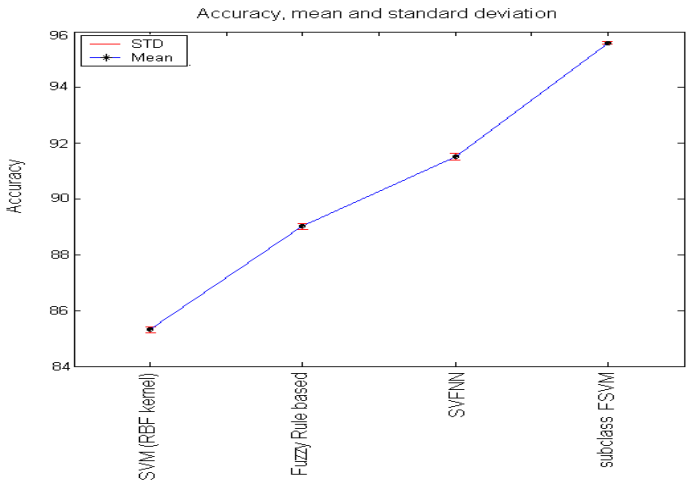


FIGURE 8. Accuracy and standard deviation in different classifiers

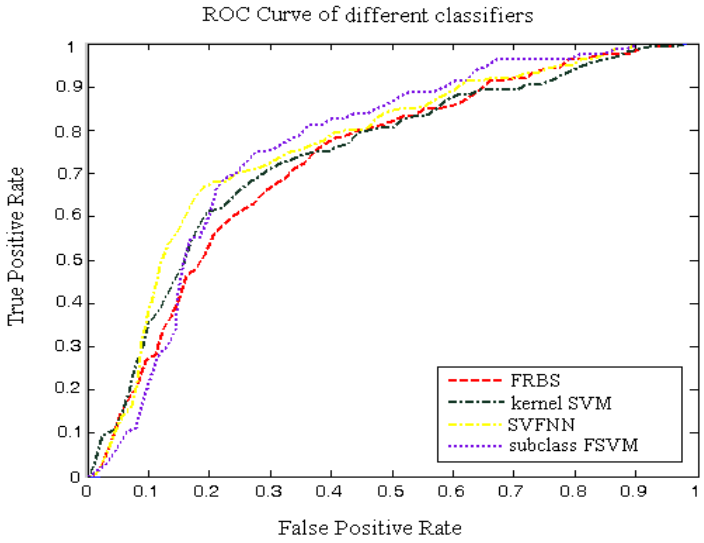


FIGURE 9. ROC curve for subclass FSVM, SVFNN, FRBS and kernel SVM classifiers. It can be seen the area under the ROC curve for subclass FSVM is the best.

6. **Conclusion**

In this paper, we have exploited the advantages of contourlet-based texture analysis along with geometrical and statistical features and employed SVFNN, FSVM

and fuzzy rule based classifiers to detect and classify breast masses into benign and malignant cases. It is shown that the contourlet transform can successfully capture structural information along multiple scales, locations, and orientations.

SVFNN combines the superior classification power of SVM in high-dimensional data spaces and the efficient human-like reasoning of FNN in handling uncertainty information. The SVM has a good generalization ability and can simultaneously minimize the empirical and expected risks for pattern classification problems. The use of the proposed fuzzy kernels provides the SVM with adaptive local representation power, and thus brings the advantages of FNN (such as adaptive learning and economic network structure) into the SVM directly. Fuzzy kernels using the variable-width fuzzy membership functions can make the SVM more efficient in terms of the number of required support vectors, and also make the learned FNN more understandable to humans. FSVM uses a fuzzification procedure on scores of component SVMs to improve voting using a fuzzy reasoning (instead of majority voting used in standard SVM) to resolve non-classifiable regions between hyperplans and improve classification rate. We conclude that FSVM has a strong capability to classify the mass and non mass images and may be considered as a suitable alternative for this application. Digital mammogram images using contourlet coefficients, statistical and geometrical features along with subclass fuzzy-SVM classifier are also introduced.

References

[1] R. A. Aliev, B. G. Guirimov and R. R. Aliev, *A neuro-fuzzy object classifier with modified distance measure estimator*, Iranian Journal of Fuzzy Systems, **1(1)** (2004), 5-15.

[2] K. Bovis, S. Singh, J. Fieldsend and C. Pinder, *Identification of masses in digital mammograms with MLP and RBF nets*, IEEE Trans. on Image Processing, **1** (2005), 342-347.

[3] E. J. Candes and D. L. Donoho, *Curvelets: a surprisingly effective non adaptive representation for objects with edges*, Saint-Malo Proceedings, Nashville, TN: Vanderbilt Univ, 2000.

[4] O. Cordon and M. J. del Jesus and F. Herrera, *Genetic learning of fuzzy rule based classification systems cooperating with fuzzy reasoning methods*, Technical Report, DECSAI-970130, 1997.

[5] M. N. Do and M. Vetterli, *The contourlet transform: an efficient directional multi-resolution image representation*, IEEE Trans. on Image Processing, **14(12)** (2005), 2091-2106.

