



Characterization of clustered microcalcifications in digitized mammograms using neural networks and support vector machines

A. Papadopoulos^{a,b}, D.I. Fotiadis^{b,*}, A. Likas^b

^aDepartment of Medical Physics, Medical School, University of Ioannina, GR 45110 Ioannina, Greece

^bDepartment of Computer Science, University of Ioannina, Unit of Medical Technology and Intelligent Information Systems, and Biomedical Research Institute—FORTH, GR 45110 Ioannina, Greece

Received 12 November 2003; received in revised form 29 September 2004; accepted 11 October 2004

KEYWORDS

Support vector machine;
Microcalcification cluster classification;
Mammography

Summary

Objective : Detection and characterization of microcalcification clusters in mammograms is vital in daily clinical practice. The scope of this work is to present a novel computer-based automated method for the characterization of microcalcification clusters in digitized mammograms.

Methods and material : The proposed method has been implemented in three stages: (a) the cluster detection stage to identify clusters of microcalcifications, (b) the feature extraction stage to compute the important features of each cluster and (c) the classification stage, which provides with the final characterization. In the classification stage, a rule-based system, an artificial neural network (ANN) and a support vector machine (SVM) have been implemented and evaluated using receiver operating characteristic (ROC) analysis. The proposed method was evaluated using the Nijmegen and Mammographic Image Analysis Society (MIAS) mammographic databases. The original feature set was enhanced by the addition of four rule-based features.

Results and conclusions : In the case of Nijmegen dataset, the performance of the SVM was $A_z = 0.79$ and 0.77 for the original and enhanced feature set, respectively, while for the MIAS dataset the corresponding characterization scores were $A_z = 0.81$ and 0.80 . Utilizing neural network classification methodology, the corresponding performance for the Nijmegen dataset was $A_z = 0.70$ and 0.76 while for the MIAS dataset it was $A_z = 0.73$ and 0.78 . Although the obtained high classification performance can be successfully applied to microcalcification clusters characterization, further studies must be carried out for the clinical evaluation of the system using

* Corresponding author. Tel.: +30 26510 98803; fax: +30 26510 298889.
E-mail address: fotiadis@cs.uoi.gr (D.I. Fotiadis).

larger datasets. The use of additional features originating either from the image itself (such as cluster location and orientation) or from the patient data may further improve the diagnostic value of the system.

© 2004 Elsevier B.V. All rights reserved.

1. Introduction

Several methodologies have been developed in order to improve radiologists' efficiency in the diagnostic interpretation of mammograms. The successful development of computer aided diagnosis (CAD) systems would be of great value, if these systems can provide a reliable second opinion to the radiologist. CAD systems integrate image analysis and artificial intelligent techniques, aiming to provide accurate, objective and reproducible mammogram interpretation procedures.

The problem of mammogram interpretation using CAD systems can be decomposed into two sub-problems. The first deals with the detection and localisation of regions of interest (ROIs), which include suspicious lesions. The second, and more difficult sub-problem, is the characterization of the identified lesions as malignant or benign [1]. A successful characterisation could contribute to the reduction of unnecessary biopsies.

The most common approach for the development of CAD systems involves feature extraction procedures performed either by a computer system or manually by the radiologists [2,3]. The computed features are subsequently fed to a categorization scheme. Automatic feature extraction procedures utilize image analysis techniques for the computation of feature vectors characteristic of structures detected at the segmentation stage. Several types of feature extraction methods can be found in the literature such as morphological [4–7], texture [8–10], fractal [11,12], histogram statistics [13] and wavelets [14–17]. Morphological features are the most commonly used due to their similarity with the characteristics taken into account by the radiologists. The patient's age and a set of 13 typical morphological descriptors is used in ref. [18] resulting in a high classification performance.

Several methodologies have been proposed for the microcalcification characterization problem, such as decision trees [6,19], linear discriminant analysis [5,8], k-nearest neighbours [20–23] and artificial neural networks (ANNs) [2,3,24–30]. In general it is very difficult to compare the efficiency of the above methods since they have been tested in different mammographic datasets using different performance measures. A review of the existing detection and classification methodologies can be found in ref. [31].

In this work an automated system for the characterization of microcalcification clusters as malignant or benign is presented. The method consists of three stages: the cluster detection stage described in a previous work [32], the feature computation stage and the final classification stage. Two different classification schemes have been implemented and tested based on ANNs and support vector machines (SVMs). It must be noted that SVMs are used for the first time for the cluster characterization problem. Originally, 33 features of a 54-feature set were selected. Moreover, a new type of features is defined, called rule-based features, which are obtained from the 2D graphical representation of all pairs of features. The addition of four rule-based features resulted in an enhanced feature set that consisted of 37 features. The performance of the classifiers has been evaluated using the receiver operating characteristic (ROC) methodology [33] and the classification rate. The obtained results provide high classification performance and thus our method can be considered quite promising.

2. Methods and material

2.1. Image dataset

In this study the Nijmegen [34] and MIAS [35] mammographic image databases were used. The Nijmegen database consists of 40 images of both craniocaudal (CC) and medio-lateral oblique (MLO) views from 21 patients. The digitisation sampling aperture is 0.05 mm, the sampling distance is 0.1 mm and the size of each image is 2048 × 2048 pixels. Twelve bits are used for each pixel representation and we have rescaled the images to 8 bit depth (256 grey levels) using a noise equalization table set provided with the database. The MIAS dataset contains 20 mammograms. Each one is a MLO view and is digitized with a spatial resolution of 50 μm and 8 bit grey depth. In both datasets, the microcalcification clusters have been annotated in each image by expert radiologists using a circle enclosing the abnormality. The total number of annotated clusters in the Nijmegen dataset is 105, which corresponds to 76 malignant and 29 benign microcalcification clusters. The MIAS database includes 25 annotated clusters (12 malignant and 13 benign microcalcification clusters).

