CHARACTERIZATION OF CLUSTERED MICROCALCIFICATIONS IN DIGITIZED MAMMOGRAMS USING NEURAL NETWORKS AND SUPPORT VECTOR MACHINES

A. Papadopoulos, D.I. Fotiadis and A. Likas

20–2002

Preprint, no 20–02/2002

Department of Computer Science
University of Ioannina
45110 Ioannina, Greece
Characterization of Clustered Microcalcifications in Digitized Mammograms using Neural Networks and Support Vector Machines

A. Papadopoulos\textsuperscript{1}, D.I. Fotiadis\textsuperscript{2} and A. Likas\textsuperscript{2}

\textsuperscript{1} Department of Medical Physics, Medical School, University of Ioannina, 
\textsuperscript{2} Department of Computer Science, University of Ioannina, 
GR 45110 Ioannina, Greece

Correspondence should be addressed to
Dr. D. I. Fotiadis
Department of Computer Science
University of Ioannina
45110 Ioannina, GREECE
Tel:+30 6510 98803, fax: +30 651 098889
c-mail:fotiadis@cs.uoi.gr
Abstract

We have developed an automated method for the characterization of microcalcification clusters in digitized mammograms as malignant or benign. 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 an expert rule-based system, a neural network and a support vector machine (SVM) were implemented and evaluated using ROC analysis. We have conducted experiments using the Nijmegen database. The original feature set was enhanced by the addition of four rule-based features. The performance of SVM was $A_z = 0.77$ and $A_z = 0.79$ for the original and enhanced feature set, respectively, while the corresponding performance of the neural network was $A_z = 0.70$ and $A_z = 0.76$.

Keywords- Support vector machines, Neural network, Classification, Microcalcification cluster, mammography
1 Introduction

Several methodologies have been developed in order to improve radiologists' efficiency in the diagnostic interpretation of mammograms. The successful development of CAD (Computer Aided Diagnosis) systems will be of great value, when these systems are used as second opinion to the radiologist. CAD systems integrate image analysis and artificial intelligent techniques aiming at providing with 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 the more difficult sub-problem is the characterization of the identified lesions as malignant or benign (LANYI, 1988). Such a successful characterisation can contribute to the reduction of unnecessary biopsies.

The most common approach for the development of CAD systems includes feature extraction algorithms materialized either using computer systems or through manual extraction by radiologists (BAKER et al., 1995; WU et al., 1993) that compute important features subsequently fed into a classifier. 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 (TAYLOR et al., 1999; NAKAYAMA et al., 1999; DENGLER et al., 1993; ZHAO, 1993), texture (CHAN et al., 1998; ROGOVA et al., 1999; MEERSMAN et al., 1998), fractal (LEFEBVRE
et al., 1995, Li et al., 1997), histogram statistics (Gavriolidis et al., 2000) and wavelets (Yu et al., 2000; Strickland et al., 1996; Wang et al., 1998; Yoshida et al., 1994). Morphological features are the most commonly used due to their similarity with the characteristics taken into account by the radiologists.

Several methodologies have been presented for the lesion characterization problem, such as decision trees (Taylor et al., 1999, Bottema et al., 2000), linear discriminant classifier (Chan et al., 1998; Nakayama et al., 1999), k-nearest neighbours (Veldkamp et al., 2000; Soltanian et al., 2001) and neural networks (Baker et al., 1995; Wu et al., 1993; Jiang et al., 1996a; Jiang et al., 1996b; Schmidt et al., 1999; Shen et al., 1994; Lee et al., 2000; Hara et al., 2001; Verma, 1999). 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.

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 an previous work (Papadopoulos et al., 2002), the feature computation stage and the final classification stage. Two different classification schemes have been implemented and tested based on neural networks and SVMs. It must be noted that SVMs are used for the first for the cluster characterization problem. Originally, among a set of 54 features we have selected 33 of them. In addition, we have defined a new type of features, called rule-based features, which are obtained using 2D graphical representation with respect to all pairs of features. This resulted in an enhanced feature set consisted of 37 features.
The performance of the classifiers has been evaluated using the receiver operating characteristic (ROC) methodology (METZ, 1986) and the classification rate. The obtained results are very encouraging and thus our method can be considered very promising.

2 Materials and Methods

2.1 Image Dataset
We use the Nijmegen mammographic image database (KARSEMEIJER, 1993). It consists of 40 images of both craniocaudal (CC) and oblique (MLO) views from 21 patients. The digitisation sampling aperture is 0.05 mm and the sampling distance 0.1 mm and the size of each image is 2048 × 2048 pixels. 12 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 which it is provided with the database. The microcalcification clusters have been annotated in each image by expert radiologists using a circle enclosing the abnormality. The total number of annotated clusters is 105, which correspond to 76 malignant and 29 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 with the final diagnosis.