[6] I. El-Naqa, Y. Yang, M. Wernick, N. Galatsanos and R. Nishikawa, *A support vector machine approach for detection of microcalcifications*, IEEE Trans. on Medical Imaging, **21(12)** (2002), 1552-1563.

[7] E. A. Fischer, J. Y. Lo and M. K. Markey, *Bayesian networks of BI-RADS descriptors for breast lesion classification*, IEEE EMBS, San Francisco, **4** (2004), 3031-3034.

[8] O. J. Freixenet, A. Bosch, D. Raba and R. Zwiggelaar, *Automatic classification of breast tissue*, Lecture Notes in Computer Science, Pattern Recognition and Image Analysis, (2000), 431-438.

[9] W. H. Land, J. L. Wong Daniel, W. McKee, T. Masters and F. R Anderson, *Breast cancer computer aided diagnosis (CAD) using a recently developed SVM/GRNN oracle hybrid*, IEEE International Conference on Systems, Man and Cybernetics, 2003.

[10] C. T. Lin, C. M. Yeh, S. F. Liang, J. F. Chung and N. Kumar, *Support-vector-based fuzzy neural network for pattern classification*, IEEE Trans. on Fuzzy Systems, **14(1)** (2006), 31-41.

[11] A. O. Malagelada, *Automatic mass segmentation in mammographic images*, PhD Thesis, Universitat de Girona, Spain, 2004.

[12] E. G. Mansoori, M. J. Zolghadri and S. D. Katebi, *Using distribution of data to enhance prformance of fuzzy classification systems*, Iranian Journal of Fuzzy Systems, **4(1)** (2007), 21-36.

[13] E. G. Mansoori, M. J. Zolghadri, S. D. Katebi, H. Mohabatkar, R. Boostani and M. H. Sadreddini, *Generating fuzzy for protein classification*, Iranian Journal of Fuzzy Systems, **5(2)** (2008), 21-33.

[14] F. Moayedi, Z. Azimifar, R. Boostani and S. Katebi, *Contourlet based mammography mass classification*, Lecture Notes in Computer Science, Image Analysis and Recognition, **4633** (2007), 923-934.

[15] F. Moayedi, R. Boostani, Z. Azimifar and S. Katebi, *A support vector based fuzzy neural network approach for mass classification in mammography*, International Conference on Digital Signal Processing, Britain, 2007.

[16] R. Mousa, Q. Munib and A. Mousa, *Breast cancer diagnosis system based on wavelet analysis and fuzzy-neural netwrok*, IEEE Trans. on Image Processing, **28(4)** (2005), 713-723.

[17] D. Y. Po and N. Do, *Directional multiscale modeling of images using the contourlet transform*, IEEE Trans. on Image Processing, (2006), 1-11.

[18] D. Raba, A. Oliver, J. Marti, M. Peracaula and J. Espunya, *Breast segmentation with pectoral muscle suppression on digital mammograms*, Springer-Verlag: Medical Imaging: Pattern Recognition and Image Analysis, **3523** (2005), 471-478.

[19] M. Roffilli, *Advanced machine learning techniques for digital mammography*, Technical Report, Department of Computer Science University of Bologna, Italy, 2006.

[20] M. S. B. Sehgal, I. Gondal and L. Dooley, *Support vector machine and generalized regression neural network based classification fusion models for cancer diagnosis*, proceedings in Fourth IEEE International Conference on Hybrid Intelligent System, Computer Society, 2004.

[21] L. Semler and L. Dettori, *A comparison of wavelet-based and ridgelet-based texture classification of tissues in computed tomography*, International Conference on Computer Vision Theory and Applications, 2006.

[22] L. Semler, L. Dettori and J. Furst, *Wavelet-based texture classification of tissues in computed tomography*, IEEE International Symposium on Computer-Based Medical Systems, 2005.

[23] J. L. Starck, E. J. Candes and D. L. Donoho, *The curvelet transform for image denoising*, IEEE Trans. on Image Processing, **11(6)** (2002), 670-684.

[24] C. Varelaa, P. G. Tahocesb, A. J. Mndezc, M. Soutoa and J. J. Vidala, *Computerized detection of breast masses in digitized mammograms*, Computers in Biology and Medicine, **37(2)** (2007), 214-226.

[25] W. Xiaodan and W. Chongming, *Using membership function to improve multi-class SVM classification*, ICSP Proceeding, China, 2004.

[26] Z. Yu and C. Bajaj, *A fast and adaptive for image contrast enhancement*, IEEE International Conference on Image Processing, 2004.

[27] M. Zhu and A. M. Martinez, *Subclass discriminant analysis*, IEEE Trans. on Pattern Analysis and Machine Intelligence, **28(8)** (2006), 1247-1286.

Fatemeh Moayedi*, Reza Boostani, Ali Reza Kazemi and Serajodin Katebi, Vision and Image Processing Laboratory, School of Electrical and Computer Engineering, Shiraz University, Shiraz, Iran

E-mail address: moayyedi,boostani,Kazemi,Katebi@cse.shirazu.ac.ir

Ebrahim Dashti, Board of Science, Azad Universitiy Branch of Jahrom, Iran

E-mail address: sayed.dashti@jia.ac.ir

*Corresponding author