Superior Performances of the Neural Network on the Masses Lesions Classification through Morphological Lesion Differences

U. Bottigli, R.Chiarucci, B. Golosio, G.L. Masala, P. Oliva, S.Stumbo, D.Cascio, F. Fauci, M. Glorioso, M. Iacomi, R. Magro and G. Raso

Abstract—Purpose of this work is to develop an automatic classification system that could be useful for radiologists in the breast cancer investigation. The software has been designed in the framework of the MAGIC-5 collaboration.

In an automatic classification system the suspicious regions with high probability to include a lesion are extracted from the image as regions of interest (ROIs). Each ROI is characterized by some features based generally on morphological lesion differences. A study in the space features representation is made and some classifiers are tested to distinguish the pathological regions from the healthy ones.

The results provided in terms of sensitivity and specificity will be presented through the ROC (Receiver Operating Characteristic) curves. In particular the best performances are obtained with the Neural Networks in comparison with the K-Nearest Neighbours and the Support Vector Machine: The Radial Basis Function supply the best results with 0.89 ± 0.01 of area under ROC curve but similar results are obtained with the Probabilistic Neural Network and a Multi Layer Perceptron.
.

Keywords—Neural Networks, K-Nearest Neighbours, Support Vector Machine, Computer Aided Detection

I. INTRODUCTION

BREAST cancer is reported as one of the first causes of women mortality [1] and an early diagnosis in asymptomatic women makes it possible to reduce the breast cancer mortality: in spite of a growing number of detected

G.L. Masala is with the Struttura Dipartimentale di Matematica e Fisica dell'Università di Sassari e Sezione INFN di Cagliari, Italy, Via Vienna 2, Sassari, 07100, Italy (corresponding author; phone:+39079229486; fax: +39079229482; e-mail: giovanni.masala@ca.infn.it).

U.Bottigli, was with the Struttura Dipartimentale di Matematica e Fisica dell'Università di Sassari e Sezione INFN di Cagliari. He is now with the Dipartimento di Fisica dell'Università di Siena e Sezione INFN di Cagliari, Italy (e-mail: bottigli@uniss.it).

R. Chiarucci is with the Dipartimento di Fisica dell'Università di Siena, Italy (e-mail: riccardochiarucci@alice.it).

B. Golosio, P. Oliva, S. Stumbo are with the Struttura Dipartimentale di Matematica e Fisica dell'Università di Sassari e Sezione INFN di Cagliari, Italy (e-mail: golosio@uniss.it; oliva@uniss.it; stumbo@uniss.it).

D.Cascio, F.Fauci, M. Glorioso, M. Iacomi, R.Magro and G.Raso are with Dipartimento di Fisica e Tecnologie Relative, Università di Palermo, Palermo, Italy and INFN, Sezione di Catania, Italy (e-mail: donato@difter.unipa.it, ffauci@mail.unipa.it, rmagro@unipa.it, graso@unipa.it).

cancers, the death rate for this pathology decreased during the last 10 years [2], thanks to the screening programs and the relative early diagnosis [3] which consist in a mammographic examination performed for 49-69 years old women. Mammography is widely recognized as the best imaging modality for the early detection of the abnormalities which indicate the breast cancer presence [4]; it is realized by screen-film modality or, more recently, by digital detectors [5]-[7]. It has been estimated that radiologists involved in screening programs fail to detect up to approximately 25% of the breast cancers visible on retrospective reviews; more this percentage increases if minimal signs are considered [8]-[10].

With the framework of the project MAGIC-5 (Medical Application on Grid Infrastructure Connection), a collaboration among Italian physicists and radiologists, it was possible to built a large database of digitized mammographic images; all these images was used to develop a CAD (Computer Aided Detection) for medical applications, such as breast cancer detection through mammographic images. This collaboration group has developed an integrated station that is available either to digitize the analogical images or to archive or to perform statistical analysis. Furthermore this prototype of station can represent also a very good system for mammographic educational programs. Using the whole database, several analysis can be performed by the MAGIC-5 tools [11]-[12].

The mammographic images (18x24 cm^2, digitized by a CCD linear scanner with a 85 μm pitch and 4096 gray levels) are fully characterized: pathological ones have a consistent description which includes the radiological diagnosis and the histological data, while non pathological ones correspond to patients with a follow up of at least three years [11]. The goal is the automated analysis of masses lesions, i.e. the search for objects in the image, usually characterized by peculiar shapes. Some example of masses lesions are made in figure 1.

