

Robustness of a CAD System on Digitized Mammograms

Antonio García-Manso¹, Carlos J. García-Orellana¹,
Ramón Gallardo-Caballero¹, Nico Lanconelli², Horacio González-Velasco¹,
and Miguel Macías-Macías¹

¹ Pattern Classification and Image Analysis Group (CAPI),
University of Extremadura,
Avenida de Elvas, s/n., Badajoz, Extremadura, Spain
antonio@capi.unex.es

² Medical Imaging Group, Physics Dpt. - Bologna University, Viale B., Pichat 6/2,
40127 Bologna, Italy

Abstract. In this paper we study the robustness of our CAD system, since this is one of the main factors that determine its quality. A CAD system must guarantee consistent performance over time and in various clinical situations. Our CAD system is based on the extraction of features from the mammographic image by means of Independent Component Analysis, and machine learning classifiers, such as Neural Networks and Support Vector Machine. To measure the robustness of our CAD system we have used the digitized mammograms of the USF's DDSM database, because this database was built by digitizing mammograms from four different institutions (four different scanner) during more than 10 years. Thus, we can use the mammograms digitized with one scanner to train the system and the remaining to evaluate the performance, what gives us a measure of the robustness of our CAD system.

Keywords: ICA, NN, SVM, CAD.

1 Introduction and Purpose

The goal of a CAD system is to be able to analyze an image and indicate the location of possible lesions, if any. But also, there are several factors that should provide a CAD system to detect and diagnose masses in mammograms, such as: high sensitivity to detect the largest possible number of cancers, high specificity to reduce the number of false positives per image, acceptable call rate, early detection to increase the chances of survival, low processing time and robustness [1]. That is, the system must guarantee consistent performance along time and in various clinical situations. We have designed and implemented a CAD system to detect and classify masses in mammograms. We have used the USF's DDSM database [2] to train and test our CAD system.

The DDSM contains mammograms obtained from examinations between October 1988 and February 1999 in four different clinical sites: Massachusetts General Hospital (MGH) in Boston, Wake Forest University School of Medicine

(WFU) in North Carolina, Sacred Heart Hospital in Pensacola (SH), Florida, and Washington University School of Medicine in St. Louis Medical Center (WU). These mammograms were digitized using four different scanners: DBA M2100 ImageClear with a resolution of $42\mu m$, Howtek with a resolution of $43.5\mu m$, Lumisys 200 laser with a resolution of $50\mu m$ and Howtek MultiRAD 850 with a resolution of $43.5\mu m$. With that we can get an idea of the heterogeneous that is the used dataset. But, to normalize and avoid this heterogeneity in the used dataset we use the calibration curves which are available of each scanner to obtain the mammographic images in optical densities. In that way, at least in theory, we will have all dataset normalized in the same conditions. But the number of prototypes regarding to each class digitized with each scanner can be very different considering a scanner, or other. And, also, the way in which were indicated the ground truth over the mammograms could be very different [3], because of the long period of time (more than 10 years) during which was built the DDSM database. That is, in the DDSM database we can see different styles when the radiologists indicate the lesions on mammograms. First, because, surely, many radiologists were involved, and in addition, because it was done considering informed mammograms of four different institutions.

We propose a system to detect masses in mammograms as a two-class pattern recognition problem (mass or normal tissue), but, in our proposed approach, no modelling has been used. In contrast, we have used features extraction based on Independent Component Analysis (ICA), for its ability to obtain a basis functions adapted to the problem, especially to the natural images [4,5]. Thus, we have obtained some basis functions (basis images) to expand the original image (original patch), where the coefficients of this expansion will be used to form the input vectors to the classifiers.

The rest of our paper is organized as follows. Section 2 introduces the general concepts of feature extractions, the classifiers and the dataset used in our experiments. Section 3 includes a description of our methodology. Section 4 describes our results. Finally, Section 5 presents the main conclusions of this work.

2 Methods

In this section, we provide a brief description of the mammogram database utilized. Additionally we describe the procedure implemented to build a set of mass prototypes and the main characteristics of the selected image feature extractor. Finally, we provide a short description of the used classifiers.

2.1 Data and Prototype Creation

The Digital Database for Screening Mammography [2] is a resource available to the mammographic image analysis research community. Contains a total of 2,620 cases. Each case provides four screening views, mediolateral oblique (MLO) and craniocaudal (CC) projections of left and right breasts. Therefore, the database has a total of 10,480 images.

