WBCD breast cancer database classification applying artificial metaplasticity neural network

A. Marcano-Cedeño *, J. Quintanilla-Domínguez, D. Andina

Group for Automation in Signals and Communications, Technical University of Madrid, Madrid, Spain

ARTICLE INFO

Keywords:
Neural network
Backpropagation algorithm
Pattern recognition
Metaplasticity
Breast cancer
AMMLP

ABSTRACT

The correct diagnosis of breast cancer is one of the major problems in the medical field. From the literature it has been found that different pattern recognition techniques can help them to improve in this domain. These techniques can help doctors form a second opinion and make a better diagnosis. In this paper we present a novel improvement in neural network training for pattern classification. The proposed training algorithm is inspired by the biological metaplasticity property of neurons and Shannon’s information theory. During the training phase the Artificial metaplasticity Multilayer Perceptron (AMMLP) algorithm gives priority to updating the weights for the less frequent activations over the more frequent ones. In this way metaplasticity is modeled artificially. AMMLP achieves a more efficient training, while maintaining MLP performance. To test the proposed algorithm we used the Wisconsin Breast Cancer Database (WBCD). AMMLP performance is tested using classification accuracy, sensitivity and specificity analysis, and confusion matrix. The obtained AMMLP classification accuracy of 99.26%, a very promising result compared to the Backpropagation Algorithm (BPA) and recent classification techniques applied to the same database.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

The correct patterns classification of breast cancer is an important real-world medical problem. Breast cancer has become one of the major causes of mortality around the world and research into cancer diagnosis and treatment has become an important issue for the scientific community. The etiologies of breast cancer remain unclear and no single dominant cause has emerged (Ioanna, Evalgelos, & George, 2000; Rodrigues, Chang, & Suri, 2006). Prevention is still a mystery and the only way to help patients survive is by early detection. If the cancerous cells are detected before spreading to other organs, the survival rate for patients is more than 97% (American Cancer Society Hompage, 2008). A major class of problems in medical science involves the diagnosis of disease, based upon various tests performed upon the patient. For this reason the use of classifier systems in medical diagnosis is gradually increasing. There is no doubt that evaluation of data taken from patients and decisions of experts are the most important factors in diagnosis. But, the different artificial intelligence techniques for classification also help experts a great deal. Classification systems, minimizing possible errors that might be made because of fatigued or inexperienced experts, provide more detailed medical data for examination in a shorter time (Subashini, Ramalingam, & Palanivel, 2009).

There has been research on medical diagnosis of breast cancer with WBCD using Artificial Neural Networks (ANNs) in literature, and most has reported high classification accuracy. Quinlan (1996) reached 94.74% classification accuracy using 10-fold cross-validation with the C4.5 decision tree method. Hamilton, Shan, and Cercone (1996) obtained 94.99% accuracy with the RIAC method, while Ster and Dobnikar (1996) obtained 96.8% with the linear discreet analysis method. The accuracy obtained by Nauck and Kruse (1999) was 95.06% with neuron-fuzzy techniques. In Pena-Reyes and Sipper (1999), the classification technique used the fuzzy-GA method, reaching a classification accuracy of 97.36%. In Setiono (2000), the classification was based on a feed forward neural network rule extraction algorithm. The reported accuracy was 98.10%. In Albrecht, Lappas, Vinterbo, Wong, and Ohno-Machado (2002), a learning algorithm that combined logarithmic simulated annealing with the perceptron algorithm was used and the reported accuracy was 98.8%. In Abonyi and Szefiert (2003), an accuracy of 95.57% was obtained with the application of the supervised fuzzy clustering technique. In Übeyli (2007) five different classifiers, support vector machine, probabilistic neural network, recurrent neural network, combined neural network and Multilayer perceptron neural networks, were applied and respective accuracies of 99.54%, 98.61%, 98.15%, 97.40% and 91.92% were obtained. In Polat and Güneş (2007), Least Square
SVM (LS-SVM) was used and 98.53% accuracy was obtained. Gujjarro-Berdias, Fontenla-Romero, Perez-Sanchez, and Fraguela (2007) presented a learning algorithm applying linear-leastsquares, reaching a classification accuracy of 96.0% over the entire WBCD. Akay (2009) reached 99.51% classification accuracy using a SVM-based method combined with feature selection. Karabatak and Cevdet-Ince (2009), presented an automatic diagnosis system for detecting breast cancer based on Association Rules (AR) and Neural Networks (NNs), obtaining a classification accuracy of 97.4%. The accuracy obtained by Peng, Yang, and Jiang (2009) was 99.5% applying a hybrid method that combines filter and wrapper methods.