In this paper a study of several representation of the dataset available is made. We report the results obtained with some classifiers based on Neural Network [13]-[14] like a Radial Basis Function, a Probabilistic Neural Network and a Multi Layer Perceptron in comparison with a K-Nearest Neighbours
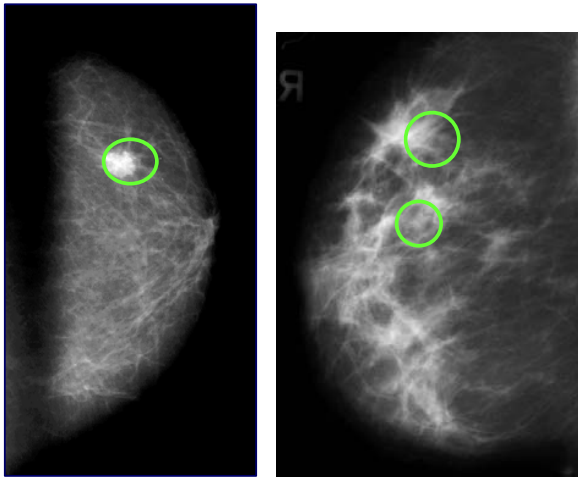
Fig. 1 Three examples of masses lesions in the original digital mammograms; the only masses present are signed in green circle

and a Support Vector Machine [15]-[23], used to select the pathological masses lesions from the region of interest (ROI).

II. METHODS

The CAD system here reported is an expert system based on three steps: a *ROI-hunter*, a *Features Extractor Module* and a *Classifier*. We also try a features reduction step before classication.

A. ROI-hunter

The ROI-hunter was already described in ref. [12]. The aim of this stage is to reduce the data amount to process by searching for Regions Of Interest (ROIs) that include a lesion with high probability. Only selected regions are stored for the next processing steps, rather than the whole mammogram as shown in Fig. 2.
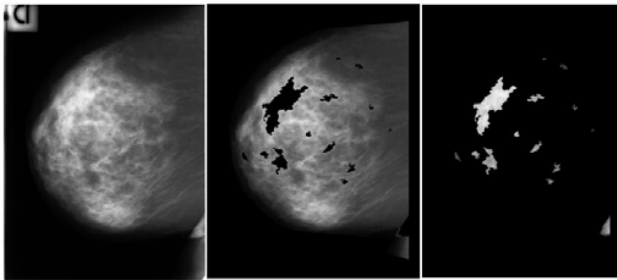

Fig. 2 The original mammogram (left), the remaining   image (middle), the selected patterns containing the ROIs (right)

The contour search is carried out by using the threshold operator. More generally this operator assigns the value I^+ to the pixel with the intensity above a pre-fixed threshold, and the value I^- to the pixel with the intensity less than the same threshold:

$$\text{if} \quad I_{x,y} \geq \text{threshold} \quad \Rightarrow \quad I_{x,y} = I^+$$
$$\text{if} \quad I_{x,y} < \text{threshold} \quad \Rightarrow \quad I_{x,y} = I^-$$

where $I_{x,y}$ is the intensity value for the pixel with coordinates (x,y).

One way to fix the threshold value is to draw an histogram of the pixel intensity on the whole image. Usually, this histogram has two peaks: the first one refers to the image background and the second one to the ROIs. The threshold may be chosen at the intensity corresponding to the crossover between the two peaks. This is a static selection criterion. In this paper we prefer to use a dynamical selection method as explained at point 3.

An iterative procedure (ROI Hunter), based on the search of relative intensity maximum inside a square window, has been implemented to select the ROIs. In the literature the masses lesions size is characterized by diameters varying in the approximate range 2 - 40 mm; in our case these two limits correspond to the square windows limit : A_{min} (25x25), A_{max} (501x501), in pixel. All the ROIs with area less than A_{min} are removed.