2.2 Cluster Detection Procedure
The objective of this stage is the identification of clusters of microcalcifications. The procedure we followed has been described in an earlier work (PAPADOPOULOS et al., 2002) and is based on a hybrid intelligent system combining rule-based and neural network methods. Fig. 1 shows the sequential 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 are applied to indicate potential microcalcification objects. Morphological descriptors are used to extract the regions of interest (ROIs). Then, for all objects and clusters contained in every ROI we compute several discriminating morphological and textural features, which are used as input to the classification system. This system is a hybrid intelligent system based on a combination of a rule-based and a neural network component providing with characterization of the ROI as cluster of microcalcifications or not. For the Nijmegen database the detection module provides 115 regions of interest corresponding to 80 malignant and 35 benign clusters.

2.3 Features Computation and Selection

For each detected cluster 54 features have been identified, which are shown in Table 1. Those features refer either to an individual microcalcification (microcalcification feature) or as averages of the five largest microcalcifications included in a cluster (cluster feature).

In order to reduce the number of features a feature selection procedure based on ROC analysis has been followed to identify the most discriminative features. The ROC curve (METZ, 1986) has been plotted for each feature and the area $A_z$ under the curve
is computed. Using a quite low threshold value $A_z = 0.52$ we selected 33 features having $A_z$ value higher than the threshold. The majority of features are the same to the features used for the detection procedure described above. The use of extra features underlines the increased difficulty of the classification of the clusters to benign or malignant compared to the removal of the false positive detected samples.

2.4 Classification methods

The aim of the classification process is the characterization of each cluster as malignant or benign using the selected features. In this work we have employed rule based expert systems, neural networks and support vector machines.

2.4.1 Rule based expert system

The common approach in this class of systems includes the use of rules applying thresholds is selected features. However, in our case it is not possible to define discriminative rules of this type. Thus, we followed the following approach: Initially, a 2-D graphical representation of the dataset in two feature space has been performed for each pair of features (e.g. see Fig. 2). It was understood that four of those plots have the highest discriminative value. That means that in the two feature space a straight line (defining a linear decision rule) can be drawn which defines a region (half space) containing mostly (more than 90%) points belonging to the same class. The pairs of features corresponding to those plots 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.
This set of rules cannot be used as an independent classifier due to its poor performance (see Table 2). However, the distance of a point from the corresponding linear boundary constitutes an additional feature to be used in the classification system. It must be also noted that each of these distance features has a sign which indicates the half space in which the data point lies. The incorporation of the additional four features into the initial set of selected features results in an enhanced feature set (it consists of 37 features).

2.4.2 

Neural Classifier

The neural network classifier we have used is a feedforward network with sigmoid hidden nodes (Multilayer Perceptron – MLP). The selected neural network architecture consists of one hidden layer with fifteen sigmoid nodes, and an output layer with one sigmoid node, whose value indicates malignant or benign microcalcification cluster. Principal component analysis (PCA) is implemented in order to reduce the size of the input feature vector. The output of the PCA is an reduced feature vector composed of seven features that contribute 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 neural network taining in order to select the one with the best classification ability (BISHOP, 1996). The training procedure is terminated either when that training error is less than $10^{-5}$ or when 2000 iterations have been performed. The highest performance was obtained using quasi-Newton one-step-secant (OSS) algorithm (BATTITI, 1992). The two fold cross validation method is utilized for the performance assessment. When the
enhanced feature set is used the classification performance, using the same neural network architecture and PCA, is improved.

2.4.3 Support Vector Machines

Another category of classification methods that has received special attention recently is the Support Vector Machines (BURGES, 1998; CRISTIANINI et al., 2000). SVMs have not used previously for clusters of microcalcification characterization but only for their detection (BAZZANI et al., 2000; EL-NAQA et al., 2002). 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, an SVM aims to the minimization of the empirical risk $R_{emp}$ and at the same time, the maximization of the distances (geometric margin $M$) of the data points from the corresponding linear decision boundary (Fig. 3). $R_{emp}$ is defined as