The proposed method for the characterization of the microcalcifications as malignant or benign has been implemented in three stages. Initially, a cluster detection procedure is used to identify clusters of microcalcifications. Next, important features of those clusters are computed. In the final stage the features are used as input to a classification system to provide the final diagnosis.

2.2. Cluster detection procedure

The objective of this stage is the identification of clusters of microcalcifications. The procedure has been described in an earlier work [32] and is based on a hybrid intelligent system combining rule-based and ANN methods. Figure 1 summarizes the stages of the detection procedure. Initially, a pre-processing procedure is applied in order to remove the useless radiological marks as well as the background of the image. Then, background correction along with contrast enhancement is applied to indicate potential microcalcification objects. Morphological descriptors are used to extract the region of interest (ROI). For all objects and clusters contained in every ROI we compute several discriminative morphological and textural features, which are used as input to the false positive reduction procedure. This system is a hybrid intelligent system based on a combination of a rule-based and an ANN component, and provides a

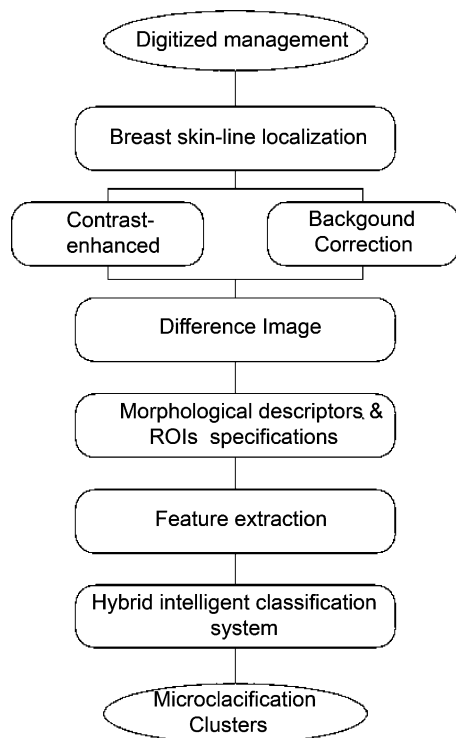


Figure 1 The microcalcification cluster detection system.

characterization of each ROI, either as a true cluster of microcalcifications or a false positive detection.

The Nijmegen database reference file reports 105 cluster areas. Applying our detection algorithm we have identified 10 additional areas because in some cases two distinct ROIs had been detected into a single database annotated circle. Two of the detected areas are false negative cases of the detection algorithm which were added to the classification system manually. From the total of 115 ROIs, 80 are malignant and 35 are benign. False positives are not included in the 115 ROIs since those are excluded after the detection phase.

The MIAS database includes 25 annotated cluster areas. The detection procedure results in the identification of 34 cluster areas which include all the annotated clusters. The additional ROIs were extracted due to the fact that two different ROIs were identified within the same annotated circle. From the 34 detected clusters, 18 are malignant and 16 are benign. All false positive findings were excluded at the end of the detection procedure.

2.3. Features computation and selection

For each detected cluster 54 features have been identified as shown in Table 1. Those features were computed either from individual microcalcifications or constitute averages of the five largest microcalcifications included in a cluster. The features given in Table 1, are cluster features which are extracted either using microcalcification features (e.g. mean value of microcalcification area, S.D. of microcalcifications area, mean microcalcification background intensity) or considering each cluster as a separate object (e.g. cluster area, cluster entropy, cluster elongation).

In order to reduce the number of features, a feature selection procedure based on ROC analysis was followed to identify the most discriminative features. The ROC curve [33] was plotted for each feature and the area A_z under the curve is computed. Initially in the case of Nijmegen database, 33 features were selected having A_z value higher than the defined threshold, as listed in Table 1. The criterion for feature elimination is that $0.50 \leq A_z \leq 0.52$. It must be noted that a high percentage of the 33 features were used as features for the detection procedure described above (22 features were utilized in the detection scheme). The use of extra features underlines the increased difficulty of the classification of the clusters as benign or malignant, compared to the cluster detection problem. In the case of MIAS dataset, the same 33 features are utilized since their A_z scores are above 0.60 and none of the rest resulted in a considerable

Table 1 Features for cluster categorization

Microcalcification cluster classification features

Area of the cluster convex hull	Mean microcalcifications intensity
Cluster area	Mean perimeter of microcalcifications in cluster
Cluster eccentricity	Minor cluster's axis (equivalent ellipse)
Cluster elongation	Neighbouring with a larger cluster
Cluster entropy	Number of microcalcifications in cluster
Clusters' equivalent diameter	Orientation of cluster
Extent of cluster	Solidity of cluster
Filled area in cluster	Spreading of microcalcifications in cluster
Major cluster's axis (equivalent ellipse)	STD of distances from cluster centroid
Mean contrast	STD of microcalcification compactness
Mean distance from cluster centroid	STD of microcalcification elongation
Mean local microcalcification background	STD of microcalcification intensity
Mean microcalcification area	STD of microcalcifications area
Mean microcalcification background intensity	STD of microcalcifications contrast in cluster
Mean microcalcification compactness	STD of microcalcifications perimeter in cluster
Mean microcalcification elongation	The length of the cluster convex hull
Mean microcalcifications eccentricity	

A_z score. Thus, for generalization and simplicity reasons the same feature set is used.

