

Feature Selection for the Classification of Both Individual and Clustered Microcalcifications in Digital Mammograms Using Genetic Algorithms

Rolando R. Hernández-Cisneros
ITESM-Center for Intelligent Systems
Av. Eugenio G. Sada 2501 Sur
Monterrey, NL 64849 Mexico
A00766380@itesm.mx

Hugo Terashima-Marín
ITESM-Center for Intelligent Systems
Av. Eugenio G. Sada 2501 Sur
Monterrey, NL 64849 Mexico
terashima@itesm.mx

ABSTRACT

Breast cancer is one of the main causes of death in women and early diagnosis is an important means to reduce the mortality rate. The presence of microcalcification clusters are primary indicators of early stages of malignant types of breast cancer and its detection is important to prevent the disease. This paper uses a procedure for the classification of microcalcification clusters in mammograms using sequential Difference of Gaussian filters (DoG) and a Genetic Algorithm (GA) for feature selection. We found that the use of Genetic Algorithms (GAs) for selecting the features from microcalcifications and microcalcification clusters that will be the inputs of a feedforward Neural Network (NN) results mainly in improvements in overall accuracy, sensitivity and specificity of the classification.

Categories and Subject Descriptors

I.2.6 [Artificial Intelligence]: Learning—*Connectionism and Neural Nets*

General Terms

algorithms, experimentation

Keywords

microcalcifications classification, feature selection, genetic algorithms, evolutionary neural networks, mammography, computer-aided diagnosis

1. INTRODUCTION

Breast cancer is one of the main causes of death in women and early diagnosis is an important means to reduce the mortality rate. Mammography is one of the most common techniques for breast cancer diagnosis, and microcalcifications are one among several types of objects that can be

detected in a mammogram. Microcalcifications are calcium accumulations typically 100 microns to several mm in diameter, and they sometimes are early indicators of the presence of breast cancer. Microcalcification clusters are groups of three or more microcalcifications that usually appear in areas smaller than 1 cm^2 , with a probability of becoming a malignant lesion.

However, the predictive value of mammograms is relatively low, compared to biopsy. This low sensitivity [6] is caused by the low contrast between the cancerous tissue and the normal parenchymal tissue, the small size of microcalcifications and possible deficiencies in the image digitalization process. The sensitivity may be improved having each mammogram checked by two or more radiologists, with the consequence of making the process inefficient by reducing the individual productivity of each specialist. A viable alternative is replacing one of the radiologists by a computer system, giving a second opinion [1], [18].

A computer system intended for microcalcification detection in mammograms may be based on several methods, like wavelets, fractal models, support vector machines, mathematical morphology, bayesian image analysis models, high order statistic, fuzzy logic, etc. The method we selected for this work is the Difference of Gaussian Filters (DoG). DoG filters are adequate for the noise-invariant and size-specific detection of spots, resulting in a DoG image. This DoG image represents the microcalcifications if a thresholding operation is applied to it. The use of DoG for detection of potential microcalcifications has been addressed successfully by Dengler, Behrens and Desaga [5] and Ochoa [14].

In [16], we previously developed a procedure that applies a sequence of Difference of Gaussian Filters, in order to maximize the amount of detected probable individual microcalcifications (signals) in the mammogram, which are later classified in order to detect if they are real microcalcifications or not. Finally, microcalcification clusters are identified and also classified in order to determine which ones are malignant and which ones are benign. Both classifiers were simple feedforward NNs.

Neural networks (NNs) have been successfully used for classification purposes in medical applications, including the classification of microcalcifications in digital mammograms. Unfortunately, for a NN to be successful in a particular domain, its architecture, training algorithm and the domain variables selected as inputs must be adequately chosen. Designing a NN architecture is a trial-and-error process; several

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MedGEC 2006, Seattle, WA, USA

Copyright 200X ACM X-XXXXX-XX-X/XX/XX ...\$5.00.

parameters must be tuned according to the training data when a training algorithm is chosen and, finally, a classification problem could involve too many variables (feature most of them not relevant at all for the classification procedure itself).

Genetic algorithms (GAs) may be used to address the problems mentioned above, helping to obtain more accurate NNs with better generalization abilities. GAs have been used for searching the optimal weight set of a NN, for designing its architecture, for finding its most adequate parameter set (number of neurons in the hidden layer(s), learning rate, etc.), among others tasks. Exhaustive reviews about evolutionary artificial neural networks (EANNs) have been presented by Yao [19] and Balakrishnan and Honavar [2]. In particular, this paper describes the use of GAs for selecting the most relevant features extracted from individual microcalcifications, and from microcalcification clusters, that will become the inputs of two simple feedforward NNs for the classification of microcalcifications and microcalcification clusters in digital mammograms respectively, expecting to improve its accuracy. We compare this approach to feature selection with the one we used in [16], which was based on a forward sequential search, sequentially adding inputs to the NN while its error decreases, and stopping when it starts to increase. The GA provides a broader, parallel search of the feature space, simultaneously managing a population of feature subsets.