Cases are categorized in four major groups: *normal*, *cancer*, *benign* and *benign without callback*.

All cases in the DDSM database were reported by experienced radiologists providing various BIRADS parameters (density, assessment and subtlety), BIRADS abnormality description and proven pathology. For each abnormality identified, the radiologists draw free form digital curves defining ground truth regions. We use these regions to define squared regions of interest (ROIs) for use as prototypes of mass. Each DDSM case includes additional information such as patient age, date of study and digitization or digitizer's brand.

The DDSM database contains 2,582 mass prototypes including benign and malignant masses. Some of them were located on the border of the mammograms. Consequently, only 2,324 prototypes could be considered, namely, those which might be taken centered in a square without stretching. Some mass prototype examples are shown in Figure 1.

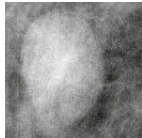
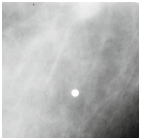
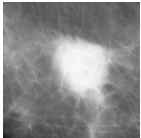
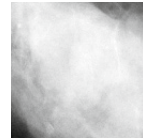
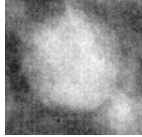
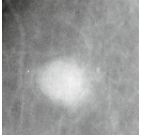
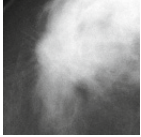
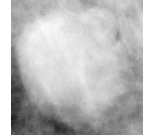
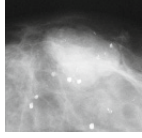
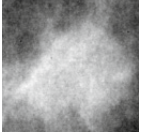
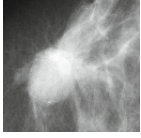
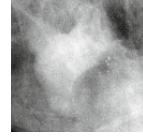
	Shape	Edges			
		Circumscribed	Ill-defined	Spiculate	Obscured
Oval					
	Case3090-RCC	Case0107-LCC	Case4178-LCC	Case0418-LCC	
	Round				
		Case3391-RCC	Case4021-RCC	Case0339-RCC	Case1576-RCC
Lobulate					
		Case0145-RCC	Case1908-RCC	Case0457-RMLO	Case0418-LMLO

Fig. 1. Mass samples for each shape and margin combination. Each ROI has been resized to a common size of 128×128 pixels. Mammogram case and view is located under each ROI.

Regions of Interest. Ground truth regions are defined in the database by a chain code which generates a free hand closed curve. We use the chain code to determine the smallest square region of the mammogram that includes the manually defined region. Therefore, if the mass is located near one edge of the mammogram, this procedure may not be able to obtain a squared region from

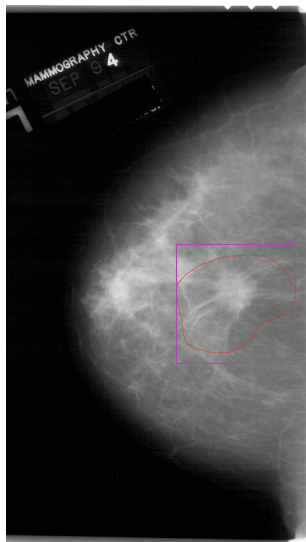


Fig. 2. Ground truth region defined by radiologist (red solid line) and considered ROI (purple box) on a DDSM mammogram

the image and the mass is discarded as a valid prototype. Figure 2 shows the ground truth region coded by the radiologist (red solid line) and the area to be used as ROI (purple box).

USF's DDSM mammograms were digitized with four different scanners for which optical density calibration and spatial resolution are known [2]. Three scanners provide a linear optical response and the fourth one of logarithmic type. To eliminate the dependence of the origin of each digitized mammogram, all obtained ROIs were converted to optical density using the referenced calibration parameters.

The generated regions have different sizes but the selected image feature extractor needs to operate on regions with the same size. So, we need to reduce the size of the selected regions to common sizes. The reduction of ROIs to a common size has demonstrated to preserve mass malignancy information [6,7,8]. To determine the optimum region size, we resized each ROI to two sizes: 32×32 , 64×64 pixels. We also tried other sizes such as 128×128 pixels, but the performance obtained with this size was not better than that obtained with the two smaller sizes, whereas the computation time was much greater. Resizing has been carried out using the bilinear interpolation algorithm provided by the OpenCV library [9].