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

where $x_i \in \mathbb{R}^N$, $i = 1, \ldots, l$, 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 $\mathbb{R}^N$ are mapped nonlinearly into a higher dimensional space $F$ by the function (a priori selected) $\Phi: \mathbb{R}^N \rightarrow F$. In this space (feature space) the decision hyperplane is computed. The training algorithm uses only the dot products $(\Phi(x_i) \cdot \Phi(x_j))$ in $F$. If there exists a “kernel function” $K$ such that $K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)$, then only the knowledge of $K$ is required by the training algorithm. The decision function is defined as

$$f(x) = \sum_{i=1}^{l} y_i \alpha_i K(x_i, x) + b$$
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 the others it holds that \( a_i = 0 \). 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 of major importance for the performance of the classifier. Several types of kernels have been reported in the literature, such as the polynomial type of degree \( p \), \( K(x_i, x) = (x_i \cdot x + 1)^p \) and the Gaussian kernel \( K(x_i, x) = e^{-\|x-x\|^2/2\sigma^2} \) (\( \sigma \) is the kernel width). Each kernel function should fulfill Mercer's condition (BURGES, 1998; VAPNIK, 1995).

We use the SVM algorithm provided by the LIBSVM library (CHANG et al., 2002), which has been proven to be stable, computationally inexpensive and highly competitive compared with other SVM codes [PLATT, 1998; KEERTHI et al., 1999; JOACHIMS, 1998; CHANG et al., 2001].

We have tested the Gaussian kernel for several values of the standard deviation \( \sigma \). In order to apply the SVM training algorithm the following parameters must be adjusted: the regularization parameter \( C \) and the termination criterion \( \varepsilon \). To perform parametrization we have applied the training algorithm for the following values of the parameters: \( \gamma \in \{10^{-5}, 10^{-4}, ..., 0.01, 0.5\} \), \( C \in \{1, 10, ..., 10^3\} \) and \( \varepsilon \in \{10^{-5}, ..., 10^{-1}\} \).

3 Experimental Results
The above described methodology is evaluated on the Nijmegen mammographic dataset. The database reference file reports 105 cluster areas. Applying the detection algorithm we have identified 15 more areas because in some cases two separate ROIs had been detected into one database annotated circle. From the total 120 ROIs, 84 are malignant and 36 are benign. To assess the performance of our method we have employed ROC analysis (computation of $A_2$). Moreover, in order to compare our results with previous reported we have computed the best classification rate (BCR=ratio of the sum of true positives and true negatives over the total number of samples) for a range of threshold values.

For the rule based classifier, if a decision rule is valid then the cluster is classified as malignant, otherwise as unclassified. Using the expert system with the four linear decision rules we achieved the correct characterization of 44 (52%) malignant clusters and the false characterization of 2 (5.8 %) benign. The decision is based on the 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 characterizations are shown in Table 2. It is clear that the obtained performance is not acceptable.

The neural network classifier has been used for the initial and the enhanced feature sets. To evaluate the performance of the method we have used two fold cross validation and we performed ROC analysis. For each training set (fold) we have applied the training algorithm ten times with different initial weight values. For the initial feature set $A_{\text{max}}=0.70$. The mean $A_2$ is 0.65 and the standard deviation is 0.04. For the enhanced feature set the classification performance is improved remarkably.
resulting in $A_{z\text{max}}=0.76$ with mean $A_z=0.71$ and standard deviation 0.04. The BCR values were found 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 poor resulting $A_{z\text{max}}=0.72$ mean 0.66 and standard deviation 0.05.

Similarly, the SVM classifier has been used with the initial, enhanced and four rule based feature sets. The hyperparameters providing with the best $A_z$ performance of the SVM scheme are: $C=25\times10^4$, $\varepsilon=0.001$ 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 for malignant and benign, respectively (Fig. 5). In the case of enhanced feature set we have obtained $A_z=0.77$, while in the case of using only the four rule based features we have obtained $A_z=0.67$. The BCR values were 0.81 and 0.78 for the original and enhanced feature set, respectively. In contrast to the neural network case the use of the enhanced dataset does not lead to performance improvement.

4 Discussion and conclusions

A methodology for the characterization of microcalcification clusters in digitised mammograms in malignant or benign has been developed. In the final stage of the methodology two major classes of classifiers have been used: neural networks and support vector machines. SVMs using Gaussian kernel function provided with the best performance (classification rate 0.81 and $A_z=0.79$) for the original feature set, which consists of 33 features.
Comparison of our methodology with other reported in the literature is not straightforward because experiments were conducted on different datasets. Using human extracted feature characterization $A_x=0.89$ was reported (Baker et al., 1995; Wu et al., 1993). Manual specification resulted in $A_x=0.83$ (Jiang et al., 1996a) and $A_x=0.89$ (Chan et al., 1998). Experiments with the Nijmegen database are reported in (Lee et al., 2000) where an automated method is presented exhibiting sensitivity 0.77 with specificity 0.90. For the same dataset $BCR=0.75$ is reported in (Verma, 1999).