 The steps of the algorithm are:

1. *starting from the right top corner of the mammogram, a raster scanning is performed to find the coordinates (x_o,y_o) of an intensity maximum I_m (the initial centre of the candidate lesion). Its value is accepted if it is also a relative maximum in a box A_{min} (25x25 pixels);*

2. *an iso-intensity contour, including the relative maximum intensity pixel, is drawn at a threshold value $I_{th} = I_m / 2$; this contour defines a ROI with area A_R;*

3. *the threshold I_{th} is dynamically changed by increasing/decreasing its value if the Area A_R of the corresponding ROI is greater/smaller than the limit area A_{max} (501x501 pixels), until the difference between two consecutive thresholds is equal to one. At each step, the threshold is changed by an amount which is one half of the previous one.*

4. *the ROI is removed and stored for a further analysis; the corresponding "hole" left in the mammogram is set to zero;*

5. *go to step 1 to find next (x_o,y_o) coordinates of a relative intensity maximum.*

B. The Features Extractor Module

 In this paper twelve features are extracted from the segmented masses. The criteria for the features selection are based on morphological lesion differences [24]-[29]. For example the excessive lengthening is often symptom of pathology absence. In Table I the complete extracted features list is reported.

TABLE I

MORPHOLOGICAL FEATURES EXTRACTED FROM THE REGION OF INTEREST

Fractal index	Area
Eccentricity	Contour Gradient Entropy
Average Intensity	Standard Deviation of Intensity
Average Radial Length (ARL)	Standard Deviation of ARL
Entropy of intensity distribution	Anisotropy
Inertial Momentum	Circularity

The features extraction [12]-[16] plays a fundamental role in many pattern recognition tasks. Some features give geometrical information as eccentricity, area and average radial length; others provide shape parameters as fractal index and inertial momentum. In order to verify the feature discrimination capability between the two classes (pathologic or healthy patients), the feature value histograms are drawn. As an example in Fig. 3, 4 and 5 the histograms of average radial length, entropy of intensity distribution and circularity are shown.
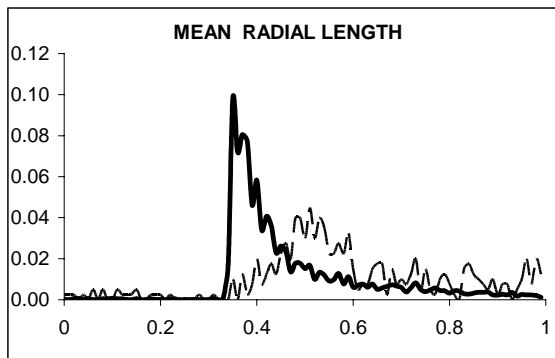

Fig. 3 "Mean radial length" feature distribution for pathological (dashed) and healthy (continuous) ROIs
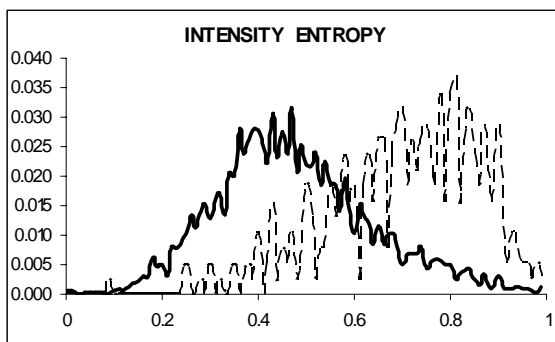

Fig. 4 "Entropy of intensity" feature distribution for pathological (dashed) and healthy (continuous) ROIs

The variability range varies from feature to feature and the healthy/pathological data histograms show a different shape. Moreover, the number of healthy ROI is about seven times bigger than the number of pathological ROI. For this reason each feature histogram is normalized both to its minimum and maximum values, and to the total ROI number in each class. A completely overlapped histogram for the healthy/pathological

classes, corresponds to a feature not being able to discriminate the presence of a pathology, while a reduced value of the overlapping area indicates an increased discrimination capability.
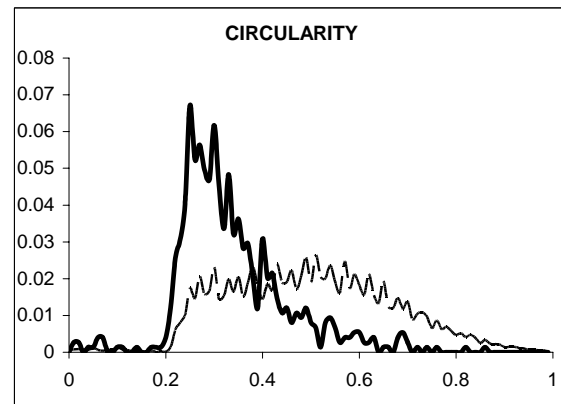

Fig. 5 "Circularity" feature distribution for pathological (dashed) and healthy (continuous) ROIs

More dataset extracted from the CALMA database [11] are reported in the table II. It is indicated the composition (positive samples vs total samples) of the training set, validation set and testing set.