2.2 Independent Component Analysis

The original motivation of ICA is to solve problems known as *blind source separation*(BBS). These problems consist in the following: suppose that we have n

signals. The objective is to develop the signals registered by the sensors (\mathbf{x}_i) as a linear combination of n sources (\mathbf{s}_j), in principle unknown [10].

$$x_i = \sum_{j=1}^n a_{ij} s_j$$

The goal of ICA is to estimate the mixture matrix $\mathbf{A} = (a_{ij})$, along with the sources \mathbf{s}_j . The ICA model supposes that the observed signals are a linear transformation of hidden sources: $\mathbf{x} = \mathbf{A} \cdot \mathbf{s}$. In general, the mixture matrix \mathbf{A} is invertible, so we have:

$$\mathbf{x} = \mathbf{A} \cdot \mathbf{s} \Rightarrow \mathbf{s} = \mathbf{W} \cdot \mathbf{x} \quad \text{with} \quad \mathbf{W} = \mathbf{A}^{-1}$$

It is important to remark that:

- The key of ICA estimation is to suppose that hidden sources (\mathbf{s}) are non-gaussian and statistically independent.
- We cannot determine the variances (energies) of the independent components.
 - Therefore, the magnitudes of the s_i can be freely normalized.
 - We cannot determine the order of the independent components.

We can use this technique for feature extraction since the components of \mathbf{x} can be regarded as the characteristics representing the objects (patterns) [10]. We have used the FastICA algorithm [11] proceeding as follows:

- We start with N samples (patches, N vectors of dimension p) forming the patches matrix (\mathbf{x}) where each row is a patch (\mathbf{i}), therefore, the dimensions of this matrix are $N \times p$.
- First, the data are centered by subtracting their averages. That is, to each element is subtracted the mean of its column ($\mathbf{m} = E\{\mathbf{x}\}$) so as to make (\mathbf{x}) a zero-mean variable, which implies that \mathbf{s} is zero-mean as well. After estimating the mixing matrix ($\mathbf{A} = (a_{ij})$) with centered data, we can complete the estimation by adding the mean vector of \mathbf{s} back to the centered estimates of \mathbf{s} . The mean vector of \mathbf{s} is given by $\mathbf{A}^{-1}\mathbf{m}$.
- Then a whitening process is applied. This transformation consist in to uncorrelate the data so that their variances are equal to 1 and the covariance matrix is the identity matrix. The dimension of this new whitening matrix is $p \times p$.
- To reduce the size of the input space is applied Principal Component Analysis (PCA) [12], ordering the array of eigenvectors (whitening matrix) by its eigenvalues from highest to lowest and discarding those with lower eigenvalue that will be those with a smaller contribution the variance. Taking q ($q < p$) first components we obtain the matrix \mathbf{K}_{PCA} of dimension ($q \times p$).
- Now, taking as input this matrix and applying the ICA algorithm is obtained the ICA transformation matrix of dimension ($q \times q$).

- Finally considering a new matrix (\mathbf{W}_T), multiplication of the previous two ($\mathbf{W}_T = \mathbf{K}_{PCA}^T \cdot \mathbf{W}$ ($p \times q$)), in which each row is a vector of the new base, q characteristics can be extracted of each original input (\mathbf{i}) simply by multiplying the matrix \mathbf{W}_T for each of them.

$$\mathbf{c} = \mathbf{i} \cdot \mathbf{W}_T$$

Following this process we can express, as many other transformations (wavelets, Gabor filters, ... [13]), the image (or a image patch) as a linear superposition of some basis functions (basis images in our case) $a_i(x, y)$:

$$I(x, y) = \sum_{i=1}^p a_i(x, y) c_i \tag{1}$$

Where the c_i are image-dependent coefficients. This expression is similar to the ICA model, and we can visualize this idea in Figure 3. In that way, estimating a basis images using ICA, we could obtain a basis adapted to our data.