In this study, the AMMLP algorithm is proposed for classifying the breast cancer lesions as benign or malignant. This method consists of simulating the biological property of metaplasticity on MLP with Backpropagation (Andina et al., 2009).

We modeled this interpretation on the NNs training phase. Our AMMLP algorithm has been compared with a classical Backpropagation Algorithm (BPA) as well as with recently proposed algorithms applied on the WBCD database. Our test results prove our method to be a superior or at least an interesting alternative.

### 2. Wisconsin breast cancer database overview

Breast cancer is a malignant tumour that develops from cells of the breast. Although scientists know some of the risk factors (i.e. aging, genetic risk factors, family history, menstrual periods, not having children, obesity) that increase a woman's chance of developing breast cancer, they do not yet know what causes most breast cancers or exactly how some of these risk factors cause cells to become cancerous. Research is underway to learn more and scientists are making great progress in understanding how certain changes in DNA can cause normal breast cells to become cancerous. (Jerez-Aragones, Gomez-Ruiz, Ramos-Jimenez, Munoz-Perez, & Alba-Conejo, 2003).

In this study, the Wisconsin Breast Cancer Database (UCI Machine Learning Repository) was analyzed. The WBCD dataset consists of 699 instances taken from Fine Needle Aspirates (FNA) of human breast tissue. Each record in the database has nine attributes. The nine attributes are detailed in Table 1. The measurements are assigned an integer value between 1 and 10, with 1 being the closest to benign and 10 the most anaplastic. Associated with each sample is its class label, which is either benign or malignant. This dataset contains 16 instances with missing attribute values. Since many classification algorithms have discarded these data samples, for the ease of comparison, the same method is followed and the remaining 683 samples are taken for use. Therefore, the class has a distribution of 444 (65.0%) benign samples and 239 (35.0%) malignant samples.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Attribute description</th>
<th>Values of attribute</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Clump thickness</td>
<td>1–10</td>
<td>4.44</td>
<td>2.83</td>
<td></td>
</tr>
<tr>
<td>2 Uniformity of cell size</td>
<td>1–10</td>
<td>3.15</td>
<td>3.07</td>
<td></td>
</tr>
<tr>
<td>3 Uniformity of cell shape</td>
<td>1–10</td>
<td>3.22</td>
<td>2.99</td>
<td></td>
</tr>
<tr>
<td>4 Marginal adhesion</td>
<td>1–10</td>
<td>2.83</td>
<td>2.86</td>
<td></td>
</tr>
<tr>
<td>5 Single epithelial cell size</td>
<td>1–10</td>
<td>2.23</td>
<td>2.22</td>
<td></td>
</tr>
<tr>
<td>6 Bare nuclei</td>
<td>1–10</td>
<td>3.54</td>
<td>3.64</td>
<td></td>
</tr>
<tr>
<td>7 Bland chromatin</td>
<td>1–10</td>
<td>3.45</td>
<td>2.45</td>
<td></td>
</tr>
<tr>
<td>8 Normal nucleioli</td>
<td>1–10</td>
<td>2.87</td>
<td>3.05</td>
<td></td>
</tr>
<tr>
<td>9 Mitoses</td>
<td>1–10</td>
<td>1.60</td>
<td>1.73</td>
<td></td>
</tr>
</tbody>
</table>

### 3. Brief review of artificial neural networks

#### 3.1. Artificial neural networks

ANNs are biologically inspired and mimic the human brain. The neurons are interconnected with connection links which have weights which are multiplied by the signal transmitted in the network. The output of each neuron is determined by using an activation function such as sigmoid and step. Usually nonlinear activation functions are used. NN’s are trained by experience. When an unknown input to the network is applied, it can generalize from past experiences and produce a new result (Andina, 2007; Haykin, 1994; Hagan, Demuth, & Beale, 1996; Kandel, Schwartz, & Jessell, 2000).