TABLE II

COMPOSITION OF THE MAMMOGRAPICH DATASET

Dataset	# of samples (ROIs)	# of positive samples (ROIs)
Training set	4230	318
Validation set	4230	315
Testing set	4230	320

C. The features reduction

A study of the representation space is made through two types of Component Analysis [15]-[16]: Principal Component Analysis (PCA) and the Independent Component Analysis (ICA). The PCA algorithm uses the Karhunen Loève transformation to reduce the features with the main eigenvalues of the covariance matrix of the mean vectors. While PCA seek directions in a features spaces that best represent the data in a sum squad error sense, ICA instead seeks direction that are most independent from each other. In particular a Fast Independent Component Analysis (FASTICA) is used to provide a computationally quick method to estimate the unobserved independent components [30]. It is a method to decompose a multi-dimensional dataset into a set of statistically independent non-gaussian variables. The algorithm iteratively maximises an approximation to the negentropy of the projected data. Negentropy is based on the information-theoretic quantity of entropy which measures the "randomness" of an observed variable. Since gaussian variables have the largest entropy among all random variables

of equal variance entropy can be used to define a measure of non-gaussianity i.e. negentropy. In practice this quantity can be time consuming to calculate.

In the figures 6 and 7 the original features space is plotted. We show only some examples of significant projection of the multidimensional spaces (original features of input).

Although PCA algorithm found the principal component, the population of positive samples and negative samples remain rather overlapped after the transformation. In the figures 8 and 9 are shown some projection of the PCA representation space.
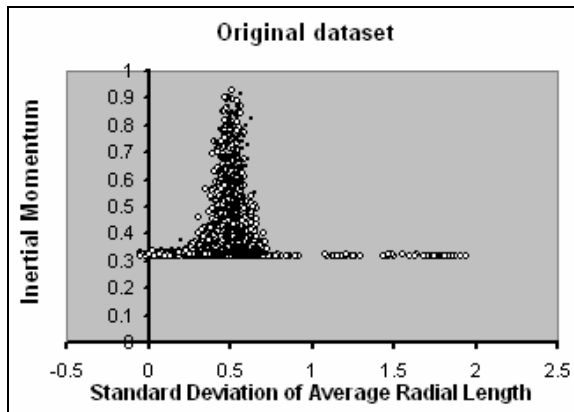

Fig. 6 Feature representation of the original dataset for pathological (black points) and healthy (white points) ROIs
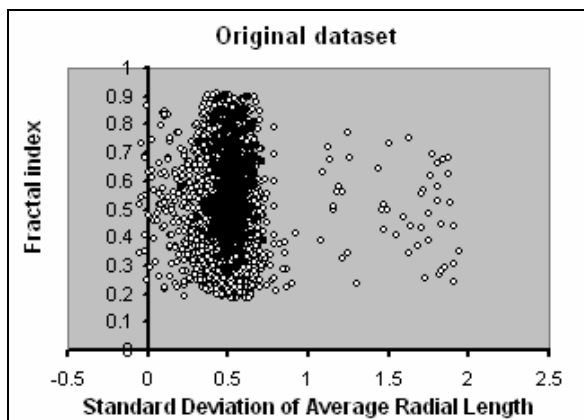

Fig. 7 Feature representation of the original dataset for pathological (black points) and healthy (white points) ROIs
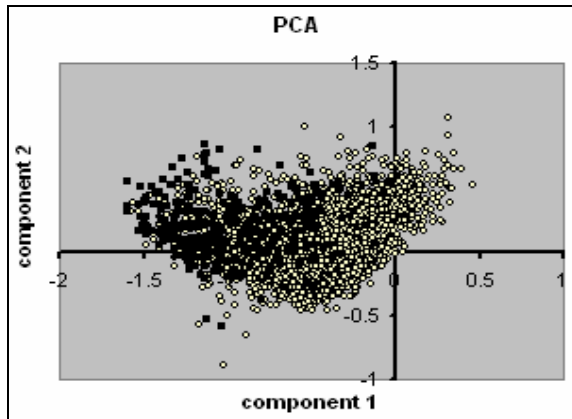