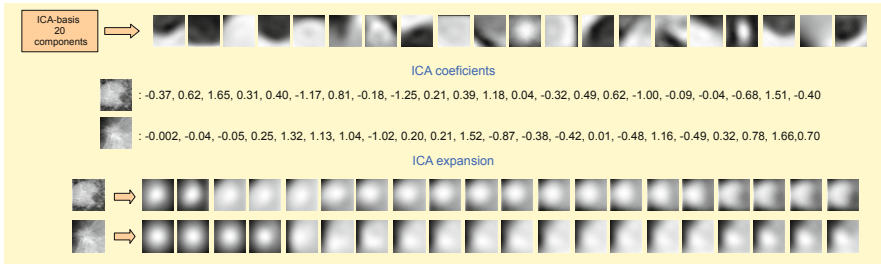


Fig. 3. Example ICA basis and expansion of two mass prototypes

In Figure 3, one can see an example of an ICA basis of 20 components that operates on prototypes with dimension 64×64 pixels. In this figure, there are 20 components extracted by the ICA basis for two mass prototypes. Then one can also see its subsequent expansion using the extracted coefficients and the ICA basis of the form shown in Eq. 1. That is, we can decompose or expand our image ($I(x, y)$) by using a base image ($a_i(x, y)$) multiplied by coefficients (c_i) and adding the mean vector $\mathbf{m} = E\{\mathbf{x}\}$ (not showed in the figure).

2.3 Classification Algorithm

The classification algorithm has the work of learning from data. Usually, a model with excessive complexity leads to poor generalization results. In the learning process is convenient to use, at least, two independent sets of patterns: one for training and another for testing. Or, as in this work, we have used three independent sets of patterns: one for training, one for avoid the overtraining and another for testing [14]. We have used Neural Networks and Support Vector Machine (SVM) [15] classifiers.

Neural Networks (NN). We have used the classical feed-forward multilayer perceptron with a single hidden layer and, a variant of Back-Propagation (BP) algorithm named Resilient Back-Propagation (Rprop) [16] to adjust the weights. Rprop is a local adaptive learning scheme performing supervised batch-learning in multilayer perceptron with faster convergence than the standard BP algorithm. The basic principle of Rprop is to eliminate the negative affect of the size of the partial derivative on the update process. As a consequence, only the sing of the derivative is considered to indicate the direction of the weight update [16]. The library of functions of the Stuttgart Neural Network Simulator environment [17] were used to generate and train the NN classifiers. To avoid local minima during the training process each setting was repeated four times changing randomly the initial weights in the net. Furthermore, the number the neurons in the hidden layer could change between 50 and 650 in steps of 50.

Support Vector Machines. The goal of SVM is to find a model (based on the training prototypes) which is able to predict the class membership of the prototypes of the test subset, based on the value of their characteristics.

Given a labeled training set of the form $(\mathbf{x}_i, \mathbf{y}_i)$, $i = 1, \dots, l$ where $\mathbf{x}_i \in \mathbb{R}^n$ and $\mathbf{y} \in \{1, -1\}^l$, the SVM algorithm requires the following optimization problem to be solved:

$$\begin{aligned} \min_{w,b,\xi} \quad & \frac{1}{2}w^T w + C \sum_{i=1}^l \xi_i \\ \text{where} \quad & y_i (w^T \phi(\mathbf{x}_i) + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0 \end{aligned} \tag{2}$$

This algorithm works by projecting the training vectors \mathbf{x}_i onto a higher-dimensional space than the original. The final dimension of this space depends on the complexity of the input space. Thus SVM finds a linear separation by means of a hyperplane with a maximal (i.e., optimal) margin of separation between classes in this higher dimensional space.

The parameter C ($C > 0$) shown in the model is a penalty term to control the error, and $K(\mathbf{x}_i, \mathbf{x}_j) \equiv \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$ is a kernel function to project the input data onto to a higher dimensional space. We have used LibSVM [18] library in this work with a radial basis function (RBF: $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$, $\gamma > 0$) as kernel function. To find the optimal configuration of the parameters in the algorithm γ could change between -5 and 20 in step of 0.5 and the penalty parameter C between -5 and 10 , also, in steps of 0.5 .

3 System

In this section, we provide a description of the main steps of our system. The first task is to obtain the prototypes of masses and normal tissue. The prototypes of masses are obtained as was explained in Section 2.1 and the prototypes of normal

tissue were selected randomly from the normal mammograms. This normal tissue prototypes were caught originally with sizes that randomly ranging from the smallest to the largest of the sizes found in the DDSM database for masses. Then, applying the FastICA algorithm [11] as is described in Section 2.2 and using *log cosh* function to estimate neg-entropy were obtained the ICA basis (ICA-based feature extractor). To obtain the optimal configuration of the system were generated different ICA basis to extract different number of features (from 10 to 65 in steps of 5) from the original patches and, in addition, operating over patches of different dimensions (as said before, 32×32 and 64×64).