The rest of this document is organized as follows. In the second section, the mammography database we selected for this study is described. In the third section, the methodology is discussed. The fourth section deals with the experiments and the main results of this work. Finally, in the fifth section, the conclusions are presented, and some comments about future work are also made.

2. THE MAMMOGRAPHY DATABASE

There are several mammography databases publicly available for research purposes, and the most known of them are the mammography database from the Mammographic Image Analysis Society (MIAS) [17], the Digital Database for Screening Mammography (DDSM) from the University of South Florida [9] and the Nijmegen digital mammography database.

The mammograms used in this project were provided by The Mammographic Image Analysis Society (MIAS). The MIAS is an organization of UK research groups interested in the understanding of mammograms. The MIAS database contains 322 images, all medio-lateral (MLO) view, digitized with a scanning microdensitometer (Joyce-Loebl, SCANDIG3) with resolutions of 50 microns/pixel and 200 microns/pixel. In this work, the images with a resolution of 200 microns/pixel were used. The data has been reviewed by a consultant radiologist and all the abnormalities have been identified and marked. The truth data consists of the location of the abnormality and the radius of a circle which encloses it.

The abnormalities represented in the database include calcifications, circumscribed masses, spiculated masses, ill-defined masses, architectural distortions and asymmetry. Several normal cases are also included. From the totality of the database, only 25 images contain microcalcifications. Among these 25 images, 13 cases are diagnosed as malignant and 12 as benign. Some related works have used this

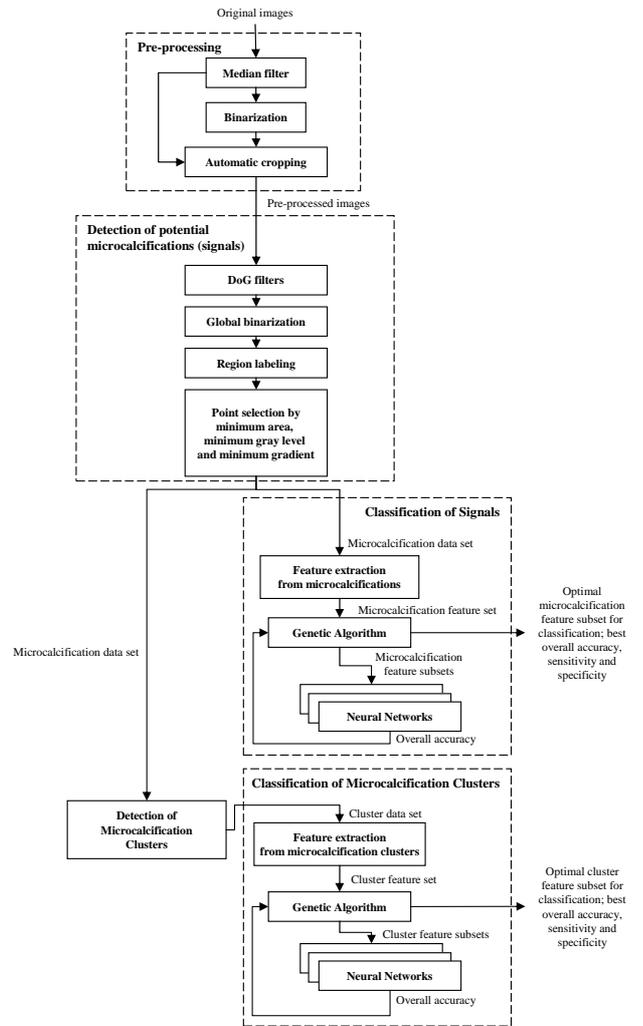


Figure 1: Diagram of the proposed procedure.

same database [4], [7], [10], [13].

3. METHODOLOGY

The general procedure receives digital mammograms as input, and it is conformed by five stages: pre-processing, detection of potential microcalcifications (signals), classification of signals into real microcalcifications, detection of microcalcification clusters and classification of microcalcification clusters into benign and malignant. The diagram of the proposed procedure is shown in Figure 1. This procedure is similar to the one we developed in [16], except that in this case, we experiment with the use of GAs for selecting the most relevant features that are expected to improve the accuracy of the classification process. As end-products of the procedure, we obtain two feedforward NNs for classifying microcalcifications and microcalcifications clusters respectively, which in this case, are products of the evolutionary approaches that are proposed. These networks have an adequate number of inputs for receiving the subsets of features that maximize the overall accuracy of the classification.