Fig. 8 Feature representation of the dataset through Principal Component Analysis for pathological (black points) and healthy (white points) ROIs
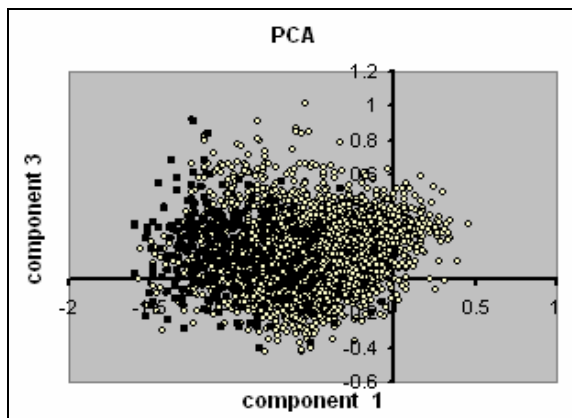

Fig. 9 Feature representation of the dataset through Principal Component Analysis for pathological (black points) and healthy (white points) ROIs

Typical algorithms for ICA use centering, whitening and dimensionality reduction as preprocessing steps in order to simplify and reduce the complexity of the problem for the actual iterative algorithm. By using FASTICA on the same dataset we optimized the algorithm with a preliminary features reduction through Whitening method [15]; just an example of this first reduction is shown in figure 10.
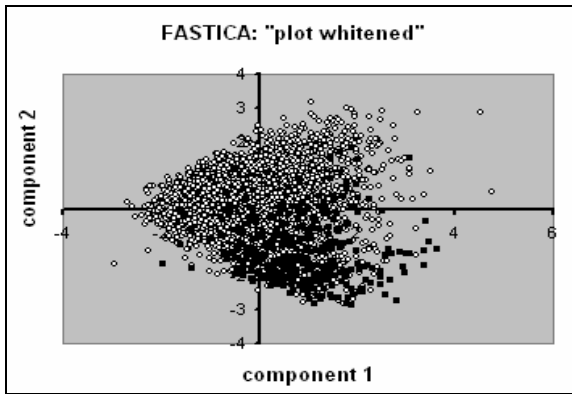
Fig. 10 Feature representation of the dataset through Whitening method before FASTICA algorithm for pathological (black points) and healthy (white points) ROIs

Then the independent component are found optimizing FASTICA parameters [30] but also in this case positive samples and negative samples remain rather overlapped after the transformation; the example is shown in figure 11.
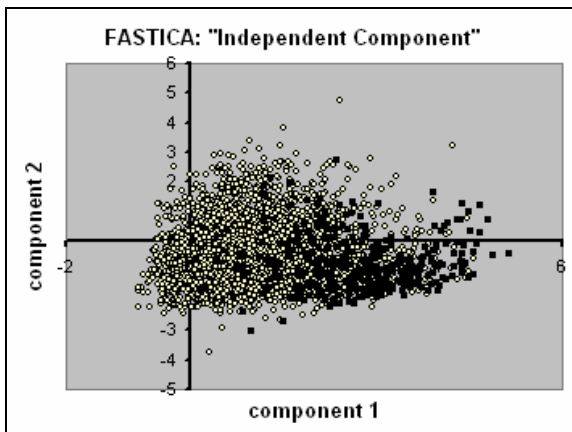

Fig. 11 Feature representation of the dataset through Fast Independent Component Analysis for pathological (black points) and healthy (white points) ROIs

D. The Classifier

We make a comparative study of three different neural networks, a deterministic classifier and a classifier that realize an optimal evolution of the theory of linear discriminative functions; so we present the following classifiers:

- Multi Layer Perceptron (MLP). The selected MLP is a feed-forward back-propagation supervised neural network trained with gradient descent learning rule with "momentum", so as to quickly move along the direction of decreasing gradient, thus avoiding oscillations around secondary minima.

- Probabilistic Neural Network (PNN). This network provides a general solution to pattern classification problems by following an approach developed in statistics, called Bayesian classifiers. The probabilistic neural network uses a supervised training set to develop distribution functions within a pattern layer. These functions, in the recall mode, are used to estimate the likelihood of an input feature vector being part of a learned category, or class. The learned patterns can also be combined or weighted with the a priori probability, also called the relative frequency, of each category to determine the most likely class for a given input vector.