We did a double training process, on the one hand, we trained NN classifiers and, on the other hand, we trained SVM classifiers. After training process were obtained the results shown in Figure 3, where can be seen the results obtained over the test subsets in a 10-fold cross validation scheme. In that way, we find the best configuration of the feature extractor. This study was made with a total of 5,052 prototypes: 1,197 of malignant masses, 1,133 of benign masses and 2,722 of normal tissue. We found that the optimal configuration for the ICA-based feature extractor, for a NN classifier, was a feature extractor that operated on prototypes of 64×64 pixels extracting 10 components (average success 86.33%). And, for a SVM classifier, the best configuration was for a feature extractor that operated also on prototypes of 64×64 pixels extracting 15 components (average success 88.41%). Therefore, the results shown in the following section were obtained using these configurations for the ICA-based feature extractor in each case.

4 Results

In this work, our main interest was to evaluate the robustness of our CAD system. We have included all the prototypes of masses found in the DDSM which could be obtained as a square shape without stretching them by determining the smallest square region that includes the complete ground truth. The distribution of prototypes is shown in Table 1. As can be seen in this table, the number of prototypes on the learning and test sets is quite different depending on the considered scanner, being the most heterogeneous distribution for the DBA M2100 scanner. For this scanner, no prototypes were found of benign masses and the number of prototypes of malignant masses is much lower than the normal tissue prototypes in the learning set. This, as will be seen in the results, is a big handicap to train the classifiers.

In Table 1 one can see that the number of normal tissue prototypes from a DBA scanner is more than half of total normal tissue prototypes. This is due to that in the DDSM database there are 12 volumes of normal mammograms with different number of cases by each volume, with four mammograms by case. From these 12 volumes 6 were digitized with a DBA scanner, among them, those that have a bigger number of cases. Therefore, it seems clear that if we have selected the normal tissue prototypes randomly from the normal mammograms the number of prototypes from a DBA scanner should be, at least, the half of

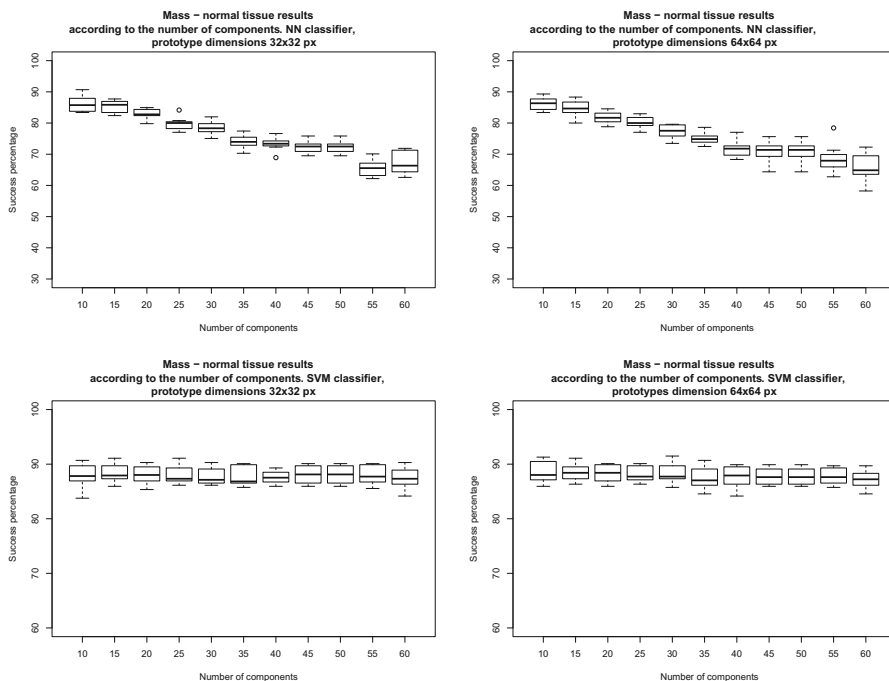


Fig. 4. Choosing the best configuration to the feature extractor. The top row shows the results when we use a NN classifier, while the bottom row shows the results for a SVM classifier. In both cases, prototypes of 32×32 first column and 64×64 , second column.

the total of normal tissue prototypes. On the other hand, the number of volumes of “cancer” from a DBA scanner is only two and of “benign” zero. The number of prototypes is more equilibrated for the rest of scanners.