3.1 Pre-processing

This stage has the aim of eliminating those elements in the images that could interfere in the process of identifying microcalcifications. A secondary goal is to reduce the work area only to the relevant region that exactly contains the breast.

The procedure receives the original images as input. First, a 3x3 median filter is applied in order to eliminate the background noise, keeping the significant features of the images. The output is the filtered image. The size of the mask was chosen empirically, trying to avoid the loss of local detail.

Next, binary images are created from each filtered image, where each pixel in the binarized image is determined by a window centered in the corresponding pixel in the original image. If the mean gray level of the window is below a certain threshold (established empirically) a zero value is placed in the binary image; otherwise, a unitary value is placed. After observing the histograms of the mean gray level corresponding to windows of sizes 8x8, 16x16, 32x32 and 64x64, it was observed that the histograms corresponding to windows of sizes 8x8 and 16x16 were bimodal, and the visual selection of a threshold was easier. The selected size for this window was 16x16 and the threshold was set to 15. The manual selection of the binarization threshold is allowed by the small size of the data set used in this work. For larger and diverse mammography sets, an automated procedure for the selection of this threshold is proposed as future work.

In this stage, the binarized images are intended solely for helping the automatic cropping procedure to delete the background marks and the isolated regions, so the image will contain only the region of interest. The result of this stage is a smaller image, with less noise.

3.2 Detection of potential microcalcification (signals)

The main objective of this stage is to detect the mass centers of the potential microcalcifications in the image (signals). The pre-processed image of the previous stage is the input of this procedure. The optimized difference of two gaussian filters (DoG) is used for enhancing those regions containing bright points. A gaussian filter is obtained from a gaussian distribution, and when it is applied to an image, eliminates high frequency noise, acting like a smoothing filter. A DoG filter is built from two simple gaussian filters. These two smoothing filters must have different variances. When two images, obtained by separately applying each filter, are subtracted, then an image containing only the desired frequency range is obtained. The DoG filter is obtained from the difference of two gaussian functions, as it is shown in equation (1), where x and y are the coordinates of a pixel in the image, k is the height of the function and σ_1 and σ_2 are the standard deviations of the two gaussian filters that construct the DoG filter.

$$DoG(x, y) = k_1 e^{-(x^2+y^2)/2\sigma_1^2} - k_2 e^{-(x^2+y^2)/2\sigma_2^2} \quad (1)$$

The resultant image after applying a DoG filter is globally binarized, using an empirically determined threshold. In Figure 2, an example of the application of a DoG filter is shown. A region-labeling algorithm allows the identification of each one of the points (defined as high-contrast regions detected after the application of the DoG filters, which cannot be considered microcalcifications yet). Then, a segmentation algorithm extracts small 9x9 windows, containing the

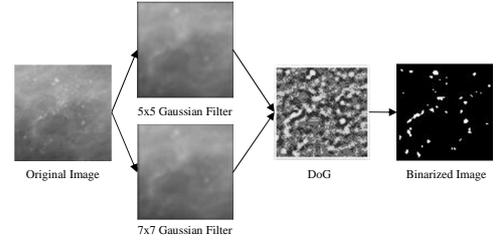


Figure 2: Example of the application of a DoG filter (5x5, 7x7).

region of interest whose centroid corresponds to the centroid of each point. The size of the windows is adequate for containing the signals, given that at the current resolution of 200 microns, the potentially malignant microcalcifications (whose diameter is typically 100 microns to several mm) have an area of 5x5 pixels on average [15].

In order to detect the greater possible amount of points, six gaussian filters of sizes 5x5, 7x7, 9x9, 11x11, 13x13 and 15x15 are combined, two at a time, to construct 15 DoG filters that are applied sequentially. Each one of the 15 DoG filters was applied 51 times, varying the binarization threshold. The points obtained by applying each filter are added to the points obtained by the previous one, deleting the repeated points. The same procedure is repeated with the points obtained by the remaining DoG filters. All of these points are passed later to three selection procedures.

These three selection methods are applied in order to transform a point into a signal (potential microcalcification). The first method performs selection according to the object area, choosing only the points with an area between a predefined minimum and a maximum. For this work, a minimum area of 1 pixel (0.0314 mm^2) and a maximum of 77 pixels (3.08 mm^2) were considered. The second method performs selection according to the gray level of the points. Studying the mean gray levels of pixels surrounding real identified microcalcifications, it was found they have values in the interval [102, 237] with a mean of 164. For this study, we set the minimum gray level for points to be selected to 100. Finally, the third selection method uses the gray gradient (or absolute contrast, the difference between the mean gray level of the point and the mean gray level of the background). Again, studying the mean gray gradient of point surrounding real identified microcalcifications, it was found they have values in the interval [3, 56] with a mean of 9.66. For this study, we set the minimum gray gradient for points to be selected to 3. The result of these three selection processes is a list of signals (potential microcalcifications) represented by their centroids.