- Radial Basis Function (RBF) network. It has a static gaussian function as the nonlinearity for the hidden layer processing elements. The gaussian function responds only to a small region of the input space where the gaussian is centered. The key for a successful implementation of these networks is to find suitable centers for the gaussian functions. This can be done with a supervised learning, but an unsupervised approach usually produces better results. For this reason we implement RBF networks as a hybrid supervised-unsupervised topology.

- K-Nearest Neighbors (K-NN) classifier. For this type of deterministic classifier it is necessary to have a training set which is not too small and a good discriminating distance. KNN performs well in multi-class simultaneous problem solving. There exists an optimal choice for the value of the parameter K which brings to the best performance of the classifier. This value of K is often approximately close to $N^{1/2}$.

- The SVM algorithm creates a hyperplane that separates the data into two classes with the maximum-margin. Given training examples labeled either "yes" or "no", a maximum-margin hyperplane is identified when it splits the "yes" from the "no" training examples, such that the distance between the hyperplane and the closest examples (the margin) is maximized. There is a way to create non-linear classifiers by applying the kernel trick to maximum-margin hyperplanes. The resulting algorithm is formally similar, except that every dot product is replaced by a non-linear kernel function. This allows the algorithm to fit the maximum-margin hyperplane in the transformed feature space. The transformation may be non-linear and the transformed space high dimensional; thus though the classifier is a hyperplane in the high-dimensional feature space it may be non-linear in the original input space.

III. RESULTS

For each classifier the training and validation sets are used for the optimization phase. The feature reduction through PCA allows obtaining same results with respect to the original dataset but it reduces slightly the computing time. Using FASTICA we have not good performance, so this method is not suitable. In both cases, classifiers on raw data give better performances compared to PCA and FASTICA. This is a

strong argument against using component analysis for this dataset.

Using sensitivity (percentage of pathologic ROIs correctly classified) and specificity (percentage of healthy ROIs correctly classified), the results obtained with this analysis are described in terms of the ROC (Receiver Operating Characteristic) curve [31]-[32], which shows the true positive fraction (sensitivity), as a function of the false positive fraction (1-specificity) obtained varying the threshold level of the ROI selection procedure. In this way the ROC curve allows to the radiologist to detect masses lesions with predictable performance, so that he can set the CAD sensitivity value.

In particular we obtain for the classifiers the following configuration:

- KNN is optimized for K = 21 on validation set; furthermore an additional threshold allow to obtain the ROC curve.

- SVM supplies the best performances for a sigmoidal kernel

- MLP has 12 hidden neurons for the 12 input previously described.

- RBF has 40 neurons in the radial layer with 1.9 of spread parameter

- PNN supplies best performances with 0.24 of spread parameter

The results of the classifiers are supplied in the diagram through the ROC curve calculated on the testing set after the optimization through PCA on the training set-validation set.

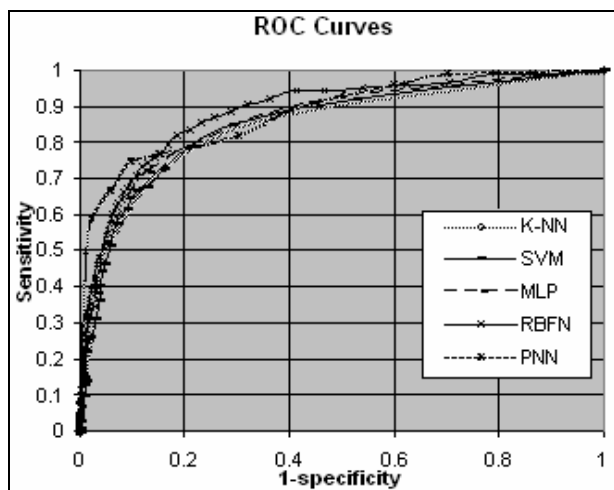Finally in Fig. 12 are shown the results of the classifiers on the testing set.

Fig. 12 Receive Operating Characteristic curves for several classifiers tried

Also the area under the curve [21]-[22], obtained in relation to the same ROC curves calculated on the test values, are reported in Table III.

