ELSEVIER

Artificial Intelligence in Medicine 25 (2002) 149–167

Artificial Intelligence in Medicine

www.elsevier.com/locate/artmed

An automatic microcalcification detection system based on a hybrid neural network classifier

A. Papadopoulos[a,*], D.I. Fotiadis[b], A. Likas[b]

[a]*Department of Medical Physics, Medical School, University of Ioannina, GR 45110 Ioannina, Greece*
[b]*Department of Computer Science, University of Ioannina, GR 45110 Ioannina, Greece*

Received 29 August 2001; received in revised form 19 November 2001; accepted 17 January 2002

Abstract

A hybrid intelligent system is presented for the identification of microcalcification clusters in digital mammograms. The proposed method is based on a three-step procedure: (a) preprocessing and segmentation, (b) regions of interest (ROI) specification, and (c) feature extraction and classification. The reduction of false positive cases is performed using an intelligent system containing two sub-systems: a rule-based and a neural network sub-system. In the first step of the classification schema 22 features are automatically computed which refer either to individual microcalcifications or to groups of them. Further reduction in the number of features is achieved through principal component analysis (PCA). The proposed methodology is tested using the Nijmegen and the Mammographic Image Analysis Society (MIAS) mammographic databases. Results are presented as the receiver operating characteristic (ROC) performance and are quantified by the area under the ROC curve (A_z). In particular, the A_z value for the Nijmegen dataset is 0.91 and for the MIAS is 0.92. The detection specificity of the two sets is 1.80 and 1.15 false positive clusters per image, at the sensitivity level higher than 0.90, respectively.

Keywords: Microcalcification detection; Hybrid neural network; Computer-aided detection (CAD); Mammography

1. Introduction

Breast cancer is currently one of the leading causes of death among women worldwide. Regular mammographic screening programs for women of certain age or high-risk groups are taking place in a number of countries on a nation-wide basis or as projects organized from several institutes [1,15,30,31,48]. Although some researchers doubt about the real

* Corresponding author. Tel.: +30-651-98803; fax: +30-651-98889.
E-mail address: fotiadis@cs.uoi.gr (A. Papadopoulos).

0933-3657/02/$ – see front matter
PII: S0933-3657(02)00013-1

effectiveness of population screening programs [17] the majority of them contribute to the mortality reduction [19,40,49]. Early detection is the key for improving breast cancer prognosis [44]. Mammography is the most common procedure for detecting non-palpable cancers. It is an inexpensive practice being highly effective even when the size of the breast abnormality is minimal [42,62]. One of the early signs of breast cancer is the presence of microcalcification clusters at the mammogram of asymptomatic women. However, a number of such findings especially the microcalcifications that have small size and low contrast could be missed or misinterpreted by doctors. Thus, the task of the radiologist is tedious in the case where a significant number of mammograms require fast and accurate interpretation. For this reason, a reliable automated computer-aided diagnosis system (CAD) could be very useful, providing a valuable "second opinion" to a radiologist, especially to a non-expert one [34,41]. Nevertheless, the application of CAD systems in clinical use should be done after extensively high assessments of their effectiveness in terms of sensitivity and reduction of false positive rate.

In the literature, several techniques have been proposed to detect the presence of microcalcifications using various methodologies. Concerning image segmentation and specification of regions of interest (ROIs), several methods have been proposed such as classical image filtering and local thresholding [9,12,39,45], techniques based on mathematical morphology [13,60], stochastic fractal models [25,26], wavelet analysis [3,7,22,23,46,52,56,57] and multiscale analysis based on a specialized Gaussian and Peitgen [32]. Furthermore, various classification methodologies have been reported for the characterization of ROI such as, rule-based systems [9,12], fuzzy logic systems [11], statistical methods based on Markov random fields [20] and support vector machines [3]. Nevertheless, the most work reported in the literature employs neural networks for cluster characterization [10,27,33,37,51,54,55,58,59,61]. Typically, a neural network accepts as input features computed for a specific region of interest and provides as output a characterization of the region as true microcalcification cluster or not. Recently, neural networks have also been used to characterize a microcalcification as malignant or benign [6,24,42,50].

In this paper, we present an intelligent system (Fig. 1) for the identification of microcalcification clusters in digitized mammographic images. The system, as it is described in Section 2, consists of three modules: the preprocessing and segmentation, (ROI) specification and the feature extraction and the classification module. The latter is a hybrid classification schema composed of a rule-based and a neural network sub-system. The proposed system is fast and accurate in the detection of ROIs. We employ an additional feature for ROI characterization that is related with the existence of a small ROI in the neighborhood of a large one. In addition, we have found that performance is improved in the case where principal component analysis (PCA) is used to reduce the number of features. The method provides satisfactory results in two well-known datasets: the Nijmegen and the Mammographic Image Analysis Society (MIAS) mammographic databases as described in Section 3. It must be noted that the proposed hybrid system performs better compared to the case where either the rule-based or the neural network subsystem are solely employed for classification. The proposed methodology could be an essential part of an integrated CAD technique, which could assist radiologists in mammogram analysis and diagnostic decision making. The system successfully combines

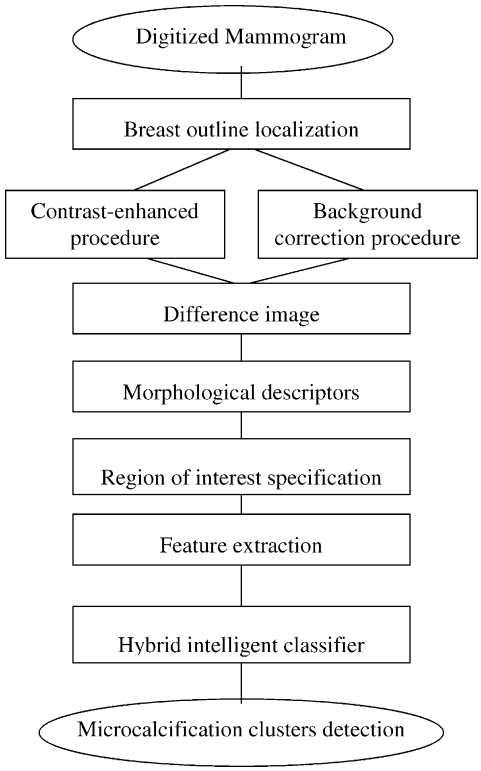

Fig. 1. The microcalcification cluster detection system.

intelligent methods and image processing techniques which contribute to the enhancement of mammographic diagnosis sensitivity and reduction of negative biopsies.