3.3 Classification of Signals into Real Microcalcifications

The objective of this stage is to identify if an obtained signal corresponds to an individual microcalcification or not. With this in mind, a set of features are extracted from the signal, related to their contrast and shape. From each signal, 47 features are extracted: seven related to contrast, seven related to background contrast, three related to relative contrast, 20 related to shape, six related to the moments of the contour sequence and the first four Hu invariants. These features are shown in Table 1.

Table 1: Summary of features extracted from the signals (potential microcalcifications).

Signal contrast	Maximum gray level, minimum gray level, median gray level, mean gray level, standard deviation of the gray level, gray level skewness, gray level kurtosis
Background contrast	Background maximum gray level, background minimum gray level, background median gray level, background mean gray level, standard deviation of the background gray level, background gray level skewness, background gray level kurtosis.
Contrast	Absolute contrast, relative contrast, proportional contrast.
Shape features	Area, convex area, background area, perimeter, maximum diameter, minimum diameter, equivalent circular diameter, fiber length, fiber width, curl, circularity, roundness, elongation1, elongation2, eccentricity, aspect ratio, compactness1, compactness2, compactness3, solidity.
Contour sequence moments	CSM1, CSM2, CSM3, CSM4, mean radii, standard deviation of radii.
Invariant geometric moments	IM1, IM2, IM3, IM4.

In order to process signals and accurately classify the real microcalcifications, we decided to use NNs as classifiers. In the first section, we mentioned that one of the difficulties of working with conventional feedforward NNs is that a classification problem could involve too many variables (features), and most of them may not be relevant at all for the classification process itself. In [16], we used a method which consisted of two feature selection processes [11]: the first process attempts to delete the features that present high correlation with other features, and the second process uses a derivation of the forward sequential search algorithm, which is a sub-optimal search algorithm, adding features to a NN while the error is decreasing and stopping when it increases again. After these processes were applied, only three features were selected and used for classification: absolute contrast (the difference between the mean gray levels of the signal and its background), standard deviation of the gray level of the pixels that form the signal and the third moment of contour sequence. Moments of contour sequence are calculated using the signal centroid and the pixels in its perimeter, and are invariant to translation, rotation and scale transformations [8].

Expecting to achieve greater accuracy in the classification, we use a different method, this being based on a GA for selecting features. The chromosomes of the individuals in the GA contain 47 bits, one bit for each extracted feature, and the value of the bit determines whether that feature will be used in the classification or not [3]. The individuals are evaluated by constructing and training a feedforward NN (with a predetermined structure), and the number of inputs of this NN is determined by the subset of features to be included, coded in the chromosome. The accuracy of each network is used to determine the fitness of each individual. When the GA stops either because the generations limit has been reached or because improvements on the evaluation of the best individual has not been observed during five consecutive generations, we obtain the NN with the best performance in terms of the overall accuracy, and the subset of features that are relevant for the classification.

3.4 Detection of Microcalcification Clusters

During this stage, the microcalcification clusters are identified. The detection and posterior consideration of every

microcalcification cluster in the images may produce better results in a subsequent classification process, as shown in [16]. Because of this, an algorithm for locating microcalcification cluster regions where the quantity of microcalcifications per cm^2 (density) is higher, was developed. This algorithm keeps adding microcalcifications to their closest clusters at a reasonable distance until there are no more microcalcifications left or if the remaining ones are too distant for being considered as part of a cluster. Every detected cluster is then labeled.

3.5 Classification of Microcalcification Clusters into Benign and Malignant

This stage has the objective of classifying each cluster in one of two classes: benign or malignant. This information is provided by the MIAS database.

From every microcalcification cluster detected in the mammograms in the previous stage, a cluster feature set is extracted. The feature set is constituted by 30 features: 14 related to the shape of the cluster, six related to the area of the microcalcifications included in the cluster and ten related to the contrast of the microcalcifications in the cluster. These features are shown in Table 2.