2.4. Classification methods

The aim of the classification stage is the characterization of each cluster as malignant or benign using the selected features. In this work we have employed rule-based expert systems, ANNs and SVMs.

2.4.1. Rule-based expert system

The common approach in rule-based systems involves the use of rules for applying thresholds in selected single features. However, in our case it is not possible to identify discriminative rules of this type. Thus, the following approach was adopted: Initially, for each pair of features, a 2D graphical representation of the dataset in the two-feature space was performed as shown in Figure 2. This generates more than 4002-D plots. From the visual inspection of the 2D plots, it was found that four of

those plots have the highest discriminative value. This means that in a two-feature space a straight line (defining a linear decision rule) can be drawn which defines a region (half space) containing a sufficient number of points belonging mostly (more than 90%) to the same class. Thus, the definition of the linear decision value is empirical.

The pairs of features corresponding to those plots in the case of the Nijmegen database are: (a) mean microcalcification cluster eccentricity—mean contrast, (b) mean local background—mean distance from cluster centroid, (c) mean contrast—mean compactness and (d) standard deviation of the microcalcification distances from the cluster centroid—equivalent diameter of the cluster. In the case of MIAS dataset the corresponding feature pairs are: (a) mean microcalcification cluster eccentricity—mean contrast, (b) mean local background—mean distance from cluster centroid, (c) mean contrast—number of microcalcifications in cluster and (d) mean local background—standard deviation of the microcalcification distances from the cluster centroid. Two of the four feature pairs, (a) and (b), are the same for both datasets.

Both sets of rules cannot be directly used as an independent classifier due to the resulting poor performance as presented in Table 2. However, the distance of a point from the corresponding linear boundary can be considered as an additional feature to be used by the classification system. It must be also noted that each of these distance features has a sign which indicates the half space (with respect to the line) in which the data point lies. The incorporation of the additional four features into the original selected set of features results in an enhanced feature set, consisting of 37 features.

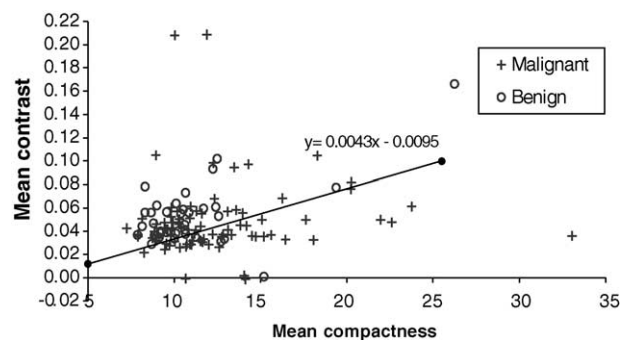


Figure 2 2D plot of mean contrast vs. mean compactness and the corresponding linear decision boundary.

Table 2 Performance of the rule-based classifier

Dataset	Pairs of features	Malignant (true characterization)	Benign (false characterization)
Nijmegen	Mean cluster eccentricity—mean contrast	28	1
	Mean local background—mean distance from cluster centroid	44	5
	Mean contrast—mean compactness	33	3
	Standard deviation of the microcalcification distances from cluster centroid—equivalent diameter of the cluster	41	4
	MIAS	Mean cluster eccentricity—mean contrast	5
MIAS	Mean local background—mean distance from cluster centroid	11	1
	Mean contrast—number of microcalcifications in cluster	9	1
	Mean local background—standard deviation of the microcalcification distances from the cluster centroid	9	1

2.4.2. Neural classifier

The selected ANN classifier is a feedforward multi-layer perceptron with sigmoid hidden nodes. The ANN architecture consists of one hidden layer with fifteen sigmoid nodes as shown in Figure 3, and an output layer with one sigmoid node, whose value indicates a malignant or a benign microcalcification cluster. Principal component analysis (PCA) has been implemented in order to reduce the size of the input feature vector. The output of the PCA is a reduced feature vector composed of seven features as shown in Figure 4, providing the best classification performance. PCA eliminates features contributing to more than 3% of the total variation of the original feature set.

Those features are normalized to zero mean and unit variance. Gradient decent, resilient backpropagation, conjugate gradient and quasi-Newton methodologies were employed for ANN training in order to select the one with the best classification

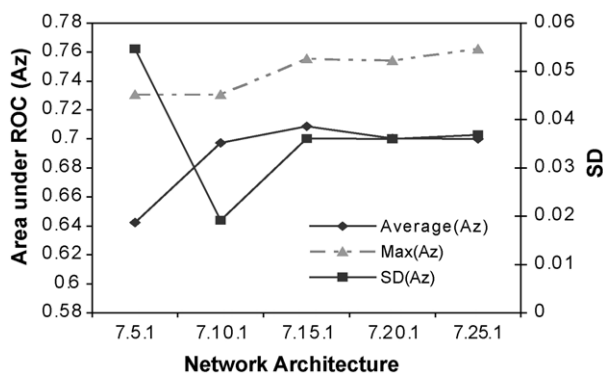


Figure 3 The performance of one-hidden layer network architectures for several numbers of hidden nodes.

ability [36]. The training procedure is terminated either when the training error is less than 10^{-5} or when 2000 iterations have been performed. The training error used is the mean square error—which is the average squared error between the network output and the target output for all the training patterns (training) or the test patterns (evaluation). Best results were obtained using the quasi-Newton one-step-secant (OSS) algorithm [37]. The two-fold cross validation method was used for the performance assessment. When the enhanced feature set is used the classification performance, using the same ANN architecture and PCA, is improved.