2. Material and methods

2.1. Image datasets

For the development and evaluation of the proposed system we used the Nijmegen [20] and the MIAS [47] databases. The first contains 40 mammograms of both craniocaudal and oblique views from 21 patients. Digitization has been carried out using an Eikonix 1412 CCD camera with 0.1 mm pixel size and 12 bit grey depth. The size of each image is 2048×2048 pixels. For each image a lookup table is provided for conversion-rescaling from 12 to 8 bit format based on noise characteristics [20]. One or more microcalcification clusters are annotated in each mammogram by expert radiologists using a circle enclosing the abnormality. The total number of annotated clusters in the database is 105. It must be noted that the Nijmegen dataset digitization characteristics are different from the MIAS dataset and we resampled the Nijmegen images to change the pixel size from 100 to 50 μm,

because our software originally was developed to handle 50 μm images. The resampling technique was actually a magnification process in which each one of the initial pixels was divided to four keeping the same intensity value.

The second dataset contains 20 images and has been developed by the MIAS (c) [47]. Each mammographic image is obtained from the medio-lateral oblique view and is digitized with spatial resolution 50 μm and 8 bit grey depth. A circle enclosing the abnormality indicates each cluster area. The database contains 25 annotated clusters.

The proposed system is implemented in three stages. The first is related to image segmentation, the second with the identification of candidate ROIs, and the third with the characterisation of each ROI as cluster of microcalcifications or not.

2.2. Preprocessing and segmentation module

In a typical mammogram several different areas are present such as the image background, the tissue area, and informative marks. At the beginning of preprocessing it is necessary to locate the breast region. For this reason we apply a skin-line segmentation procedure by setting equal to zero the image pixels with intensity less than 20 (for 0–255 Gy levels). Most of those pixels belong to the background area, although a small number exists belonging to the tissue area close to the breast surface. This thresholding procedure results in a binary image of white objects on a black background. Neighbouring white pixels with connectivity of eight are grouped together to form objects corresponding either to the breast region or to marks and film artifacts. The largest object corresponds to the breast region (Fig. 2) and close to the breast outline a number of very small objects appear. These are actually part of the breast region but, due to thresholding, they appear as distinct objects. To deal with this problem, we apply morphological dilation with a structure element radius of 30 pixels (∼1.5 mm) [16]. This results in an expansion of breast

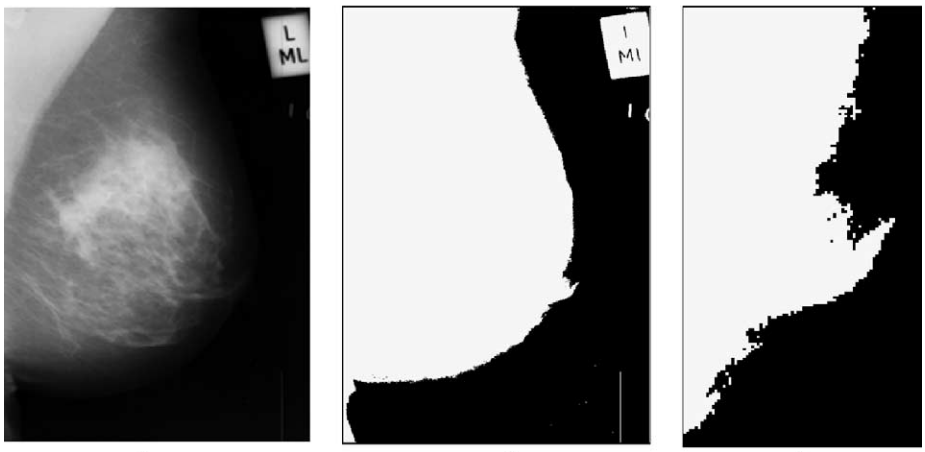

Fig. 2. (a) An original mammogram, (b) the different objects appearing in the binary image and (c) a zoom view in the area of the breast skinline.

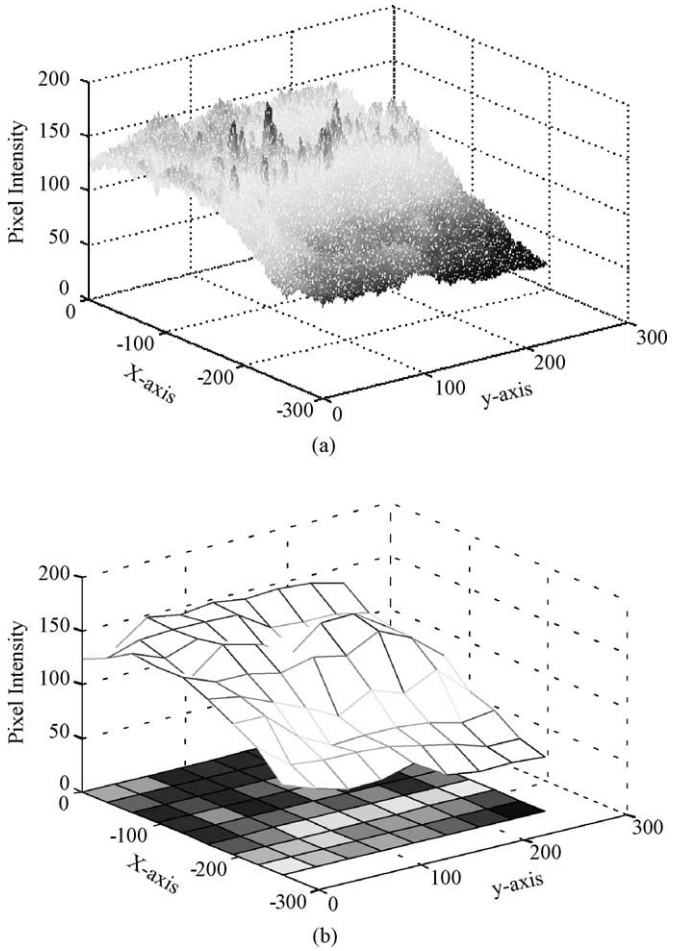

Fig. 3. (a) Three-dimensional intensity representation of a 300×300 pixel area, (b) calculated object's background intensity of the same area.

region outline, which includes all the nearby located objects. All the pixels that do not belong to the expanded breast area are set to zero, resulting in the removal of background, marks and artifacts. The artifacts located at the boundary of the breast region, at the chest side, forming a thick line are eliminated too. The minimum rectangle containing the breast region is automatically drawn and it is used in the subsequent processing stages.

At first, the mammogram is considered as a three-dimensional plot with the third axis (z) corresponding to the intensity of each pixel (Fig. 3a). The whole image is split into 30×30 sub-regions and, using bicubic interpolation, a second plot is obtained representing the intensity level of the local background (Fig. 3b). The interpolated image is subtracted from the original mammogram producing a third image with each pixel value providing the difference between the original and local background pixel values. The pixels with positive

values are identified and a percentage of them (5%) with the highest values is selected producing a binary image and also specifying a threshold value (the lowest value among the selected pixels). The reason for the above selection is that the objects of interest (microcalcifications) are characterized by higher intensity compared to their background. In a typical image, the number of selected pixels is quite large and in subsequent processing a fraction of them will be removed. If the amount of the selected pixels is lower than 10% of the total number of pixels of the cropped mammogram, the pixels with intensity higher than half of the previously specified threshold are added. In such a way an adequate number of pixels are included in the obtained binary image (A). This case occurs when the mammogram exhibits very low contrast usually due to erroneous exposure conditions.