In order to process microcalcification clusters and accurately classify them into benign or malignant, we decided again to use NNs as classifiers. In order to determine which ones of the 30 extracted features from the clusters are relevant for their classification, the first method we used in [16] was performed in this work. This method consisted of two feature selection processes [11]: the first process attempts to delete the features that present high correlation with other features, and the second process uses a derivation of the forward sequential search algorithm, which is a sub-optimal search algorithm, adding features to a NN while the error is decreasing and stopping when it increases again. After these processes were applied, only three cluster features were selected for the classification process: minimum diameter, minimum radius and mean radius of the clusters. The minimum diameter is the maximum distance that can exist between two microcalcifications within a cluster in such a way that the line connecting them is perpendicular to the maximum diameter, defined as the maximum distance between two microcalcifications in a cluster. The minimum

Table 2: Summary of features extracted from the microcalcification clusters.

Cluster shape	Number of calcifications, convex perimeter, convex area, compactness, microcalcification density, total radius, maximum radius, minimum radius, mean radius, standard deviation of radii, maximum diameter, minimum diameter mean of the distances between microcalcifications, standard deviation of the distances between microcalcifications.
Microcalcification Area	Total area of microcalcifications, mean area of microcalcifications, standard deviation of the area of microcalcifications, maximum area of the microcalcifications, minimum area of the microcalcifications, relative area.
Microcalcification Contrast	Total gray mean level of microcalcifications, mean of the mean gray levels of microcalcifications, standard deviation of the mean gray levels of microcalcifications, maximum mean gray level of microcalcifications, minimum mean gray level of microcalcifications, total absolute contrast, mean absolute contrast, standard deviation of the absolute contrast, maximum absolute contrast, minimum absolute contrast.

radius is the shortest of the radii connecting each microcalcification to the centroid of the cluster and the mean radius is the mean of these radii.

Trying to improve the accuracy in the classification of the microcalcification clusters, we also applied GAs to the feature selection task. The chromosomes of the individuals in this GA contain 30 bits, one bit for each extracted feature from the clusters, and the value of the bit determines whether that feature will be used in the classification or not. The individuals are evaluated by constructing and training a feedforward NN, where the number of inputs of the NN is determined by the subset of features to be included, which is coded in the chromosome. For solving nonlinearly separable problems, it is recommended at least one hidden layer in the network, and according to Kolmogorov’s theorem [12], and considering the number of inputs (n), the hidden layer contains $2n + 1$ neurons. The output layer has only one neuron. The accuracy of each network is used to determine the fitness of each cluster. When the GA is stopped either because the generations limit has been reached or because an improvement on the evaluation of the best individual has not been observed during five consecutive generations, we obtain the NN with the best performance (in terms of the overall accuracy) and the subset of cluster features that are the most relevant for the classification.

4. EXPERIMENTS AND RESULTS

All the programs were written in MATLAB Version 7.0.0.19920 (R14), and executed on a PC with a single 2.1 GHz Intel Pentium IV processor with 1 Gb of memory. The following subsections explain the experimentation and results of every stage of the study.

4.1 From Pre-processing to Feature Extraction

As it was mentioned in the previous section, only 25 images from the MIAS database contain microcalcifications. Among these 25 images, 13 cases are diagnosed as malignant and 12 as benign. Three images were discarded because the positions of the microcalcifications clusters, marked in the additional data that comes with the database, were outside the boundaries of the breast. So, only 22 images were finally used for this study, and they were passed through the pre-processing stage first (application of a 3x3 median filter, binarization and trimming).

In the second phase, six gaussian filters of sizes 5x5, 7x7, 9x9, 11x11, 13x13 and 15x15 were combined, two at a time, to construct 15 DoG filters that were applied sequentially. Each one of the 15 DoG filters was applied 51 times to the pre-processed images, varying the binarization threshold in the interval $[0, 5]$ in increments of 0.1. The points obtained by applying each filter were added to the points obtained by the previous one, deleting the repeated points. The same procedure was repeated with the points obtained by the remaining DoG filters. These points passed through the three selection methods for selecting signals (potential microcalcification), according to region area, gray level and the gray gradient. The result was a list of 1,242,179 signals (potential microcalcifications) represented by their centroids.

The additional data included with the MIAS database define, with centroids and radii, the areas in the mammograms where microcalcification clusters are located. It is supposed that signals within these areas are mainly microcalcifications, but there are many signals that lie outside the marked areas. With these data and the support of expert radiologists, all the signals located in these 22 mammograms were pre-classified into microcalcification, and not-microcalcifications. From the 1,242,179 signals, only 4,612 (0.37%) were microcalcifications, and the remaining 1,237,567 (99.63%) were not. Because of this imbalanced distribution of examples of each class, an exploratory sampling was made. Several samplings with different proportions of each class were tested and finally we decided to use a sample of 10,000 signals, including 2,500 real microcalcifications in it (25%).