2.4.3. Support vector machines

Another category of classification methods that has recently received considerable attention is the use of support vector machines [38,39]. SVMs have not

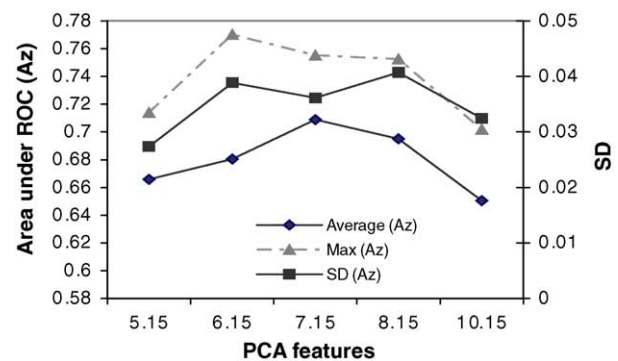


Figure 4 The performance of ANNs (one-hidden layer, fifteen nodes) for several values of inputs (PCA features) is plotted. The maximum, average and standard deviation of A_z is presented for each network.

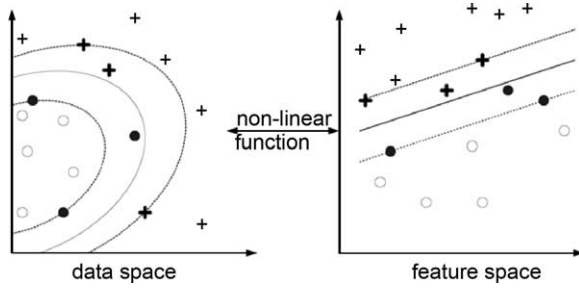


Figure 5 A non-linear SVM maps the data from the feature space D to the high dimensional feature space F by the non-linear function ϕ .

been used previously for the characterization of microcalcification clusters but only for their detection [40,41]. SVMs are based on the definition of an optimal hyperplane, which linearly separates the training data so that minimum expected risk is achieved. In contrast with other classification schemes, a SVM aims to minimize the empirical risk R_{emp} and at the same time, maximize the distances (geometric margin) of the data points from the corresponding linear decision boundary as shown graphically in Figure 5. R_{emp} is defined as

$$R_{\text{emp}}(a) = \frac{1}{2l} \sum_{i=1}^l |y_i - f(x_i, a)| \quad (1)$$

where $x_i \in R^N$, $i = 1, \dots, l$, is the training vector belonging to one of two classes, l is the number of training points, $y_i \in \{-1, 1\}$ indicates the class of x_i , and f is the decision function. The training points in the space R^N are mapped nonlinearly into a higher dimensional space F by the function (*a priori* selected) $\phi: R^N \rightarrow F$. It is in this space (feature space) where the decision hyperplane is computed. The training algorithm uses only the dot products ($\phi(x_i) \phi(x_j)$) in F . If a “kernel function” K exists, such that

$$K(x_i, x_j) = \phi(x_i) \phi(x_j) \quad (2)$$

in which only the knowledge of K is required by the training algorithm. The decision function is defined as

$$f(x) = \sum_{i=1}^l y_i a_i K(x_i, x) + b \quad (3)$$

where a_i are the weighting factors and b denotes the bias. After training, the condition $a_i > 0$ is valid for only a few examples, while for most $a_i = 0$. Thus, the final discriminant function depends only on a small subset of the training vectors which are called support vectors.

The selection of the kernel K is very important for the performance of the classifier. Several types of

kernels have been reported in the literature, such as the polynomial type of degree p ,

$$K(x_i, x) = (x_i(x+1))^p \quad (4)$$

and the Gaussian kernel

$$K(x_i, x) = e^{-\|x_i - x\|^2 / 2\sigma^2} \quad (5)$$

where σ is the kernel width. Each kernel function should fulfill *Mercer's condition* [38,42].

In this study the SVM training algorithm provided by the LIBSVM library [10] was implemented. It has been proven to be stable, computationally inexpensive and highly competitive compared with other SVM codes [43–47]. As in the ANN case, the two-fold cross validation method was employed for the performance evaluation. The number of PCA components used in the SVM classification system is seven, as in the ANN scheme, since for this number the best performance was obtained.

The Gaussian kernel and conducted experiments for several values of the standard deviation σ were selected. In order to apply the SVM training algorithm, the regularization parameter C and the termination criterion ε must also be adjusted. To perform parameterization the training algorithm was applied for the following values of the parameters: $\gamma \in \{10^{-6}, 10^{-5}, \dots, 0.01, 0.5\}$, $C \in \{1, 10, \dots, 10^5\}$ and $\varepsilon \in \{10^{-5}, \dots, 10^{-1}\}$, where $\gamma = \frac{1}{2\sigma^2}$.

3. Experimental results

The above described methodology has been evaluated on two well established mammographic datasets, the Nijmegen and the MIAS datasets. ROC analysis was employed to assess the performance of the method in both datasets. Moreover, in order to compare the results with those reported in the literature, the best classification rate (BCR) was computed, which is the ratio of the sum of true positives and true negatives over the total number of samples for a range of decision threshold values.