Next a contrast enhancement filter is applied with 9×9 kernel having central element equal to 80 and all the other elements equal to -1 [38,43]. Five percent of the pixels having the highest intensity are selected, producing a second binary image (B). The outcome of the segmentation module is an image produced by the logical summation (AND) of the two binary images A and B. It contains the pixels that have high intensity values and, at the same time, quite high intensity values in comparison with the background intensity of their local neighbourhood (Fig. 4).

2.3. Regions of interest specification

In the segmented image obtained in the previous stage, neighbouring pixels with connectivity of eight are grouped together to create possible microcalcification objects. Objects containing one or two pixels are rejected since they are considered as artifacts [10]. Since the diagnostic information is based on the existence of groups of objects, individual objects (possibly artifacts) should be removed. The elimination of these artifacts is achieved through the use of morphological operators. The application of the erosion operator (with structure element a 3×3 kernel of unit value) results in the removal of all objects apart from those that have at least one innermost pixel that is not part of their boundary. In this way, only inner pixels that belong to large objects remain. These pixels correspond to the centres of ROIs, which are generated using the dilation operator with a 3×3 structure element of unit value. The dilation is repeated 50 times in order to produce a ROI with sufficient area around the object.

The smallest possible size of ROI is 101×101 pixels and appears when the central pixel of an object is isolated and no other central pixel is located at a distance smaller than 100 pixels (which is the maximum allowed distance in order for two distinct objects to belong in the same ROI). This selection takes into account the mean distance among microcalcifications in a cluster [4]. A ROI that is not of minimum size is considered as having been generated from a group of objects located in the same neighbourhood. In such case, two or more ROIs will be combined and a new enlarged ROI will be generated containing more than two of the original objects. Based on the above methodology, several ROIs are identified in the mammogram and each of them is a candidate for being a true cluster of microcalcifications.

The set of ROIs is partitionned in two groups depending on their area. The first group contains those ROIs with areas lower than 20,000 pixels ($2 \times 100 \times 100$), which is a reliable threshold value discriminating ROIs that are generated from individual objects.

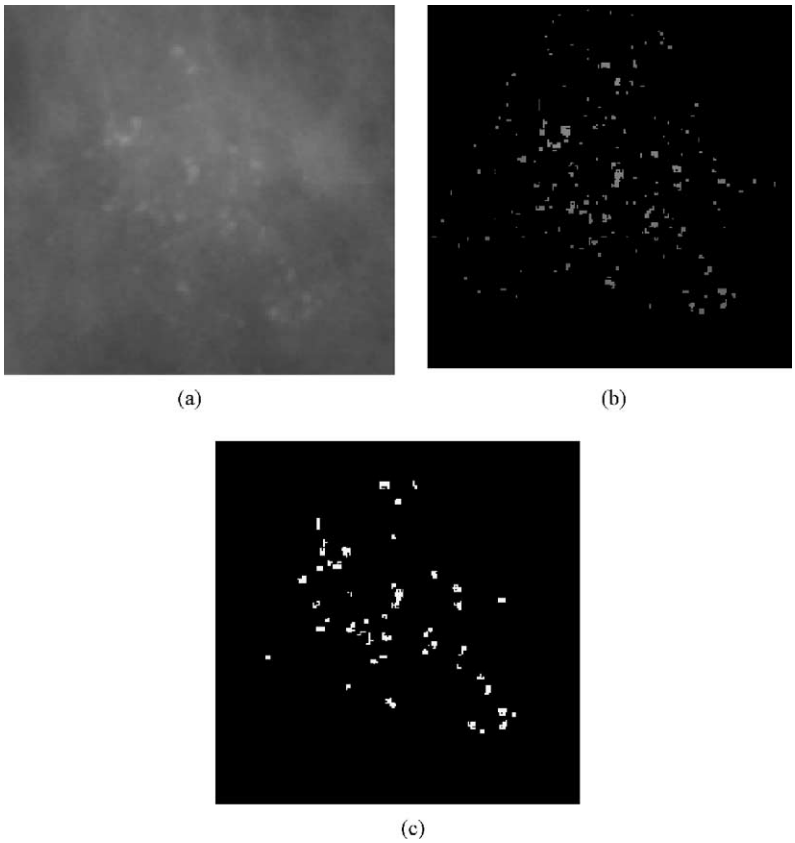

Fig. 4. (a) A part of a mammogram (original image), (b) the output of the segmentation component and (c) the binary image after small object elimination.

The second group contains the remaining ROIs which contain at least two nearby objects. This discrimination of ROIs defines a novel feature that will be used at the classification stage.

The existence of an individual object close to a ROI might be a problem in some cases. To resolve it a second dilation process is applied on the previous image, but only to the set of larger ROIs, using a 3×3 structure element in a 50-cycles repeated procedure. The resulting image contains usually one or two ROIs that consisting of at least one large ROI and perhaps some small ROIs (of the previous image) that are close to the large one.

The above procedure constitutes an attempt to identify groups of objects that are candidates for microcalcification clusters. The medical definition of clustered microcalcifications is the presence of more than three microcalcifications in 1 cm^2 area [21]. The minimum area of a ROI is 101×101 pixels or 0.25 cm^2 or a quarter of the area that is reported at the medical rule. Keeping a relative proportion, the criterion is altered to two or more. Since this rule can be used for the reduction of false positive detected ROIs, all the regions that include less than two objects are eliminated. Using the above morphological analysis, a number of ROIs is specified.

2.4. Classification module

The objective of the classification module is to categorize the specified ROIs as true microcalcification clusters or not. The large number of false positive clusters that are identified by the segmentation process makes the characterization task difficult. In order to specify the features that will be used as inputs to the classification system, at first 54 features are identified and computed characterizing either an individual microcalcification (object) or a group of them in a specific ROI. Those features fall into three categories related with the intensity, shape and texture properties of each object. It should be noted that does not exit any particular feature indicating the relation of each ROI with the mammographic image of origin since each area is treated separately. The group features are computed as the mean value of the five largest objects included in a ROI. The term largest refers to the number of pixels each object is composed, in the binary-segmented image. The selection of the five largest microcalcifications is made since a very small microcalcification does not have enough pixels for reliable feature value computation [4].

An important feature that contributes significantly to the classification accuracy of the proposed system is whether a given ROI lies in the same neighbourhood with a larger ROI. Despite the fact that this feature is not related with some established medical rule, the discrimination performance of this feature is high. The latter is a consequence of the how that the ROIs are generated. The existence of a small ROI near a large one introduces increased possibility for it either to be a true cluster or part of the large one. In either case, the inclusion of this feature increases the detection performance of the system.