After the 47 microcalcification features were extracted from each signal, the first method for feature selection, based on the forward sequential search, reduced the relevant features to only three: absolute contrast, standard deviation of the gray level of the signal and the third moment of contour sequence. A transactional database was obtained, containing 10,000 signals (2500 of them being real microcalcifications randomly distributed) and three features describing each signal. For using the second approach, using the GA, the original transactional database with all the 47 features were used.

4.2 Classification of Signals into Microcalcifications

For testing the first feature selection method, based on the forward sequential search, a simple feedforward NN with

three inputs (corresponding to the three features selected by this method: absolute contrast, standard deviation of the gray level and the third moment of contour sequence) was trained and tested. The architecture of this NN consisted of three inputs, seven neurons in the hidden layer and one output. All the units had the sigmoid hyperbolic tangent function as the transfer function. The data (input and targets) were scaled in the range $[-1, 1]$ and divided into ten non-overlapping splits, each one with 90% of the data for training and the remaining 10% for testing. A ten-fold crossvalidation trial was performed; that is, the NN was trained ten times, each time using a different split on the data and the averages of the overall performance, sensitivity and specificity were reported. These results are shown in Table 3 on the row “FSS”, representing the NN that had the best performance in terms of overall accuracy (percentage of correctly classified microcalcifications). The sensitivity (percentage of true positives or correctly classified microcalcifications) and specificity (percentage of true negatives or correctly classified objects that are not microcalcifications) of this NN are also shown.

Also, a GA was combined with NNs to select the features to train them, as described earlier. The GA had a population of 50 individuals, each one with a length of $l = 47$ bits, representing the inclusion (or exclusion) of each one of the 47 features extracted from the signals. We used a simple GA, with gray encoding, stochastic universal sampling selection, single-point crossover, fitness based reinsertion and a generational gap of 0.9. The probability of crossover was 0.7 and the probability of mutation was $1/l$, where l is the length of the chromosome (in this case, $1/l = 1/47 = 0.0213$). The initial population of the GA was always initialized uniformly at random. All the NNs constructed by the GA are feedforward networks with one hidden layer. All neurons have biases with a constant input of 1.0. The NNs are fully connected, and the transfer functions of every unit is the sigmoid hyperbolic tangent function. The data (input and targets) were normalized to the interval $[-1, 1]$. For the targets, a value of “-1” means “not-microcalcification” and a value of “1” means “microcalcification”. For training each NN, backpropagation was used, only one split of the data was considered (90% for training and 10% for testing) and the training stopped after 20 epochs. The GA ran for 50 generations, and the results of this experiment are shown in Table 3 on the row “GA”.

Table 3: Average sensitivity, specificity and overall accuracy of two NNs applied in the classification of individual microcalcifications, with features (inputs) selected using the method of forward sequential search (FSS) and using a Genetic Algorithm (GA)

Method	Sensitivity (%)	Specificity (%)	Overall (%)
FSS	76.21	81.92	81.33
GA	83.33	94.87	95.40

The best solution found is a NN with 23 inputs (five related to the contrast of the signal, four related to the background contrast, two related to the relative contrast, seven related to the shape, four moments of the contour sequence

and only one of the invariant geometric moments), corresponding to 48.94% of the original 47 extracted features. All the NNs coded in the chromosomes of the final population of the GA use 20.02 inputs on average, that is, the NNs with the best performance need only 42.60% of the original 47 features extracted from the microcalcifications.

4.3 Microcalcification Clusters Detection and Classification

The process of cluster detection and the subsequent feature extraction phase generates another transactional database, this time containing the information of every microcalcification cluster detected in the images. A total of 40 clusters were detected in the 22 mammograms from the MIAS database that were used in this study. According to MIAS additional data and the advice of expert radiologists, 10 clusters are benign and 30 are malignant. The number of features extracted from them is 30, but after the two feature selection processes already discussed in previous sections, the number of relevant features we considered relevant was three: minimum diameter, minimum radius and mean radius of the clusters, which were previously defined.

As in the stage of signal classification, a simple feedforward NN with three inputs (corresponding to the three features from the clusters selected by the processes suggested by Kozlov [11]) was trained and tested. The architecture of this NN had three inputs, seven neurons in the hidden layer and only one output. The sigmoid hyperbolic tangent function was used as the transfer function for every neuron. The data (input and targets) were scaled in the range $[-1, 1]$ and divided into ten non-overlapping splits, each one with 90% of the data for training and the remaining 10% for testing. A ten-fold crossvalidation trial was performed; that is, the NN was trained ten times, each time using a different split on the data and the averages of the overall performance, sensitivity and specificity were reported. These results are shown in Table 4 on the row “FSS”, representing the NN that had the best performance in terms of overall accuracy (percentage of correctly classified clusters). The sensitivity (percentage of true positives or correctly classified malignant clusters) and specificity (percentage of true negatives or correctly classified benign clusters) of this NN are also shown.