3.1. Nijmegen dataset

For the rule-based classifier, if a decision rule is valid, the cluster is classified as malignant, otherwise it remains unclassified. Using the expert system with the four linear decision rules for the Nijmegen dataset, the correct characterization of 44 (52%) malignant clusters and the false characterization of 2 (5.8%) benign was achieved. The decision is based on majority voting, using the characterisation provided by each rule. A cluster with two positive votes is characterised as malignant. For each rule applied independently, the obtained true and false charac-

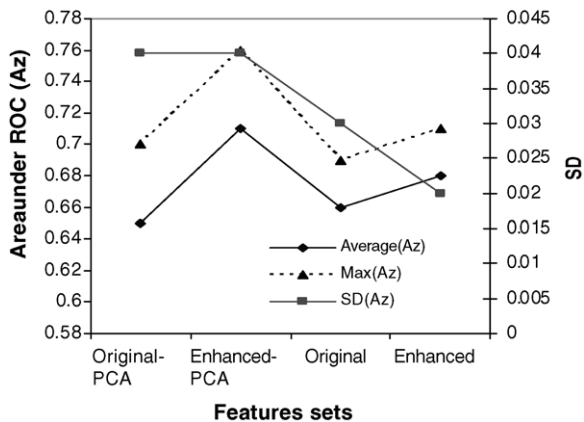


Figure 6 The performance of ANN classifier for several feature sets with and without PCA analysis (Nijmegen dataset), is plotted. The maximum, average and standard deviation of A_z is presented for each feature set.

terizations are shown in Table 2. It is evident that the performance of the rule-based system is not acceptable.

The ANN classifier has been used both for the original and the enhanced feature sets. The performance of the method was evaluated using two-fold cross validation and ROC analysis. For each training set (fold) the training algorithm was applied 10 times with different initial weight values. For the original feature set the maximum value of A_z was found $A_{zmax} = 0.70$. The mean A_z was 0.65 with standard deviation 0.04 as presented in Figure 6. For the enhanced feature set the classification performance is improved remarkably resulting in $A_{zmax} = 0.76$ with mean A_z 0.71 and standard deviation 0.04 as shown in Figure 6. The BCR values were 0.72 and 0.77 for the original and enhanced feature set. When only the four rule-based features constitute the input vector, the characterization is worse resulting $A_{zmax} = 0.72$ mean 0.66 and standard deviation 0.05.

Similarly, the SVM classifier was applied with the original, enhanced and four rule-based feature sets. The hyperparameters providing with the best A_z performance of the SVM scheme are: $C = 25 \times 10^4$, $\varepsilon = 0.001$ and $\gamma = 10^{-6}$. In the original feature set the use of PCA was beneficial for the method resulting in $A_z = 0.79$ using 20 and 18 support vectors (in this case is the same for both folds) for malignant and benign, respectively, as illustrated in Figure 7. In the case of the enhanced feature set A_z is 0.77, while when using only the four rule-based features $A_z = 0.67$. The BCR values were 0.81 and 0.78 for the original and enhanced feature set, respectively. In contrast to the ANN procedure, the use of the enhanced dataset does not lead to performance improvement.

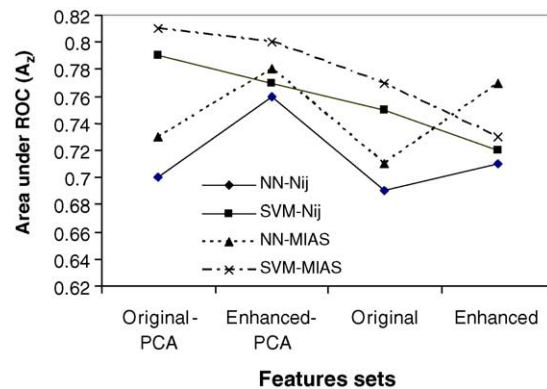


Figure 7 The performance of ANN and SVM classification methodologies for several feature sets, for both mammographic databases.

In order to evaluate the use of the PCA analysis to the classification system, additional experiments were realized for both classification systems. Several network architectures were utilized including one or two hidden layers. For the original feature set, the maximum value of A_z was $A_{zmax} = 0.69$ with mean $A_z = 0.66$ and standard deviation 0.03. For the enhanced feature set the best classification performance was $A_{zmax} = 0.71$ with mean $A_z = 0.68$ and standard deviation 0.02 as shown in Figure 6.

Figure 7 presents the performance of the SVM classifier without PCA achieving $A_z = 0.75$ and $A_z = 0.72$ for the original and the enhanced feature sets, respectively. Two groups of support vectors were generated for each training fold consisting of (26,18) and (26,17) vectors for the original feature set and (22,16) and (22,17) for the enhanced feature set.

3.2. MIAS dataset

Following the same methodology for evaluation, we achieved the correct characterization of 11 (61%) malignant clusters and the false classification of 1 (7.6%) benign. The above obtained true and false characterizations for the independent application of each rule, are quite low and thus a rule based classifier could not provide useful diagnostic information.

The ANN classifier using the same architecture as in the Nijmegen dataset was applied on the original and the enhanced feature sets. For the original feature set the maximum value of A_z was $A_{zmax} = 0.73$, the mean $A_z = 0.66$ and the standard deviation 0.05. For the enhanced feature set, A_z is higher corresponding to maximum performance ($A_{zmax} = 0.78$), the mean value is 0.73 and the stan-

standard deviation 0.05 as shown in Figure 7. The BCR values were 0.74 and 0.80 for the original and enhanced feature set, respectively. When only the four rule-based features were used, $A_{z\max} = 0.70$, the mean is 0.67 and the standard deviation is 0.03.