Feedforward neural networks are a basic type of ANN capable of approximating generic classes of functions, including continuous and integrable functions. An important class of feedforward neural networks are MultiLayer Perceptron Neural Networks (MLPNNs). The MLPNNs have features such as the ability to learn and generalize, smaller training set requirements, fast operation, and ease of implementation. Therefore, they are the most commonly used neural network architectures (Basheer & Hajmeera, 2000; Chaudhuri & Bhattacharya, 2000; Haykin, 1994).

BPA is one of the most popular training algorithms for MLP. Unfortunately BPA showed some limitations and problems during MLP training. A serious drawback of BPA (Rumelhart, Hinton, & Williams, 1986) for conventional MLP training is its slow rate of convergence and, in the search for the global minimum, the risk it runs of being trapped in a local minimum. Several modifications (Jacobs, 1998; Kruschke & Movellan, 1991; Pingzhou & Zhaoai, 2008; Zhang, Zhang, & Wu, 2009) of the original BPA have already been suggested to improve either the convergence or the performance over the original algorithm. However, none of the modifications is capable of delivering satisfactory performance for all problems, in general. Thus, the search for an approach to speed-up its convergence and/or for the improvement of general performance of the trained network still remains important. For this reason we present a novel proposal for improved BPA based on metaplasticity. In Section 5 we will describe the above algorithm.

### 4. Metaplasticity

The metaplasticity concept was defined by Abraham (1996) and is a biological concept widely known in the fields of biology, medical computer science, neuroscience, physiology, neurology and others (Abraham, 1999; Andina, Jevtic, Marcano-Cedeño, & Barrón-Adame, 2007; Jedlicka, 2002; Kinto, Del-Moral-Hernandez, Marcano-Cedeño, & Ropero-Pelaez, 2007; Ropero-Pelaez, Piqueira, Ropero-Pelaez, & Piqueira, 2007). The prefix “meta” comes from Greek and means “beyond” or “above”, and in neuroscience and others fields it is used to indicate a higher level of plasticity, expressed as a change or transformation in the way synaptic efficacy is modified. Metaplasticity is defined as the induction of synaptic changes also depending on prior synaptic activity (Abraham & Bear, 1996; Abraham, 1999). Metaplasticity is due, at least in part, to variations in the level of postsynaptic depolarization for inducing synaptic changes. These variations facilitate synaptic potentiation and inhibit synaptic depression in depressed synapses and vice versa in potentiated synapses (the direction and the degree of the synaptic change are a function of postsynaptic depolarization during synaptic activation: long-term potentiation (LTP) is obtained following low levels of postsynaptic depolarization, whereas long-term depression (LTD) is produced by stronger depolarizations).

The induction of synaptic change variation in levels of neural activity is explained in Fig. 1 (Ropero-Pelaez et al., 2007). On the
aptic weight, the curve is elongated so that the LTP threshold value.

Graphs illustrate this idea. For higher values of initial synapse weight, the curve moves to the right, reinforcing LTP. This graphical shows a family of curves in which each curve yields the variation of weight given the neuron's activation. The parameter that defines what weights are the parameters that play the most relevant role in ANN learning and performance.

The fundamental ideal of BPA is to minimize the error function which is linked to the update of weights, that: If the strategy of an ANN learning procedure is to minimize an expected error $E_M$ defined by the following expression:

$$E_M = \mathbb{E}[E(x)] dx$$

where $x$ is a random variable of training input vectors $x = (x_1, x_2, \ldots, x_n)$, where $R^n$ is the $n$-dimensional space and $E(x)$ is the expression of a given error criterion as a function of the inputs applied in ANN training to update its weights in each training iteration step, then we can perform the following manipulation:

$$E_M = \int_{R^n} E(x)f_X(x) dx = \int_{R^n} e(x)dx$$

and calculate $E_M$ through the following estimator:

$$E_M = \frac{1}{M} \sum_{i=1}^M \frac{e(x_i)}{f_X(x_i)}$$

where $x_i$, $k = 1, 2, \ldots, M$, are independent sample vectors whose Probability Density Function (pdf) is $f_X(x)$ which can be arbitrarily chosen by the designer if $f_X(x) \neq 0$, wherever $e(x) \neq 0, \forall x \in R^n$. Note that from Eq. (3) $f_X(x)$ is ideally given by:

$$(f_X(x))_{opt} = \frac{1}{E_M} e(x)$$

In Eq. (5), we evaluate $E_M$ with just one realization. Obviously, in practice it is impossible to know $f_X(x)_{opt}$, but we can use a suboptimum function. All we have to find is a function whose effect in (4) is the improvement of the training convergence (Andina, Martinez-Antorrena, & Melgar, 2004). The closer to the optimum, the better the results that should be achieved.