A GA was also used to select the features for training ANNs, as described earlier. In this case, the transactional database containing the 30 features extracted from the clusters was used. The GA had a population of 50 individuals, each one with a length of $l = 30$ bits, representing the inclusion (or exclusion) of each one of the 30 features extracted from the clusters. We used a simple GA, with gray encoding, stochastic universal sampling selection, single-point crossover, fitness based reinsertion and a generational gap of 0.9. The probability of crossover was 0.7 and the probability of mutation was $1/l = 1/30 = 0.0333$. The initial population of the GA was initialized uniformly at random. All the NNs constructed by the GA are feedforward networks with one hidden layer. All neurons have biases with a constant input of 1.0. The NNs are fully connected, and the transfer functions of every neuron is the sigmoid hyperbolic tangent function. The data (input and targets) were normalized to the interval $[-1, 1]$. For the targets, a value of “-1” means that the cluster is “benign” and a value of “1” means “malignant”. For training each NN, backpropa-

gation was used, considering 10 splits of the data as in the previous experiment (90% for training and 10% for testing) and the training stopped after 20 epochs. The GA ran for 50 generations, and the results of this experiment are shown in Table 4 on the row “GA”.

Table 4: Average sensitivity, specificity and overall accuracy of two NNs applied in the classification of microcalcification clusters, with features (inputs) selected using the method of the forward sequential search (FSS) and using a Genetic Algorithm (GA)

Method	Sensitivity (%)	Specificity (%)	Overall (%)
FSS	53.85	88.89	77.50
GA	100.00	100.00	100.00

The best solution has 9 inputs, corresponding to 30% of the original cluster feature set (five features related to the shape of the cluster, one related to the area of the microcalcifications and three related to the contrast of the microcalcifications). On average, the chromosomes of the last generation coded 14.03 inputs, that is, the NNs with the best performance only receive 46.76% of the original features extracted from the microcalcification clusters.

5. CONCLUSIONS AND FUTURE WORK

This paper presented a comparison of two methods for feature selection from individual microcalcifications and microcalcification clusters in mammograms, for their classification using a feedforward NN. The first method uses two feature selection processes [11]: the first process attempts to delete the features that present high correlation with other features, and the second process uses a derivation of the forward sequential search algorithm, which is a sub-optimal search algorithm, adding features while the error is decreasing and stopping when it increases again. The second method uses a GA for selecting the most relevant features in order to improve the accuracy, evolving a population of NNs with different subsets of the features as inputs.

We found that the use of GAs combined with NNs greatly improves the overall accuracy, the specificity and the sensitivity of the classification, when signals are classified. The best solution found is a NN with 23 inputs, corresponding to 23 extracted features (five related with the contrast of the signal, four related with the background contrast, two related with the relative contrast, seven related to the shape, four moments of the contour sequence and only one of the invariant geometric moments). We found also that all the NNs coded in the chromosomes of the final population of the GA use 20.02 inputs on average; that is, the NNs with the best performance need only 42.60% of the original 47 original features.

In the case of the classification of microcalcification clusters, we observed that the use of a GA greatly improved the overall accuracy, the sensitivity and the specificity, achieving values of 100%. The best solution has 9 inputs, corresponding to 9 extracted features from the clusters (five related to the shape of the cluster, one related to the area of the microcalcifications and three related to the contrast of the microcalcifications). On average, the best NNs architectures

receive 14.03 inputs on average, that is, they only receive 46.76% of the 30 original cluster features as inputs. Nevertheless, only 40 microcalcification clusters were detected in the 22 mammograms used in this study. The test sets used in the ten-fold crossvalidation trial were very small and in some splits, all the examples belonged to only one of the two classes so either sensitivity or specificity could not be calculated. These splits were ignored in the calculation of the respective mean.

As future work, it would be useful to include and process other mammography databases, in order to have more examples and produce transactional feature databases more balanced and complete, and test also how different resolutions could affect system effectiveness. Some of the parameters that were empirically set in this work (like the binarization threshold in the pre-processing stage) need to be determined via automated procedures when dealing with other larger mammographic databases. The size of the gaussian filters could be adapted depending on the size of the microcalcifications to be detected and the resolution of images. The correspondence between the spatial frequency of the image and the relation σ_1/σ_2 has to be thoroughly studied. Different new features could be extracted from the microcalcifications in the images and tested also.

In this study, simple GAs and NNs were used, and more sophisticated versions of these methods could produce better results. The use of real valued chromosomes, chromosomes with indirect representation (metaheuristics, NN construction rules, etc.) are other approaches that could give different results. The inclusion of simple backpropagation training in the EANNs have consequences of longer computation times, so alternatives to backpropagation should be tested in order to reduce time costs.