In addition, we utilized the SVM classifier for the three groups of features. The hyperparameters which provide the highest A_z performance are: $C = 2 \times 10^5$, $\varepsilon = 0.001$ and $\gamma = 10^{-5}$. Employing the original feature set, the SVM algorithm produced two pairs of vectors (one for each fold) composed of (16,14) and (17,16) support vectors for the malignant and benign samples, respectively. The classification performance is $A_z = 0.81$. Utilizing the enhanced feature set the classification performance is lower, $A_z = 0.80$ as presented in Figure 7. Moreover, utilizing the rule-based features only, $A_z = 0.68$. The BCR values are 0.83 and 0.82 for the original and enhanced feature set, respectively.

The use of the ANN classification system without PCA is assessed in the MIAS dataset. For the original feature set the maximum value of A_z was $A_{z\max} = 0.71$ with mean 0.69 and standard deviation 0.02. For the enhanced feature set the classification performance was $A_{z\max} = 0.73$ with mean 0.70 and standard deviation 0.03 as shown in Figure 7. The BCR values were 0.73 and 0.74 for the original and the enhanced feature sets, respectively.

Employing the SVM classifier without PCA, $A_z = 0.77$ and $A_z = 0.73$ were achieved for the original and the enhance feature set, respectively. The BCR values are 0.79 and 0.75 for the two feature sets. Two groups of support vectors were generated for each training fold which consist of (14,12) and (15,14) vectors for the original feature set and (16,13) and (18,17) vectors for the enhanced feature set.

4. Discussion and conclusions

A methodology for the characterization of microcalcification clusters in digitised mammograms as malignant or benign has been developed. In the final stage of the methodology two major classes of classifiers have been used: ANNs and SVMs. SVMs using a Gaussian kernel function provided the best performance (classification rate for Nijmegen dataset is 0.81 and $A_z = 0.79$, classification rate for MIAS dataset is 0.83 and $A_z = 0.81$) using the original feature set, which consists of 33 features. The classification performances are shown graphically in Figure 8.

Comparison of our methodology with others reported in the literature is not straightforward because experiments were conducted on different

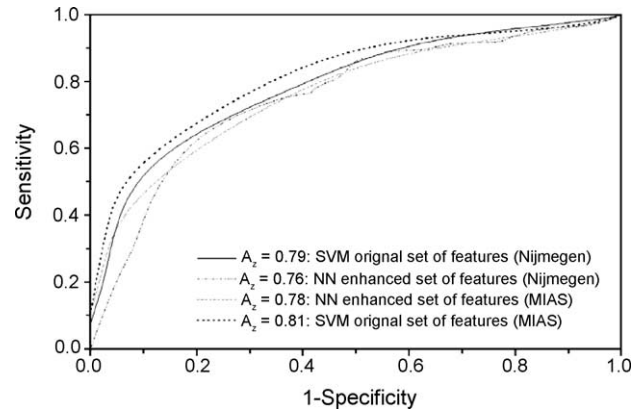


Figure 8 The ROC curves of SVM and ANN classifiers for the original and the enhanced feature sets evaluated for both mammographic datasets (best cases).

datasets. Using human extracted feature characterization $A_z = 0.89$ has been reported [2,3], while manual cluster specification resulted in $A_z = 0.83$ [25] and $A_z = 0.89$ [8]. Experiments with the Nijmegen database are reported in [27] where an automated method is presented exhibiting sensitivity 0.77 with specificity 0.90. For the same dataset BCR = 0.75 is reported in [30].

In addition, the proposed approach compares well with other methods. The k -nearest neighbour classification technique proposed by Zadeh et al. [23] results in $A_z = 0.82$ for the Nijmegen dataset. Kramer and Aghdasi [20] report a classification accuracy 100% in the Nijmegen dataset using wavelet and co-occurrence feature vector as input to their characterization system. A multiple classifier system composed of two separate (parallel) classifiers, the “ μ C-Expert” and the “Cluster Expert” is used in [48] and results in $A_z = 0.79$ for the Nijmegen dataset.

The selection of the methodology providing the best results in the microcalcification characterization is a difficult task, since most of the techniques are evaluated in different, and in some cases in custom datasets. In addition several research groups employ manually identified cluster ROIs in order to train and test their system, thus resulting in more subjective detections and performance ($A_z \leq 0.90$). It is noted that the radiologists interpretation performance corresponds to A_z values in the range 0.54–0.72 for attendants and in the range 0.53–0.66 for residents [49].

The two classification approaches, ANNs and SVMs, are based on different theoretical concepts. The ANN methodology aims at drawing a non-linear decision boundary in the data space through the minimization of a quadratic error function assuming a specific network architectural design (number of

hidden nodes). On the other hand, the SVM method draws a linear decision boundary on a higher dimensional space (specified by the kernel function) and attempts to minimize the number of training examples that fall inside the margin of the separation between the two classes. For both methodologies there are some issues to be considered during their application. In the case of SVM, the major difficulty concerns the identification of a “good” kernel function. On the other hand, the problem with ANNs is the specification of the neural architecture (number of hidden nodes). However, regularization methods have been integrated into the training techniques in order to deal with model complexity.

An advantage of SVM methodology is that the training procedure always converges to a specific solution corresponding to the global minimum of the objective function. In ANNs the existence of several poor local minima that may trap the training procedure constitutes a considerable drawback.

The present work has proposed a methodology to extract a new type of features, called the rule-based features. The addition of such features improves significantly the performance of ANNs but the same does not happen for SVMs as shown in Figure 8. We have also found that the reduction of feature dimensionality using PCA was beneficial for both classification methods. Finally, the best performance was achieved with SVMs, which offer the additional advantage that their performance does not depend on parameter initialisation, as happens with ANN methods.

The proposed approach is novel, in the sense that the classifier uses features which are not demographic or image extracted. This leads to better performance in the case of ANNs. Those features are considered meta-features and are expressed as linear rules, which are extracted visually. This approach can be proven extremely useful in other CAD systems.