What Eq. (4) expresses is that an error objective function $E(x)$ can be weighted by a proper function without affecting the final error objective for each class.

6. Artificial metaplasticity implementation in MLP training

In the case of MLP trained with BPA, it has been shown that the output for each class is the MLP inherent estimation of the a posteriori probability of the class (Rucky, Rogers, Kabrisk, Oxley, & Suter, 1990). This allows a straight implementation of metaplasticity implementation. In an MLP applied to classification of $L$ classes $H_i, i = 0, 1, \ldots, L – 1$, based on Bayes Theorem we get

$$y_i = P(H_i|x) = \frac{f_X(x|H_i)P(H_i)}{f_X(x)}$$

We can straightforwardly assume that $f_X(x|H_i) = f_X(x)$ in Eq. (4), which, for each class $l$ becomes

![Fig. 1. Changes of synaptic strength due to the postsynaptic activity in biological neurons. If postsynaptic activity is high, the curve will move to the right, reinforcing the LTP.](image1)

![Fig. 2. Show the idea of the metaplastic intuitively.](image2)
have assumed that weights in each iteration step using the following weight function: 

\[
\bar{E}_M = \frac{1}{M} \sum_{k=1}^{M} E(x_k|f_k(x)/H_k) \approx \frac{1}{M} \sum_{k=1}^{M} E(x_k) \frac{y_k}{P(H_k)}
\]  

(7)

and implement Eq. (4) by including the suboptimum function \(f_k^s\) in the error BPA equations:

\[
\delta(W) = \frac{\partial}{\partial W_{ij}} \left(1 - \left(1 - \frac{y}{\hat{y}}(s)\right)^3\right) - \frac{1}{f_k(x)} \frac{\partial}{\partial W_{ij}} W_{ij}(x)
\]

(8)

\[
\delta_j^s = \left( y - \hat{y}(s) \right) \frac{\delta_j^s}{f_k(x)}
\]

where \(s\) is the layer counter, \(s = 1, 2, \ldots, S, j\) and \(i\) the node and input counters respectively, and backpropagate the error to the other layers as usual (Andina et al., 2009). So, in the end, in the overall algorithm, the AMP is included in the training algorithm by affecting the weights in each iteration step using the following weight function:

\[
w^s(x) = \frac{A}{\sqrt{2\pi} \sigma} e^{-\frac{x^2}{2\sigma^2}}
\]

(10)

where \(N\) is the number of components of input vector \(X\) that feed the first hidden layer (for the second hidden layer, \(X\) is substituted by the first hidden layer output vector, and so on) and \(A, B\) parameters that will be estimated empirically \((A, B \in R^*)\). Note that we have assumed that a posteriori probabilities are well estimated. This diverges from reality, in which the training cannot even converge. This is very important in the beginning of training, where the output of the networks still has no valid statistical estimation from the ANN. So, although approximation given by Eq. (9) becomes more precise as learning proceeds, the assumption made by Shannon (1948) is recommendable for the first training iterations where assumption Ropero-Pelaez et al. (2007) is not valid and may result in convergence problems (Andina et al., 2009).

We have formulated the premise that, by giving more relevance to weight update for the less frequent activations over the more frequent ones, metaplasticity is being artificially modeled. This premise connects AMMLP with the information theory entropy concept: less frequent events carry more information than frequent ones. The proposed learning causes AMMLP to gain efficiency, reducing the number of iterations necessary to complete the training as compression reduces the number of bits necessary to codify information.

6.1. The AMMLP algorithm

1. Network structure used in the experiments:

(a) Number of input neurons equal to the number of attributes of database records (plus the bias input).
(b) Number of hidden layers: 1.
(c) Hidden neurons: 8.
(d) Output neurons: 1 (all classifications present two classes).
(e) Learning rate \(\eta = 1\).
(f) Activation function is sigmoidal with value between [0, 1].