Since the number of the computed features is quite large and their discriminative power varies, a feature validation together with feature selection procedure is applied. The receiver operating characteristic (ROC) curve is plotted for each feature and the area A_z under the ROC curve is computed. Features with the highest A_z are selected, resulting in a set of 22 features (Table 1). It must be noted that most of the selected features correspond to the mammographic characteristics that radiologists examine during a diagnostic procedure such as shape, density, size, distribution of the examined group or individual objects (Table 1) [53].

In the next step of the classification module the selected features are fed into a hybrid intelligent classification system, which consists of two components (Fig. 5): a rule-based and a neural network component. The rule construction procedure consists of the feature identification step as well as the selection of the particular threshold value for each feature. First, visualization of all the calculated features in two-dimensional plots, in pairs, has been employed for the selection of suitable feature threshold values that lead to the categorization of a remarkable number of ROIs. For every feature, several threshold values are examined in the range of values corresponding to that feature. For each threshold value, the number of ROIs below and above the threshold value is recorded. The ratio of the number of ROIs that belong to a specific class (normal or pathological) over the total number of the ROIs that belong to the same class should be more than 6%. In addition, the number of the false negative ROIs must be equal or less than one.

In the Nijmegen database the rule-based sub-system contains three rules employing a single feature and one rule with two features. The employed features are the standard deviation of the microcalcifications' intensity in a cluster, the mean eccentricity value in a

Table 1
Main features for cluster categorization

Microcalcification (MC) cluster classification features	Radiologists characterization features
Number of MCs in cluster	Cluster elements (separable/countable)
Cluster area	Cluster size
Mean MC area	MCs size
STD of MCs area	Shape of elements within cluster
Mean MC compactness	Shape of elements within cluster
Mean MC elongation	Shape of elements within cluster
STD of MC elongation	Shape of elements within cluster
STD of MC intensity	Density of calcifications
Mean MC background intensity	Density of calcifications
Mean contrast	Contrast of calcifications
Cluster eccentricity	Shape of cluster
Mean distance from cluster centroid	Calcification distribution
Neighbouring with a larger cluster	Cluster distribution
Cluster entropy	Calcification's distribution
Spreading of MCs in cluster	Calcification's distribution
Cluster elongation	Cluster's shape
Mean local MC background	Density of calcifications
Mean MCs intensity	Density of calcifications
STD of MC compactness	Shape of elements within cluster
STD of distances from cluster centroid	Calcification's distribution
Area of the cluster convex hull	Shape of cluster
The length of the cluster convex hull	Shape of cluster

cluster, the entropy of a cluster and the standard deviation of the distances of microcalcifications from the cluster centre and the average microcalcifications' area in a cluster. All the rules, other than the cluster entropy, contribute to the removal of false positive clusters.

In the MIAS database four rules have been obtained each one employing one feature. The features with the higher discriminative capability are the area of a cluster, the average background intensity in each cluster, the highest entropy value of the clusters in each image (relative entropy value) and the existence of large area clusters in the neighbourhood of a cluster. All the features, apart from the second one, contribute to the classification of abnormal cases. It must be noted that the cluster entropy is a common feature in both datasets, something that underlines the importance of this particular feature.

The ROIs that have been classified by the rule-based system are easily identified regions that are subsequently removed from the set that is used for training and testing of the neural network component. The remaining ROIs constitute the dataset that will be used for the construction of the neural network. The input feature vector of the latter contains the total number of features (22 features per ROI) and includes those features used by the rule-based system. The two components (rule-based and neural network) are sequentially applied in the classification scheme. Only if a ROI remains uncharacterized by the rule-based system, it is subsequently fed to the neural network module for characterization.

The neural network (Fig. 6) that is used for ROI characterisation is a feedforward neural network with sigmoid hidden nodes (Multiplayer Perceptron—MLP). In order to select an appropriate architecture (number of hidden layers and hidden nodes per layer) several

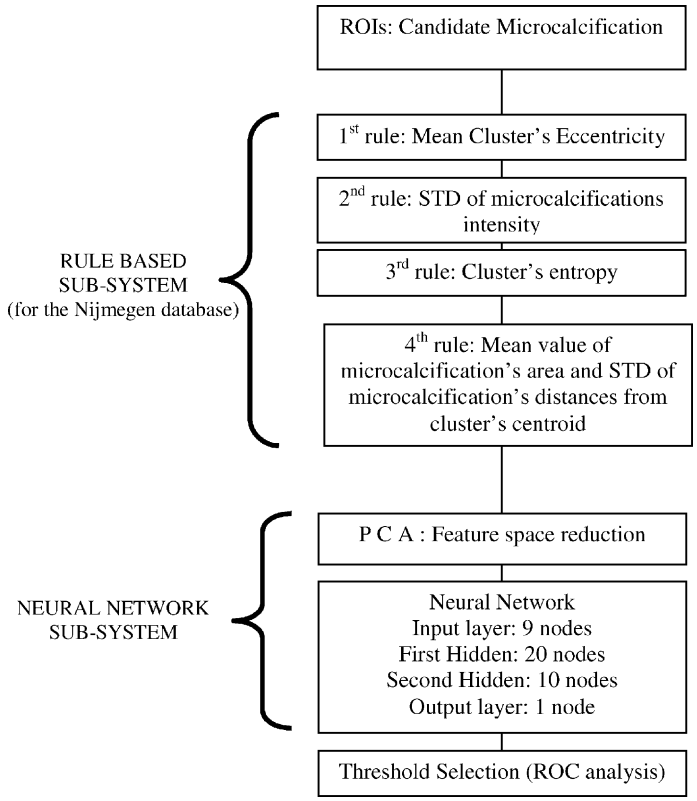

Fig. 5. The hybrid classification system.

networks were tested with one or two hidden layers and different number of hidden nodes [36]. In order to reduce the dimensionality of the input vector, a PCA was applied to eliminate the features that contribute less than 3% to the total variation of the data set. The PCA procedure transforms each 22-dimensional feature vector into a 9-dimensional feature vector that will constitute the input to the neural network. The components of each new feature vector are normalized to zero mean and unit variance.

Several training algorithms were implemented and tested: gradient descent methods, resilient backpropagation, conjugate gradient methods, and quasi Newton methods [5]. The best results are obtained using a quasi-Newton method, and more specifically, the one-step secant (OSS) algorithm [2].

To assess the performance of several architectures and training algorithms the two-fold cross validation method was employed. According to this procedure, the dataset is randomly divided into two subsets where the number of positive and negative cases in each subset is approximately equal. In a first experiment the training set corresponds to the first subset and the test set to the second one. In a second experiment the first subset corresponds to the test set and the second to the training set. The performance is calculated as the average test set performance in the two experiments.