The results, presented in Table 2, correspond to feature extractors based on ICA described in the previous Section 3 for each classifier. As expected, the most suitable distribution of prototypes is for overall results, since, as seen in Table 1, the number of prototypes in the learning set is much greater than the number of prototypes in the test set. And, in addition, the trained classifier with this distribution can “learn” prototypes from all the scanners, while, the trained classifiers with the other distributions only can learn prototypes from one scanner. Anyway, in theory this should not affect to the results, because the prototype images are transformed to optical density by using the scanner calibration parameters provided by the USF’s DDSM database authors.

In Table 2 one can see that when the number of prototypes in the learning set is low (HOWTEK scanners) the performance with SVM classifiers is a little better than with NN classifiers. This seems to agree with [19] where was made a comparison of the performance and robustness of different types of classifiers in different settings. In contrast, when the number of prototypes is large enough

Table 1. Distribution of prototypes in learning and test sets, depending on the used scanner and without considering the scanner (overall results). Along with the scanner’s name appears the identification of the institution in which they were used.

Distribution of prototypes depending on the scanner.			
Scanner	Pathology Learning Test		
HOWTEK 960 (MGH) <i>43.5μm/pixel</i> linear calibration	Malignant	345	851
	Benign	485	648
	Normal	312	2410
HOWTEK MultiRAD 850 (WU) <i>43.5μm/pixel</i> linear calibration	Malignant	107	1089
	Benign	154	979
	Normal	417	2305
DBA M2100 (MGH) <i>42μm/pixel</i> logarithmic calibration	Malignant	105	1091
	Benign	0	1133
	Normal	1668	1054
LUMISYS 200 laser (WFU & SH) <i>50μm/pixel</i> linear calibration	Malignant	639	557
	Benign	494	639
	Normal	325	2397
Overall	Malignant	1074	122
	Benign	1032	101
	Normal	2440	282

Table 2. In this table the obtained results are shown. The learning set is divided into two: the train subset and the validation (Val) subset, corresponding to the first the 80% and the second the 20% of the prototypes in the learning set.

Success results for discriminating Mass-Normal tissue						
Scanner	SVM classifier			NN classifier		
	Learning (%)		Test (%)	Learning (%)		Test (%)
	Train(80%)	Val(20%)		Train(80%)	Val(20%)	
HOWTEK 960	96.38	93.89	75.29	95.40	89.52	71.22
HOWTEK MD850	98.71	91.91	82.71	87.27	80.88	80.56
DBA M2100	99.80	97.46	58.69	99.29	96.90	57.75
LUMISYS	94.17	90.75	71.72	94.68	88.01	76.62
Overall results	92.65	88.20	88.41	90.73	88.37	86.33

(LUMISYS scanner), the performance with NN classifiers seems to be a little better than with SVM classifiers. However, considering overall results, the performance is slightly better with SVM classifiers, which contradicts the previous statement. Finally, when the number of prototypes is not enough (DBA scanner) the performance for both classifiers is quite bad.

5 Conclusions

This study about the robustness of our CAD system has been made using a very heterogeneous dataset. Data (mammograms and reports) comes from four clinical sites. The mammograms were digitized in three different institutions using, in principle, four different scanners but the DBA scanner, used at MGH, was retired due to continuing performance difficulties [20]. We can evaluate the performance obtained in each case choosing as a reference the overall performance over the test subset, because in this setting the learning set was formed by prototypes of all scanner and the classifiers can *learn* prototypes of all them. Considering a SVM classifier the differences in performance were: 14.8% for Howtek 960, 6.4% for Howtek MD850, 33.6% for DBA and 18.8% for Lumisys. And considering a NN classifier, 17.5% for Howtek 960, 6.6% for Howtek MD850, 33.1% for DBA and 11.6% for Lumisys. Taking into account all was said about the DBA scanner, we have to say that the results for this scanner are not conclusive. On the other hand, the least variation in the performance is found for Howtek MD850 scanner for both classifiers which could indicates that for this scanner had a good representation of the entire dataset. For Howtek 960 the performance variation is higher for a SVM classifier and for Lumisys the performance variation is higher for a NN classifier. Here, we can see that when the number of prototypes in the learning set is large enough, the NN classifiers obtain a better generalization capability than the SVM classifiers. While, when the number of prototypes in the learning set is low the generalization capability of the SVM classifiers is better than with NN classifiers.