The characterization procedure of the proposed methodology is fully automated. It is executed in three stages and to our knowledge exhibits better performance compared with other fully automated methods for the Nijmegen or the MIAS database. Its performance can be further improved if the cluster boundaries were more precisely identified. This constitutes a subject for future work.

Although the obtained high classification performance can be successfully applied to microcalcification clusters characterization, further studies must be carried out for the clinical evaluation of the system using larger datasets. Also, the use of additional features originating either from the image itself (such as cluster location and orientation) or from the patient data may further improve

the diagnostic value of the system. Finally, it would be interesting to experiment with image fusion techniques to combine information obtained from both mammographic views (MLO and CC).

Acknowledgement

The Nijmegen database was provided by courtesy of the National Expert and Training Centre for Breast Cancer Screening and the Department of Radiology at the University of Nijmegen, The Netherlands.

References

- [1] Lanyi M. *Diagnosis and Differential Diagnosis of Breast Calcifications*. Berlin: Springer-Verlag; 1988.
- [2] Baker JA, Kornguth PJ, Lo JY, Williford ME, Floyd CE. Breast cancer: prediction with artificial neural network based on BI-RADS standardized lexicon. *Radiology* 1995;196: 817–22.
- [3] Wu Y, Giger ML, Doi K, Vyborny CJ, Schmidt RA, Metz CE. Artificial neural networks in mammography: application to decision making in the diagnosis of breast cancer. *Radiology* 1993;187:81–7.
- [4] Dengler J, Behrens S, Desage JF. Segmentation of microcalcifications in mammograms. *IEEE Trans Med Imaging* 1993; 12:634–42.
- [5] Nakayama R, Uchiyama Y, Hatsukade I, Yamamoto K, Watanabe R, Namba K, et al. Discrimination of malignant and benign microcalcification clusters on mammograms. *J Comp Aid Diag Med Imag* 1999;3:1–7.
- [6] Taylor P, Fox J, Pokropek AT. The development and evaluation of CADMIUM: a prototype system to assist in the interpretation of mammograms. *Med Imag Anal* 1999;3:321–37.
- [7] Zhao D. Rule-based morphological feature extraction of microcalcifications in mammograms. In: Loew MH, editor. *Proc. SPIE, Medical Imaging 1993: Image Processing*. 1993.p. 7020–715.
- [8] Chan H-P, Lam KL, Petrick N, Helvie MA, Goodsitt MM, Adler DD. Computerized analysis of mammographic microcalcifications in morphological and texture feature space. *Med Phys* 1998;25:2007–19.
- [9] Meersman D, Scheunders P, Van Dyck D. Classification of microcalcifications using texture-based features. In: Karssemeijer N, Thijssen M, Hendriks J, Van Erning L, editors. *Proceedings of the 4th International Workshop on Digital Mammography*. 1998. p. 233–6.
- [10] Rogova GL, Stomper PC, Ke C-C. Microcalcification texture analysis in a hybrid system for computer-aided mammography. In: Hanson KM, editor. *Proc. SPIE, Medical Imaging 1999: Image Processing*. 1999.p. 14260–1433.
- [11] Lefebvre F, Benali H, Gilles R, Kahen E, Di Paola R. A fractal approach to the segmentation of microcalcifications in digital mammograms. *Med Phys* 1995;22:381–90.
- [12] Li H, Liu KJR, Lo SCB. Fractal modelling and segmentation for the enhancement of microcalcifications in digital mammograms. *IEEE Trans Med Imag* 1997;16:785–98.
- [13] Gavrielidis A, Lo JY, Vargas R, Floyd CE. Segmentation of suspicious clustered microcalcifications in mammograms. *Med Phys* 2000;27:13–22.