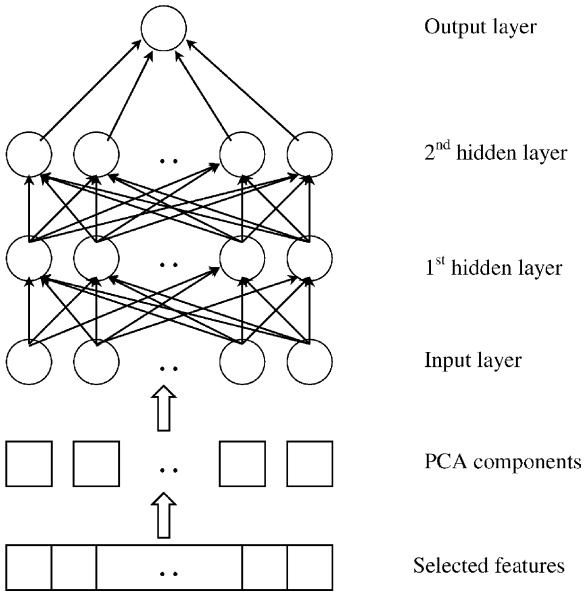

Fig. 6. The neural network sub-system.

To train a neural network we assume that positive (true) ROIs correspond to unit output while negative ones to zero output. Training is terminated either when the training error is less than a very small given value (10^{-5}) or when 2000 iterations have been performed.

During testing a threshold value is needed to classify an input case as true or false cluster based on the output value which ranges from 0 to 1. As the threshold value decreases from 1 to 0, a larger number of true positive cases is correctly characterised with an obvious increase of false positive cases. The network performance is measured using the area A_z under a ROC curve generated by plotting the true positive fraction (sensitivity) against the false positive fraction (1-specificity) of the cases for various threshold values [29]. Alternatively, the free receiver operating characteristic (FROC) curve may be used which considers the number of false positive clusters per image instead of the specificity value [8].

The finally selected network (the one with the best cross-validation performance) for the Nijmegen database has an input layer with nine nodes, two hidden layers with 20 and 10 sigmoid nodes, respectively, and an output layer with one sigmoid node. For the MIAS database, the same neural network architecture has also been used since its performance to the particular dataset was at the highest level, using the cross-validation model for training and testing.

3. Results

3.1. Nijmegen database

The segmentation process results in 446 candidate ROIs from which 115 are true. The difference is due to the fact that our system in some cases identifies two or more ROIs

contained in a single annotated ROI. Using the rule-based sub-system 215 ROIs are classified. Most of them are normal ROIs corresponding to artifacts and blood vessel type objects.The 41 ROIs are true positive (TP), 167 are true negative (TN), 5 are false positive (FP) and 2 are false negative (FN) cases. Using the rule-based sub-system, 48% of the cases are characterised corresponding to 39% of abnormal and 49% of normal cases. Concerning the relative contribution of the rules used by the rule-based sub-system, the first rule characterizes the 26.5%, the second 9.7%, the third 21.4% and the forth 42.3% of the total number of ROIs that have been classified by the rule-based system.

The performance of the hybrid system using two-fold cross-validation at the sensitivity level 0.90 is 1.8 false positive clusters per image. The use of the neural network results in 54 TP, 89 TN, 9 FN and 70 FP cases and the hybrid system results in 95 TP, 256 TN, 11 FN and 75 FP cases. The A_z value of the hybrid system is 0.912 (Fig. 7). The performance is high for a wide range of sensitivity. At the sensitivity level 0.79, the specificity is 0.86 (or 1.15 false positive clusters per image) and at the sensitivity level 0.96 the number of false positive clusters per image is 3.28 (Table 2).

It is also common in the related literature to report performance results where the complete dataset (including both the training and the test set) is used as a test set (for measuring the performance of the method), due to the limited number of available cases. In such case, as expected, the system performance is greatly improved. More specifically, at the specificity level of 1.18 false positive clusters per image, the sensitivity value obtained is 0.96 (Table 2). The A_z value is equal to 0.956 (Fig. 7).

In order to assess the performance benefits from the use of the hybrid system, we have conducted experiments to compare the method against the case where a single neural network is used a classification component instead of the hybrid system. More specifically

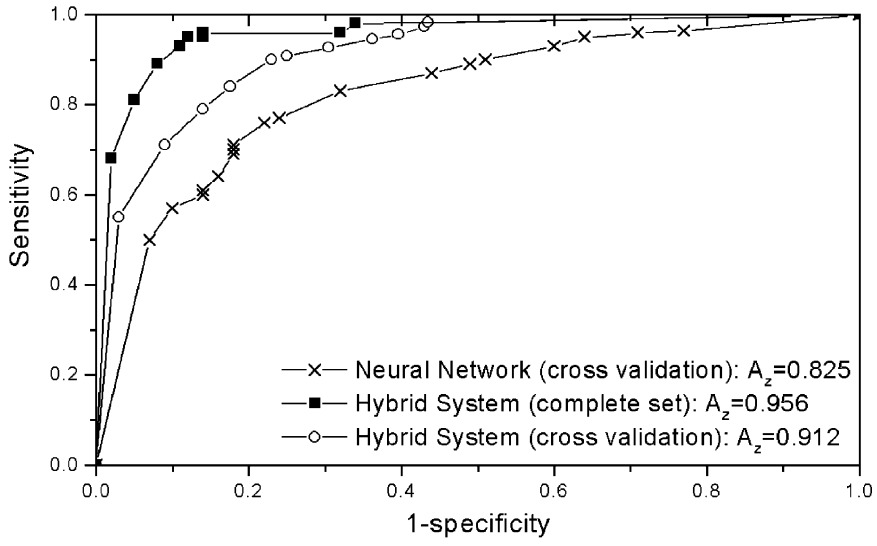

Fig. 7. ROC curves obtained with the neural network and hybrid intelligent system using the Nijmegen database with complete dataset and cross validation.

Table 2
Performance of the proposed hybrid intelligent system tested for both source databases using different training schemes (cross validation/complete set)[a]

Source database	Network training and evaluation procedures	Sensitivity	Specificity	False positive clusters/image
Nijmegen	Cross validation	0.79	0.86	1.15
		0.84	0.82	1.45
		0.90	0.77	1.80
		0.96	0.60	3.28
	Complete set	0.89	0.92	0.68
		0.93	0.89	0.95
		0.96	0.86	1.18
	No rules employment—complete set	0.83	0.68	2.70
		0.90	0.49	4.25
		0.96	0.36	5.90
MIAS	Cross validation	0.84	0.90	0.80
		0.91	0.86	1.15
		0.94	0.66	2.70
		0.97	0.55	3.55
	Complete set	0.91	0.92	0.65
		0.94	0.92	0.65
		0.97	0.88	0.95
	No rules employment—complete set	0.82	0.68	2.50
		0.91	0.62	3.00
		0.97	0.58	3.30

[a] The performance of a single network without rules employment is presented as well.

the same network architecture was used but the rule-based component was left out. The performance of this system at a sensitivity level of 0.90 was 4.25 false positive clusters per image (Table 2). The total number of ROIs has been classified as 102 TP, 163 TN, 170 FP and 11 FN cases. The A_z area under ROC curve area is 0.825 (Fig. 7). Obviously these results are inferior to those obtained with the proposed hybrid classification system.