Acknowledgments. This work has been partly supported by “Junta de Extremadura” and FEDER through projects PRI08A092, PDT09A036, GR10018 and PDT09A005.

References

1. Serio, G.V., Novello, A.C.: The advisability of the adoption of a law that would expand the definition of mammography screening to include the review of x-ray examinations by use of a computer aided detection device. Technical report, The Superintendent of Insurance in consultation with the Commissioner of Health, USA (2003)
2. Heath, M., Bowyer, K., Kopans, D., Moore, R., Kegelmeyer, P.: The digital database for screening mammography. In: Proceedings of the 5th International Workshop on Digital Mammography, pp. 212–218 (2000)
3. Horsch, A., Hapfelmeier, A., Elter, M.: Needs assessment for next generation computer-aided mammography reference image databases and evaluation studies. *International Journal of Computer Assisted Radiology and Surgery*, 1–19 (2011)
4. Hyvärinen, A., Hurri, J., Hoyer, P.O.: *Natural Image Statistics. A Probabilistic Approach to Early Computational Vision*. Springer (2009)
5. Anjali, P., Ajay, S.: A review on natural image denoising using independent component analysis (ICA) technique. *Advances in Computational Research* 2(1), 06–14 (2010)

6. Campanini, R., Dongiovanni, D., Iampieri, E., Lanconelli, N., Masotti, M., Palermo, G., Riccardi, A., Roffilli, M.: A novel featureless approach to mass detection in digital mammograms based on support vector machines. *Physics in Medicine and Biology* 49, 961–975 (2004)
7. Angelini, E., Campanini, R., Iampieri, E., Lanconelli, N., Masotti, M., Roffilli, M.: Testing the performances of different image representations for mass classification in digital mammograms. *International Journal of Modern Physics C [Computational Physics and Physical Computation]* 17(1), 113–131 (2006)
8. Hong, B.-W., Brady, J.M.: A Topographic Representation for Mammogram Segmentation. In: Ellis, R.E., Peters, T.M. (eds.) *MICCAI 2003*. LNCS, vol. 2879, pp. 730–737. Springer, Heidelberg (2003)
9. Bradski, G., Kaehler, A.: *Learning OpenCV: Computer Vision with the OpenCV Library*. O'Reilly Media (2008)
10. Hyvärinen, A., Karhunen, J., Oja, E.: *Independent Component Analysis. Adaptive and Learning Systems for Signal Processing, Communications, and Control*. John Wiley & Sons (2001)
11. Ripley, B.: *FastICA Algorithms to perform ICA and Projection Pursuit* (February 2009), <http://cran.r-project.org/web/packages/fastICA/fastICA.pdf>
12. Jolliffe, I.T.: *Principal Component Analysis*, 2nd edn. Series in Statistics. Springer (2002)
13. Gonzalez, R.C., Woods, R.E.: *Digital image processing*, 3rd edn. Prentice-Hall, Upper Saddle River (2008)
14. Bishop, C.M.: *Pattern Recognition and Machine Learning*. Springer (2006)
15. Vapnik, V.N.: *The Nature of Statistical Learning Theory*, 2nd edn. Statistics for Engineering and Information Science. Springer (2000)
16. Riedmiller, H., Braun, H.: A direct adaptive method for faster backpropagation learning. The RPROP algorithm. In: *IEEE International Conference on Neural Networks*, pp. 586–591 (1993)
17. Zell, A., Mache, N., Huebner, R., Mamier, G., Vogt, M., Schmalzl, M., Herrmann, K.: *SNNS (Stuttgart Neural Network Simulator)*. In: *Neural Network Simulation Environments*, pp. 165–186 (1994)
18. Chang, C.C., Lin, C.J.: *LIBSVM: A library for support vector machines*. *ACM Transactions on Intelligent Systems and Technology* 2(3), 27:1–27:27 (2011)
19. Lausser, L., Kestler, H.A.: *Robustness Analysis of Eleven Linear Classifiers in Extremely High-Dimensional Feature Spaces*. In: Schwenker, F., El Gayar, N. (eds.) *ANNPR 2010*. LNCS, vol. 5998, pp. 72–83. Springer, Heidelberg (2010)
20. Bowyer, K.W.: *Digital Image Database with Gold Standard and Performance Metrics for Mammographic Image Analysis Research*. Technical report, U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012 (August 1999)