TABLE III
PERFORMANCE OF THE CLASSIFIERS IN TERMS OF AREA UNDER THE ROC CURVE S

Classifiers	Area under ROC curves	Error
KNN	0.81	0.01
SVM	0.81	0.01
MLP	0.88	0.01
PNN	0.89	0.01
RBF	0.89	0.01

The results of the Table III show that the neural net (RBF, PNN) has better performances then the other classifiers for the dataset considered. A study carried out on the complementariness of the classifiers used on the dataset in under examination show (in the case of the features previously used) that the regions of decision of the three classifiers are overlapped. This fact and the better performance in comparison to the other two classifiers indicate that it is not possible in this case to combine the output of the various classifiers with techniques of multi classification system (MCS) to improve the total performances.

IV. CONCLUSIONS

In this paper a comparison of some classification system for masses lesions classification has been presented. The features are extracted through an algorithm based on morphological lesion differences. The features are used to discriminate two classes (pathological or healthy ROIs).

Principal component analysis does not reduce appreciably computing time and it does not improve the classifiers accuracy. Instead FASTICA get worse the results. In conclusion component analysis is not suitable for own dataset.

The discriminating performances of the algorithm were checked by means of a supervised neural network against other classifiers as a Support Vector Machine and a K-Nearest Neighbours and the results have been presented in terms of ROC curve. The best results are obtained with Radial Basis Function and a Probabilistic Neural Network that performs as well as a Multi Layer Perceptron. So the results show superior performances of the Neural Network on the masses lesions classification through morphological lesion differences.

The results are comparable or better than those obtained in other recent studies [11],[33]-[34] verifying that a neural network system, based on original features spaces, provides a better ability to distinguish pathological ROIs from the healthy ones.

REFERENCES

[1] Smith R.A., "Epidemiology of breast cancer", in "A categorical course in physics. Imaging considerations and medical physics responsibilities", Madison, Wisconsin, Medical Physics Publishing, 1991.

[2] Peto R., Boreham J., Clarke m., Davies c., Deral V, correspondence "UK and USA Breast cancer deaths down 25% in year 2000 at ages 20-69 years", LANCET 2000, 355, (9217) pp. 1822-1823, 2000.

[3] Bird R., Wallace T., Yankaskas B., "Analysis of cancer missed at screening mammography", Radiology 1992: 184; pp.613-617, 1992.

[4] Haus A.G. and Yaffe M. (editors), "A categorical course in physics. Technical aspects in breast imaging". Radiological Society of North America, Presented at the 79th Scientific Assembly and Annual Meeting of RSNA, 1993.

[5] Feig S.A. , Yaffe M. , "Digital mammography, computer aided diagnosis and telemammography", Radiol. Clin. N. Am. 33, 1205-1230, 1995.

[6] Keddache S., Thilander-Klang A., Lanhede B. , "Storage phosphor and film screen mammography: performance with different mammographic techniques" Eur. Radiol. 9, 591-597, 1999.

[7] Schmidt R.A., Nishikawa R.M. , "Clinical use of digital mammography: the presents and the perspects", Digit. Imaging 8/1 suppl. 74-79, 1995.

[8] Karssemejer N., "A stochastic method for automated detection of microcalcifications in digital mammograms" in Information processing in medical imaging, Springer-Verlag New York, 227-238, 1991.

[9] Karssmejer N. , "Reading screening mammograms with the help of neural networks", Nederlands Tijdschriff geneeskd, 143/45, 2232-2236, 1999.

[10] 10.C.J. Viborny, M.L. Giger, "Computer vision and artificial intelligence in mammography", AJR 162, 699-708, 1994.

[11] Bottigli U, Delogu P, Fantacci ME, Fauci F, Golosio B, Lauria A, Palmiero R, Raso G, Stumbo S, Tangaro S Search of Microcalcification clusters with the CALMA CAD station. The International Society for Optical Engineering (SPIE) 4684: 1301-1310, 2002

[12] F. Fauci, S. Bagnasco, R. Bellotti, D. Cascio, S. C. Cheran, F. De Carlo, G. De Nunzio, M. E. Fantacci, G. Forni, A. Lauria, E.Lopez Torres, R. Magro, G. L. Masala, P.Oliva, M. Quarta, G. Raso, A. Retico, S.Tangaro, Mammogram Segmentation by Contour Searching and Massive Lesion Classification with Neural Network, Proc. IEEE Medical Imaging Conference, October 16-22 2004, Rome, Italy; M2-373/1-5, 2004.