3.2. The MIAS database

The MIAS database contains 20 digitized films that include microcalcification clusters. The total number of annotated clusters is 25. The segmentation process results in 193 candidate ROIs from which 34 are true. The rule-based sub-system characterises 116 ROIs corresponding to 25 TP, 79 TN, 12 FP, and zero FN cases. The percentage of ROIs classified by the rule-based sub-system is 73% of the abnormal and 50% of the normal cases. Concerning the relative contribution of the rules used by the rule-based sub-system, the first rule characterizes the 21.5%, the second 68.1%, the third 4.3% and the forth 6.0% of the total number of ROIs that have been classified by the rule-based system.

The performance of the hybrid system using two-fold cross-validation at the sensitivity level 0.91 is 1.15 false positive clusters per image. The use of neural network at the

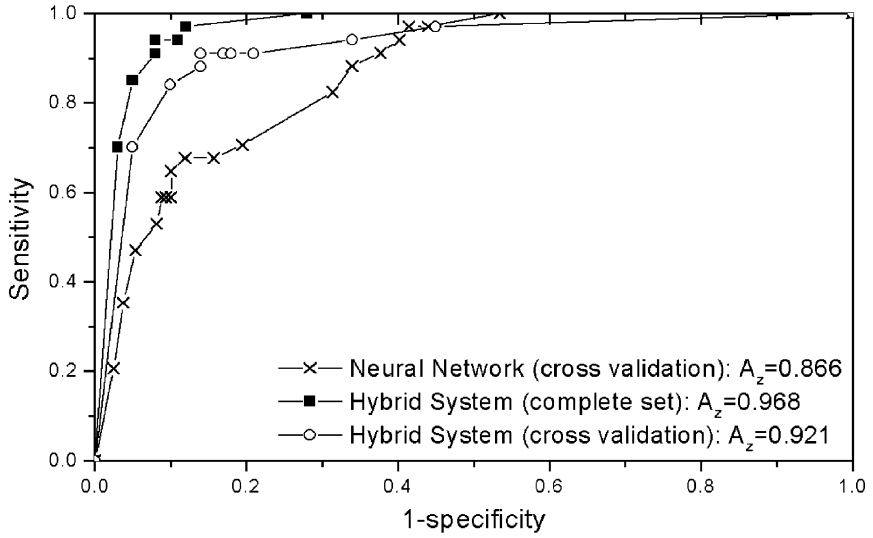

Fig. 8. ROC curves obtained with the neural network and hybrid intelligent system using the MIAS database with complete dataset and cross validation.

remaining cases results in 4 TP, 57 TN, 3 FN and 11 FP cases and the entire hybrid system results in 29 TP, 136 TN, 3 FN and 23 FP cases. The A_z value of the hybrid system is 0.921 (Fig. 8). At sensitivity level 0.84, the specificity value is 0.90 (or 0.8 false positive clusters per image) and at the sensitivity level 0.94 the specificity is 2.70 false positive clusters per image (Table 2).

If the performance of the hybrid neural network is measured in the complete data set (training and testing) the sensitivity improves to 0.94 at the specificity level of 0.92 or 0.65 false positive clusters per image (Table 2). The area A_z in this case is 0.968 (Fig. 8). When a single neural network is solely used instead of the hybrid system, the performance deteriorates significantly: at sensitivity level 0.91, the specificity is 3.0 false positive clusters per image (Table 2). The total number of ROIs has been classified as 31 TP, 99 TN, 60 FP and 3 FN cases. The A_z value is 0.866 (Fig. 8).

4. Discussion

The proposed system exhibits high performance in the detection of microcalcification clusters since it is able to identify more that 90% of the total number of clusters with a rather small number of false positive findings. The utilization of a hybrid intelligent classification component improves the performance of the system. A reduction of the false positive clusters cases is achieved without any cost for the sensitivity of the system. The absence of user adjustable parameters in the segmentation process ensures that it is straight forward to apply the method to other mammographic datasets.

Several techniques have been reported in the literature for the detection of micro-calcification clusters using various methodologies. The performance of the proposed

method is comparable with the reported results. For the Nijmegen dataset, Meersman et al. [28] using a neural network approach reported a sensitivity level 0.84 and two false positive clusters per image. Using an adaptive filtering method Gurgan et al. [18] achieved a sensitivity 1.0 with 2.3 false clusters per image. Yu [58] obtained a sensitivity 0.9 with 0.5 false clusters per image and Bazzani et al. [3] obtained a sensitivity 0.94 with 0.6 false clusters per image using an evaluation procedure that incorporated the training set. Netsch and Peitgen [32] reported sensitivity 0.84 with one false positive cluster per image. Kassemeijer [20] obtained sensitivity 0.90 with 0.8 false positive clusters per image using an extended Nijmegen dataset (containing 25 additional images).

For the MIAS dataset, Diahi et al. [14] proposed a neural network system that is fed with predefined ROIs providing a detection performance of 0.95 for the whole dataset. Norhayati et al. [35], using triple-ring filter analysis, reported a sensitivity of 0.96 with 1.8 false positive clusters per image using an extended dataset containing 24 additional images without findings.

Concerning our methodology several comments can be made. First, the cluster detection procedure is greatly accelerated compared to methodologies to which a specific rule is applied to every pixel of the whole image in order for a specific region to be selected. Instead of using sliding windows and applying a medical rule in each of them (as happens in several other systems), our methodology detects the center of each ROI and then a window is drawn around it. Therefore, search for clusters based on the application of the medical rule is performed for a limited number of windows. In addition, ROIs are detected more accurately including a larger number of microcalcifications, which are described better since each one is consisted of more pixels.

At the feature extraction step, 54 features are initially computed. Some of them are extracted from individual microcalcifications and others (group features) are the average values of the microcalcification cluster features. The 22 features are kept after ROC analysis for each feature. Most of these features are also taken into account by radiologists during the visual mammogram interpretation. In addition, we introduce a new significant feature consisting of the existence of large ROIs in the neighborhood of a given ROI. The employed number of features is further reduced using PCA whose threshold has been experimentally specified. The use of PCA in conjunction with the hybrid intelligent system constitutes a novel characteristic of the method.

The hybrid system performs better than the single schemes. The limited classification ability of a rule-based system is expected since the microcalcifications defined feature space has no absolute class borders. The neural network classifier exhibits significantly better performance than the rule-based scheme. But, even with the use of neural network classifier, the performance score is lower than the proposed hybrids' methodology. This is due to the network inability to characterize some ''unusual'' cases. Those cases probably appear as atypical combinations in the feature space. The neural network does not easily simulate such behaviors. Their removal by the rule-based system, contribute to the performance improvement of the neural network component.