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 database. Its performance can be further improved if the cluster boundaries were more precisely identified. This constitutes a subject for future work.

In our method we have proposed a methodology to extract a new type of features which called rule based features. The addition of such features improves significantly the performance of neural networks but the same does not happen for SVMs. We have also found that the reduction of the dimensionality of the feature vector 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, which is the case for neural network methods.
Although the obtained characterization performance can be considered satisfactory, further studies must be carried out for the evaluation of the system with larger datasets. Also, the use of additional features originating from the image itself, such as cluster location and orientation, and patient data can further improve the diagnostic value of the system. Finally, it will be interesting to experiment with image fusion techniques to combine information obtained from both mammographic views (MLO and CC).

Acknowledgements - 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


**Table 1: Features for cluster categorization.**

<table>
<thead>
<tr>
<th>Microcalcification (MC) cluster classification features</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area of the cluster convex hull</td>
<td>Mean MCs intensity</td>
</tr>
<tr>
<td>Cluster area</td>
<td>Mean perimeter of MCs in cluster</td>
</tr>
<tr>
<td>Cluster eccentricity</td>
<td>Minor cluster’s axis (equivalent ellipse)</td>
</tr>
<tr>
<td>Cluster elongation</td>
<td>Neighbouring with a larger cluster</td>
</tr>
<tr>
<td>Cluster Entropy</td>
<td>Number of MCs in cluster</td>
</tr>
<tr>
<td>Clusters’ equivalent diameter</td>
<td>Orientation of cluster</td>
</tr>
<tr>
<td>Extent of cluster</td>
<td>Solidity of cluster</td>
</tr>
<tr>
<td>Filled area in cluster</td>
<td>Spreading of MCs in cluster</td>
</tr>
<tr>
<td>Major cluster’s axis (equivalent ellipse)</td>
<td>STD of distances from cluster centroid</td>
</tr>
<tr>
<td>Mean contrast</td>
<td>STD of MC compactness</td>
</tr>
<tr>
<td>Mean distance from cluster centroid</td>
<td>STD of MC elongation</td>
</tr>
<tr>
<td>Mean local MC background</td>
<td>STD of MC intensity</td>
</tr>
<tr>
<td>Mean MC area</td>
<td>STD of MCs area</td>
</tr>
<tr>
<td>Mean MC background intensity</td>
<td>STD of MCs contrast in cluster</td>
</tr>
<tr>
<td>Mean MC compactness</td>
<td>STD of MCs perimeter in cluster</td>
</tr>
<tr>
<td>Mean MC elongation</td>
<td>The length of the cluster convex hull</td>
</tr>
<tr>
<td>Mean MCs eccentricity</td>
<td></td>
</tr>
<tr>
<td>Pairs of features</td>
<td>Malignant (true characterization)</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>Mean cluster eccentricity – mean contrast</td>
<td>28</td>
</tr>
<tr>
<td>Mean local background – mean distance</td>
<td>44</td>
</tr>
<tr>
<td>from cluster centroid</td>
<td></td>
</tr>
<tr>
<td>Mean contrast - mean compactness</td>
<td>33</td>
</tr>
<tr>
<td>Standard deviation of the</td>
<td></td>
</tr>
<tr>
<td>microcalcification distances from cluster</td>
<td>41</td>
</tr>
<tr>
<td>centroid – equivalent diameter of the</td>
<td></td>
</tr>
<tr>
<td>cluster</td>
<td></td>
</tr>
</tbody>
</table>
List of figure captions

**Fig 1.** The microcalcification cluster detection system.

**Fig 2.** Linear decision approximation for the development of rule-extracted features from the 2D plot of mean contrast and mean compactness features in each cluster.

**Fig 3.** The performances of one-hidden layer network architectures for several numbers of hidden nodes.

**Fig 4.** The performances of neural networks (one-hidden layer, fifteen nodes) for several PCA outputs are plotted. The maximum, the average and the standard deviation of the Az values are presented for each network.

**Fig 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 $\Phi$.

**Fig 6.** The ROC curves of SVM and neural network classifiers (best performance).
Fig 1. The microcalcification cluster detection system.
Fig 2. Linear decision approximation for the development of rule-extracted features from the 2D plot of mean contrast and mean compactness features in each cluster.
Fig 3. The performances of one-hidden layer network architectures for several numbers of hidden nodes.
Fig 4. The performances of neural networks (one-hidden layer, fifteen nodes) for several PCA outputs are plotted. The maximum, the average and the standard deviation of the Az values are presented for each network.
Fig 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 $\Phi$. 
Fig 6. The ROC curves of SVM and neural network classifiers (best performance).