[13] S. Haykin "Neural Networks – A comprehensive foundation", second edition, Prentice Hall, 1999.

[14] Massimo Buscema & Semeion Group, "Reti Neurali artificiali e sistemi sociali complessi", volume 1 Teoria e modelli 1409.1, Franco Angeli, 1999.

[15] S. Serpico, G. Vernazza, "Teorie e tecniche del riconoscimento", CUSL "Il gabbiano", 1997.

[16] O. Duda, P. E. Hart, D. G. Stark, "Pattern Classification", second edition, A Wiley-Interscience Publication John Wiley & Sons, 2001.

[17] S. J. Russel, P.Norvig, "Artificial Intelligence. A modern approach", UTET, 1998.

[18] V. N. Vapnik. "Statistical Learning Theory. Wiley", New York , 1998.

[19] M. Pontil, A. Verri "Properties of Support Vector Machines", Neural Computation, Vol. 10, pp 955-974, 1998.

[20] N. Cristianini, J. Shave-Taylor. "An Introduction to Support Vector Machine"(and other kernel-based learning methods). Cambridge University Press 2000.

[21] SVM_light software is available in the following location : ftp://ftp-ai.cs.unidortmund.de/pub/Users/thorsten/svm_light/current/ svm_light.tar.gz

[22] T. Joachims, Text Categorization with Support Vector Machines: Learning with Many Relevant Features, Proc. 10th European Conf. Machine Learning (ECML), Springer-Verlag, 1998.

[23] T. Mitchell "Machine Learning" , McGraw-Hill 1997.

[24] Timp S., Karssemeijer N., A new 2D segmentation method based on dynamic programming applied to computer aided detection in mammography, Medical Physics: 31; 958-971, 2004.

[25] Baydush A.H., Catarious D.M., Abbey C.K., Floyd C.E., Computer aided detection of masses in mammography using subregion Hotelling observers, Medical Physics: 30; 1781-1787, 2003.

[26] Tourassi G.D., Vargas-Voracek R., Catarious D.M. Jr, Floyd C.E. Jr, Computer-assisted detection of mammographic masses: A template matching scheme based on mutual information, Medical Physics: 30 (8); 2123-2130, 2003.

[27] Antonie M.L., Zaiane O.R., Coman A., Application of data mining techniques for medical image classification, Proc. of II Int. Work. On Multimedia Data Mining, USA, 2001.

[28] Vyborny CJ., Giger ML., Computer vision and artificial intelligence in mammography, AJR: 162; 699-708, 1994.

[29] Lai S., Li X., Bischof W., On techniques for detecting circumscribed masses in mammograms", IEEE Transaction on Medical Imaging: 8(4); 377-386, 1989.

[30] Aapo Hyv¨arinen, Erkki Oja, "Indipendent Component Analysis: Algorithms and Applications", Neural Networks Research Centre, Helsinki University of Technology, Finland, "Neural Networks", 13 (4-5):411-430, 2000.

[31] Hanley JA, McNeil B, The meaning and use of the area under a receiver operating characteristic (ROC) curve, Radiology: 143; 29-36, 1982.

[32] Hanley JA, McNeil B, A method of comparing the areas under receiver operating characteristic curves derived from the same cases, Radiology: 148; 839-843, 1983.

[33] U. Bottigli, B. Golosio, G. L. Masala, P. Oliva, S. Stumbo, D. Cascio, F. Fauci, R. Magro, G. Raso, R. Bellotti, F. De Carlo, S.Tangaro, I. De Mitri, G. De Nunzio, M. Quarta, A. Preite Martinez, P. Cerello, S. C. Cheran, E.Lopez Torres "Dissimilarity Application for Medical Imaging Classification" on proceedings of The 9th World Multi-Conference on Systemics, Cybernetics and Informatics WMSCI 2005, Orlando 10-13 July 2005, vol III pag 258-262, 2005.

[34] G. Masala, B. Golosio, D. Cascio, F. Fauci, S. Tangaro, M. Quarta, S. Cheran, E. L. Torres, "Classifiers trained on dissimilarity representation of medical pattern: a comparative study" on Nuovo Cimento C, Vol. 028, Issue 06, pp 905-912 , 2005.