The proposed method detects microcalcification clusters in digitized mammograms eliminating the false groups of objects having similarities with the true clusters. In both datasets that have been used for testing, the performance of our method is high for a large range of sensitivities. This is an indication of robustness giving expectations for

satisfactory performance using other datasets. The employment of different rules in the hybrid system, for the two datasets, is due to the fact that different digitization procedures had been followed in each dataset. A larger homogeneous dataset must utilize a uniform rule sub-system. However, it must be noted that the method performs well despite the fact that no processing for identification of line structures is included. The latter constitutes a reasonable direction for further improvement of our method performance.

Additional patient features other than those obtained from the image such as family history, age, etc. might be included to improve the diagnostic value of our system. The next step of our work is to use hybrid systems to perform classification of identified clusters according to their likelihood of malignancy.

5. Conclusions

A hybrid intelligent system has been developed for the identification of microcalcification clusters in digitized mammograms. The method employs two components: a rule-based and a neural network sub-system. We tested our system in the Nijmegen and the MIAS mammographic databases with satisfactory results. The achieved classification specificity is 1.80 and 1.15 false positive clusters per image, for the Nijmegen and MIAS dataset, respectively, at the sensitivity level of about 0.91.

Although the achieved performance is satisfactory further studies should be done in the elimination of falsely detected objects with line structure that are in most cases not originating from parenchymal microcalcifications. Further testing has to be performed concerning the use of other databases as well as of original mammograms obtained from the clinical routine of the collaborating hospitals or screening population projects. Finally, future work will also be directed towards the construction of a classification system that will perform discrimination between benign and malignant microcalcification clusters.

Acknowledgements

The present work is partially supported by the Greek General Secretariat of Research and Technology as part of the project EPET II—PENED: Analysis and Design of Classification Methods for Computerized-Aided Detection of Breast Cancer from Radiological Data. 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] Alexander FE, Anderson TJ, Brown HK, Forrest AP, Hepburn W, Kirkpatrick AE, et al. Fourteen years of follow-up from the Edinburgh randomised trial of breast-cancer screening. Lancet 1999;353:1093–108.

[2] Battiti R. First and second order methods for learning: between steepest descent and Newton's method. Neural Comput 1992;4(2):141–66.

[3] Bazzani A, Bevilacqua A, Bollini D, Brancaccio R, Campanini R, Lanconelli N, et al. Automated detection of clustered microcalcifications in digital mammograms using an SVM classifier. In: Proceedings of the European Symposium on Artificial Neural Networks, 2000;195–200.
[4] Betal D, Roberts N, Whitehouse GH. Segmentation and numeral analysis of microcalcifications on mammograms using mathematical morphology. Br J Radiol 1997;70(9):903–17.
[5] Bishop CM. Neural networks for pattern recognition. Oxford: Oxford University Press, 1996.
[6] Bottema MJ, Slavotinek JP. Detection and classification of lobular and DCIS (small cell) microcalcifications in digital mammograms. Pattern Recog Lett 2000;21:1209–14.
[7] Brzakovic D, Brzakovic P, Neskovic M. An approach to automated screening of mammograms. SPIE Biomed Image Processing Biomed Visual 1993;2167:868–86.
[8] Chakraborty DP, Winter LH. Free response methodology: alternate analysis and a new observer-performance experiment. Radiology 1990;174(3):873–81.
[9] Chan HP, Doi K, Galhotra S, Vyborny CJ, MacMahon H, Jokich PM. Image feature analysis and computer aided diagnosis in digital radiography. 1. Automated detection of microcalcifications in mammography. Med Phys 1987;14:538–48.
[10] Chan HP, Lo SCB, Sahiner B, Lam KL, Helvie MA. Computer-aided detection of mammographic microcalcifications: pattern recognition with an artificial neural network. Med Phys 1995;22(10):1555–67.
[11] Cheng H, Lui YM, Feiimanis RI. A novel approach to microcalcification detection using fuzzy logic techniques. IEEE Trans Med Imag 1998;17(6):442–50.
[12] Davies DH, Dance DR. Automated computer detection of clustered calcifications in digital mammograms. Phys Med and Biol 1990;35(8):1111–8.
[13] Dengler J, Behrens S, Desage JF. Segmentation of microcalcifications in mammograms. IEEE Trans Med Imag 1993;12:634–42.
[14] Diahi JG, Frouge C, Giron A, Fertil B. Artificial neural networks for detection of breast cancer in mammography. In: Proceedings of the 3rd International Workshop on Digital Mammography, Chicago, U.S.A., 9–12 June 1996. p. 329–34.
[15] Fracheboud J, de Koning HJ, Beemsterboer PM, Boer R, Hendriks JH, Verbeek AL, et al. Nation-wide breast cancer screening in the Netherlands: results of initial and subsequent screening 1990–1995. Int J Cancer 1998;75:694–8.
[16] Gonzalez RC, Woods RE. Digital image processing. Reading (MA): Addison-Wesley, 1993.
[17] Gotzsche PC, Olsen O. Is screening for breast cancer with mammography justifiable? Lancet 2000;355:129–34.
[18] Gurgan MN, Yardimci Y, Cetin AE. Microcalcification detection using adaptive filtering and gaussianity tests. In: Proceedings of the 4th International Workshop on Digital Mammography, Nijmegen, The Netherlands, 7–10 June 1998. p. 157–64.
[19] Hakama M, Pukkala E, Heikkila M, Kallio M. Effectiveness of the public health policy for breast cancer screening in Finland: population based cohort study. Br M J 1997;314:864–7.
[20] Kassemeijer N. Adaptive noise equalization and recognition of microcalcifications in mammography. Inter J Pattern Recog Artif Intel 1993;7:1357–76.
[21] Kopans DB. Breast imaging. Philadelphia, J.B. Lippincott, 1989.
[22] Lado MJ, Tahoces PG, Mendez AJ, Souto M, Vidal JJ. A wavelet-based algorithm for detecting clustered microcalcifications in digital mammograms. Med Phys 1999;26(7):1294–305.
[23] Laine AF, Schuler S, Fan J, Huda W. Mamographic feature enhancement by multiscale analysis. IEEE Trans Med Imag 1994;1905:725–38.
[24] 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.
[25] Lefebvre F, Benali H, Gilles R, Kahn E, Di Paola R. A fractal approach to the segmentation of microcalcifications in digital mammograms. Med Phys 1995;22(4):381–90.
[26] Li H, Liu KJR, Lo SCB. Fractal modelling and segmentation for the enhancement of microcalcifications in digital mammograms. IEEE Trans Med Imag 1997;16(6):785–98.
[27] Lo SC, Chan HP, Lin JS, Li H, Freedman MT, Mun SK. Artificial convolution neural network for medical image pattern recognition. Neural Networks 1995;8(7/8):1201–14.