5.1 Acknowledgments

This research was supported in part by the Instituto Tecnológico y de Estudios Superiores de Monterrey (ITESM) under the Research Chair CAT-010 and the National Council of Science and Technology of Mexico (CONACYT) under grant 41515.

6. REFERENCES

- [1] I. Anttinen, M. Pamilo, M. Soiva, and M. Roiha. Double reading of mammography screening films: one radiologist or two? *Clin. Radiol.*, 48:414–421, 1993.
- [2] K. Balakrishnan and V. Honavar. Evolutionary design of neural architectures. a preliminary taxonomy and guide to literature. Technical Report CS TR 95-01, Iowa State University, Department of Computer Sciences, 1995.
- [3] E. Cantú-Paz and C. Kamath. Evolving neural networks for the classification of galaxies. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2002)*, pages 1019–1026, San Francisco, CA, USA, 2002.
- [4] R. Chandrasekhar and Y. Attikiouzel. Digitization regime as a cause for variation in algorithm performance across two mammogram databases. Technical Report 99/05, The University of Western Australia, Centre for Intelligent Information Processing Systems, Department of Electrical and Electronic Engineering, 1999.

- [5] J. Dengler, S. Behrens, and J. F. Desaga. Segmentation of microcalcifications in mammograms. *IEEE Trans. Med. Imaging*, 12(4):634–642, 1993.
- [6] M. A. Ganott, K. M. Harris, H. M. Klaman, and T. L. Keeling. Analysis of false-negative cancer cases identified with a mammography audit. *The Breast Journal*, 5(3):166–175, 1999.
- [7] T. O. Gulsrud. Analysis of mammographic microcalcifications using a computationally efficient filter bank. Technical report, Stavanger University College, Department of Electrical and Computer Engineering, 2001.
- [8] L. Gupta and M. D. Srinath. Contour sequence moments for the classification of closed planar shapes. *Pattern Recognition*, 20(3):267–272, 1987.
- [9] M. Heath, K. Bowyer, D. Kopans, R. Moore, and P. Kegelmeyer Jr. The digital database for screening mammography. In *Proceedings of the 5th International Workshop on Digital Mammography*, Toronto, ON, Canada, 2000.
- [10] B.-W. Hong and M. Brady. Segmentation of mammograms in topographic approach. In *IEE International Conference on Visual Information Engineering*, Guildford, UK, 2003.
- [11] A. Kozlov and D. Koller. Nonuniform dynamic discretization in hybrid networks. In *Proceedings of the 13th Annual Conference of Uncertainty in AI (UAI)*, pages 314–325, Providence, RI, USA, 2003.
- [12] V. Kurkova. *Kolmogorov’s Theorem*, pages 501–502. The handbook of brain theory and neural networks. MIT Press, Cambridge, MA, USA, 1995.
- [13] S. Li, T. Hara, Y. Hatanaka, H. Fujita, T. Endo, and T. Iwase. Performance evaluation of a CAD system for detecting masses on mammograms by using the MIAS database. *Medical Imaging and Information Science*, 18(3):144–153, 2001.
- [14] E. M. Ochoa. Clustered microcalcification detection using optimized difference of gaussians. Master thesis, Air Force Institute of Technology, Wright-Patterson Air Force Base, 1996.
- [15] S. Oporto-Díaz. Automatic detection of microcalcification clusters in digital mammograms. Master thesis, Center for Intelligent Systems, Tecnológico de Monterrey, Campus Monterrey, Monterrey, Mexico, 2004.
- [16] S. Oporto-Díaz, R. R. Hernández-Cisneros, and H. Terashima-Marín. Detection of microcalcification clusters in mammograms using a difference of optimized gaussian filters. In *Proceedings of the Second International Conference on Image Analysis and Recognition, ICIAR 2005*, pages 998–1005, Toronto, ON, Canada, 2005. Springer.
- [17] J. Suckling, J. Parker, D. Dance, S. Astley, I. Hutt, C. Boggis, I. Ricketts, E. Stamatakis, N. Cerneaz, S. Kok, P. Taylor, D. Betal, and J. Savage. The Mammographic Image Analysis Society digital mammogram database. *Excerpta Medica International Congress Series*, 1069:375–378, 1994.
- [18] E. L. Thurfjell, K. A. Lernevall, and A. A. S. Taube. Benefit of independent double reading in a population-based mammography screening program. *Radiology*, 191:241–244, 1994.
- [19] X. Yao. Evolving artificial neural networks. In *Proceedings of the IEEE*, volume 87, pages 1423–1447, 1999.