- [14] Strickland RN, Hahn HI. Wavelet transforms for detecting microcalcifications in mammography. *IEEE Trans Med Imag* 1996;15:218–28.
- [15] Wang TC, Karayannis NB. Detection of microcalcifications in digital mammograms using wavelets. *IEEE Trans Med Imag* 1998;17:498–509.
- [16] Yoshida H, Doi K, Nishikawa RM. Automated detection of clustered microcalcifications in digital mammograms using wavelet transform techniques. In: Loew MH, editor. *Proc. SPIE, Medical Imaging 1994: Image processing*. 1994.p. 8680–886.
- [17] Yu S, Guan L. A CAD system for the automated detection of clustered microcalcifications in digitised mammogram films. *IEEE Trans Med Imag* 2000;19:115–26.
- [18] Kallergi M. Computer-aided diagnosis of mammographic microcalcification clusters. *Med Phys* 2004;31:314–26.
- [19] Bottema MJ, Slavotinek JP. Detection and classification of lobular and DCIS (small cell) microcalcifications in digital mammograms. *Pattern Recog Lett* 2000;21:1209–14.
- [20] Kramer D, Aghdasi F. Classification of microcalcifications in digitised mammograms using multiscale statistical texture analysis. In: *Proceedings of the South African IEEE Symposium on Communications and Signal Processing*; 1998. p. 121–6.
- [21] Soltanian-Zadeh H, Pourabdollah-Nezhad S, Rafiee F. Shape-based and texture-based feature extraction for classification of microcalcifications in mammograms. In: Sonka M, Hanson KM, editors. *Proc SPIE, Medical Imaging 2001: Image Processing*. 2001.p. 3010–310.
- [22] Veldkamp W, Karssemeijer N, Otten JDM, Hendriks JHCL. Automated classification of clustered microcalcifications into malignant and benign. *Med Phys* 2000;27:2600–8.
- [23] Zadeh HS, Nezhad SP, Rad FR. Shape-based and texture-based feature extraction for classification of microcalcification in mammograms. In: Sonka M, Hanson KM, editors. *Proc. SPIE, Med Imaging: Image Processing*. 2001.p. 3010–310.
- [24] Hara T, Yamada A, Fujita H, Iwase T, Endo T. Automated classification of mammographic microcalcifications by using artificial neural networks and ACR BI-RADS criteria. In: Sonka M, Hanson KM, editors. *Proc SPIE, Medical Imaging 2001: Image processing*. 2001.p. 17830–1787.
- [25] Jiang Y, Nishikawa RM, Metz CE, Wolverton DE, Schmidt RE, Papaioannou J, et al. A computer-aided diagnostic scheme for classification of malignant and benign clustered microcalcifications in mammograms. In: Doi K, Giger ML, Nishikawa RM, Schmidt RA, editors. *Proceedings of the 3rd International Workshop on Digital Mammography*. 1996. p. 219–24.
- [26] Jiang Y, Nishikawa RM, Wolverton DE, Metz CE, Giger ML, Schmidt RA, et al. Malignant and benign clustered microcalcifications: automated feature analysis and classification. *Radiology* 1996;198:671–8.
- [27] Lee SK, Lo CS, Wang CM, Chung PC, Chang CI, Yang CW, et al. A computer-aided design mammography screening system for detection and classification of microcalcifications. *Int J Med Inform* 2000;60:29–57.
- [28] Schmidt F, Sorantin E, Szepesvari C, Graif E, Becker M, Mayer H, et al. An automatic method for the identification and interpretation of clustered microcalcifications in mammograms. *Phys Med Biol* 1999;44:1231–43.
- [29] Shen L, Rangyyan RM, Desautels JEL. Application of shape analysis to mammographic calcifications. *IEEE Trans Med Imag* 1994;13:263–74.
- [30] Verma BK. Comparative evaluation of two neural network based techniques for classification of microcalcifications in digital mammograms. *Knowl Inform Syst Inter J* 1999;1:107–17.
- [31] Cheng HD, Cai X, Chen X, Hu L, Lou X. Computer-aided detection and classification of microcalcifications in mammograms: a survey. *Pattern Recog* 2003;26:2967–91.
- [32] Papadopoulos A, Fotiadis DI, Likas A. An automatic microcalcifications detection system based on a hybrid neural network classifier. *Artif Intell Med* 2002;25:149–67.
- [33] Metz CE. ROC methodology in radiologic imaging. *Invest Radiol* 1986;21:720–33.
- [34] Karssemeijer N. Adaptive noise equalization and recognition of microcalcifications in mammography. *Inter J Pattern Recog Artif Intell* 1993;7:1357–76.
- [35] Suckling J, Parker J, Dance D, Astley S, Hutt I, Boggis C. The mammographic images analysis society digital mammogram database. *Exerpta Med* 1994;1069:375–8.
- [36] Bishop CM. *Neural networks for pattern recognition*. Oxford: Oxford University Press; 1996.
- [37] Battiti R. First and second order methods for learning: between steepest descent and Newton's method. *Neural Comput* 1992;4:141–66.
- [38] Burges CJC. A tutorial on support vector machines for pattern recognition. *Knowl Discov Data Mining* 1998;2:1–43.
- [39] Cristianini N, Shawa-Taylor J. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge: UK: Cambridge University Press; 2000.
- [40] Bazzani A, Bevilacqua A, Bollini D, Braccaccio R, Campanini R, Lanconelli N, et al. Automated detection of clustered microcalcifications in digital mammograms using an SVM classifier. In: *Proceedings of the 8th Eur Symp on Artificial Neural Network*; 2000. p. 195–200.
- [41] El-Naqa I, Yang Y, Wernick MN, Galatsanos NP, Nishikawa R. Support vector machine learning for the detection of microcalcifications in mammograms. *IEEE Trans Med Imag* 2002; 21:1552–63.
- [42] Vapnik VN. *The Nature of Statistical Learning Theory*. Berlin: Springer-Verlag; 1995.
- [43] Chang C-C, Lin C-J. LIBSVM: Introduction and benchmarks, 2001 [Online] Available: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [44] Joachims T. Making large-scale SVM learning practical. In: Scholkopf B, Burges CJC, Smola AJ, editors. *Advances in Kernel Methods—Support Vector Learning*. Cambridge: MA: MIT Press; 1998.
- [45] Keerthi SS, Shevade SK, Bhattacharyya C, Murthy KR. Improvements to Platt's SMO algorithm for SVM classifier design. IISc, Dept of CSA, Technical Report, Bangalore, India, 1999.
- [46] Platt J. Sequential minimal optimization: a fast algorithm for training support vector machines, Microsoft Research, Technical Report MSR-TR-98-14, 1998.
- [47] Chang C-C, Lin C-J. LIBSVM: a library for support vector machines (v.2.33) 2002 [Online], Available: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [48] Foggia P, Sansone C, Tortorella F, Vento M. Automatic classification of clustered microcalcifications by a multiple classifier system. In: Kittler J, Roli F, editors. *Proceedings of the MCS 2001, Lecture Notes in Computer Science 2096*. 2001. p. 208–17.
- [49] Jiang Y, Nishikawa RM, Schmidt RA, Toledano AY, Doi K. Potential of computer-aided diagnosis to reduce variability in radiologists' interpretations of mammograms depicting microcalcifications. *Radiology* 2001;220:787–94.