[28] Meersman D, Scheunders P, Van Dyck D. Detection of microcalcifications using neural networks. In: Proceedings of the 3rd International Workshop on Digital Mammography, 1996. p. 287–90.
[29] Metz CE. ROC methodology in radiologic imaging. Invest Radiol 1986;21(9):720–33.
[30] Miller AB, Baines CJ, To T, Wall C. Canadian national breast screening study. 1. Breast cancer detection and death rates among women aged 40–49 years. Can Med Assoc J 1992;147:1459–76.
[31] Miller AB, Baines CJ, To T, Wall C. Canadian national breast screening study. 2. Breast cancer detection and death rates among women aged 50 to 59 years. Can Med Assoc J 1992;147:1477–88.
[32] Netsch T, Peitgen HO. Scale-space signatures for the detection of clustered microcalcifications in digital mammograms. IEEE Trans Med Imag 1999;18(9):774–86.
[33] Nigel RH, Nishikawa RM, Papaioannou J, Doi K. Analysis of methods for reducing false positives in the automated detection of clustered microcalcifications in mammograms. Med Phys 1998;25(8):1502–6.
[34] Nishikawa RM, Schmidt RA, Papaioannou J, Onsis R, Haldemann Heusler RA, Giger ML, et al. Performance of a prototype clinical intelligent mammography workstation, '96. In: Doi K, Giger ML, Nishikawa RM, Schmith RA, editors. Digital mammography. Amsterdam: Elsevier, 1996. p. 93–6.
[35] Norhayati I, Hiroshi F, Takeshi H, Tokiko E. Automated detection of clustered microcalcifications on mammograms: CAD system application to MIAS database. Phys Med Biol 1997;42:2577–89.
[36] Papadopoulos A, Fotiadis DI, Likas A. A hybrid neural method for microcalcification cluster detection in mammography. In: The Proceedings of 4th International Conference on Neural Networks and Expert Systems in Medicine and Healthcare, Milos Island, Greece, 20–22 June 2001. p. 90–6.
[37] Patrick EA, Moskowitz M, Mansukhani VT, Gruenstein EI. Expert learning system network for diagnosis of breast calcifications. Invest Radiol 1991;26:534–9.
[38] Pratt WK. Digital image processing. New York: Wiley/ Interscience, 1991.
[39] Qian W, Clarke LP, Kallergi M, Li H, Velthuizen R, Clark RA, et al. Tree-structured nonlinear filter and wavelet transform for microcalcification segmentation in mammography. SPIE Biomed Image Process Biomed Visual 1993;12(4):634–42.
[40] Quinn M, Allen E. Changes in incidence of and mortality from breast cancer in England and Wales since introduction of screening. Br M J 1995;311:1391–5.
[41] Roehrig J, Doi T, Hasegawa A, Hunt B, Marshall J, Romsdahl H, et al. Clinical results with R2 ImageChecker system, Digital mammography. In: Karssemeijer N, Thijssen M, Hendriks J, van Erning L, editors. Dordrecht: Kluwer academic Publishers; 1998. p. 395–400.
[42] Schmidt RA, Nishikawa RM. Digital screening mammography. Principles Practice Oncol 1994;8:1–16.
[43] 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.
[44] Smith RA. Epidemiology of breast cancer in a categorical course in physics: technical aspects of breast imaging (2nd ed). Oak Brook, IL: RSNA Publication, 1993. p. 21–33.
[45] Soni T, Zeidler JR, Ku WH. Performance evaluation of two-dimensiomal adaptive prediction filters for detection of small objects in image data. IEEE Trans Image Processing 1993;2(3):327–39.
[46] Strickland RN, Hahn HI. Wavelet transforms for detecting microcalcifications in mammography. IEEE Trans Med Imag 1996;15(2):218–28.
[47] Suckling J, Parker J, Dance D, Astley S, Hutt I, Boggis C, et al. The mammographic images analysis society digital mammogram database. Exerpta Medica 1994;1069:375–8.
[48] Tabar L, Duffy S. Malmo mammographic screening trial, Br. M. J. 7 298 6665 (1989) 48–9.
[49] Van den Akker-van Marle E, de Koning H, Boer R, van der Maas P. Reduction in breast cancer mortality due to the introduction of mass screening in The Netherlands: comparison with the United Kingdom. J Med Screen 1999;6:30–4.
[50] Veldkamp W, Karssemeijer N, Otten JDM, Hendriks JHCL. Automated classification of clustered microcalcifications into malignant and benign. Med Phys 2000;27(11):2600–8.
[51] Veldkamp W, Karssemeijer N. Automated classification of clustered microcalcifications in digital mammography. In Digital mammography 96, 1996, edited by Doi K, Giger ML, Nishikawa RM, Schmidt RA, Excerpt Medica, Amsterdam: Elsevier, 231–4.
[52] Wang TC, Karayiannis NB. Detection of microcalcifications in digital mammograms using wavelets. IEEE Trans Med Imag 1998;17(4):498–509.

[53] Whatmough P, Gale AG, Wilson A.R.M. Do radiologists agree on the importance of mammographic features? In: Proceedings of the 3rd International Workshop on Digital Mammography, 1996. p. 111–6.
[54] Wu Y, Doi K, Giger ML, Nishikawa RM. Computerized detection of clustered microcalcifications in digital mammograms: application of artificial neural networks. Med Phys 1992;19(3):555–60.
[55] 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.
[56] Yoshida H, Doi K, Nishikawa RM. Automated detection of clustered microcalcifications in digital mammograms using wavelet transform techniques. SPIE Image Processing 1994;2167:868–86.
[57] Yoshida H, Doi K, Nishikawa RM, Giger ML, Schmidt RA. An improved CAD scheme using wavelet transform for detection of clustered microcalcifications in digital mammograms. Acad Radiol 1996; 3:621–7.
[58] Yu S, Guan L. A CAD system for the automated detection of clustered microcalcifications in digitised mammogram films. IEEE Trans Med Imag 2000;19(2):115–26.
[59] Zhang W, Doi K, Giger ML, Wu Y, Nishikawa M, Schmidt RA. Computerized detection of clustered microcalcification in digital mammograms using a shift-invariant artificial neural network. Med Phys 1994;19:555–60.
[60] Zhao D. Rule-based morphological feature extraction of microcalcifications in mammograms. SPIE Med Imag 1993;1095:702–15.
[61] Zheng B, Qain W, Clarke LP. Digital mammography: Mixed feature neural network with spectral entropy decision for detection of microcalcifications. IEEE Trans Med Imag 1996;15(5):589–97.
[62] Zuckerman HC. The role of mammography in the diagnosis of breast cancer. In: Ariel IM, Cleary JB, editors. Breast cancer: diagnosis and treatment. New York: McGraw-Hill, 1987. p. 152–72.