2. Initialize all weights in weight matrix \(W\) randomly

3. Training phase

(a) AMP is modeled by applying the weight function in Eq. (10) to the BP weights updating during learning:

\[
\omega^0_j(t + 1) = \omega^0_j(t) + \eta \cdot \delta^j(t) \cdot \hat{y}^j(t) - 1
\]

(11)

\[
\delta^j(t) = (y - \hat{y}^j(t)) \cdot \delta^j(t) \cdot (1 - \hat{y}^j(t)) / f_k^s(x)
\]

(12)

where \(\omega^0_j\) are the weights of the \(j\) artificial neurons in layer \(l\) during iteration \(t\), being \(\hat{y}^j(t)\) the outputs of the \(j\) neurons in previous layer \(x\) (for the first hidden layer), and \(\delta^j(t)\) the usual error term backpropagated in BP, that for the sigmoidal activation functions case and output layer \(L\) follows the Eq. (12), where \(y\) is the desired output.

(b) Test training conditions

(i) if epochs = 2000 stop training
(ii) if Mean Squared Error, MSE = 0.01 stop training.

7. Results and discussion

The AMMLP proposed as a classifier was implemented in MATLAB® (software MATLAB version 7.4, R2007a) and on a Pentium IV computer of 3.4 GHz with 2 GB of RAM. This algorithm was applied to the Wisconsin Breast Cancer Database (WBCD).

Table 2 shows the network structure, metaplasticity parameters, epochs, MSE and numbers of patterns used in the training and testing phases.

The activation function is sigmoidal with scalar output in the range \([0, 1]\) and it is the same for all the neurons. To comparatively evaluate the performance of the classifiers, all the classifiers presented in this particular case were trained with the same training data set and tested with the same evaluation data set. The network was trained with 60% of the data, 410 samples, of which 144 were malignant and 266 benign records. The testing set, the remaining 40% of data, consisted of 233 samples of which 95 were malignant and 178 benign records.

For the experiments, we generated 100 AMMLPs with different weights whose values were random with normal distribution (mean 0 and variance 1). In each experiment 100 networks were trained in order to achieve an average result that did not depend on the initial random value of the weights of the ANN. Two different criterions were applied to stop the training: in one case it was stopped when the error reached 0.01 (the error reduces but cannot converge to 0), and in the other the training was conducted with a fixed number of 2000 epochs.

7.1. Performance evaluation methods

In this section we present results of experiments to test the behavior of the AMMLP proposed method, as well as compare it with classical Backpropagation MLP.

To measure the performance of the breast cancer diagnosis of the classifiers used in this investigation we split the evaluation it into two parts: one to determine performance result accuracies, related to classification accuracy, analysis of sensitivity and specificity, and confusion matrix; the other for performance results ROC, related to ROC curve analysis and area under the curve (AUC). We explain the methods used in each part in the following sections.

<table>
<thead>
<tr>
<th>Types</th>
<th>Network structure</th>
<th>MSE</th>
<th>Epochs</th>
<th>Metaplastic. parameters</th>
<th>Numbers patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMMLPs</td>
<td>I, H, L, O</td>
<td>0.01</td>
<td>2000</td>
<td>A</td>
<td>410</td>
</tr>
<tr>
<td>BPNNs</td>
<td>9, 8</td>
<td>0.01</td>
<td>2000</td>
<td>NA*</td>
<td>410</td>
</tr>
</tbody>
</table>

* NA: not apply.
7.1.1. Performance result accuracy

- Classification accuracy: In this study, classification accuracy for the data sets are measured using the equation:

\[
\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN}
\]

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively.

- True positive (TP): An input is detected as a patient with breast cancer, as diagnosed by the expert clinicians.
- True negative (TN): An input is detected as normal and also labeled as a healthy person by the expert clinicians.
- False positive (FP): An input is detected as a patient with breast cancer, although labeled as a healthy person by the expert clinicians.
- False negative (FN): An input is detected as normal, although diagnosed by the expert clinicians as having breast cancer.

- Sensitivity and specificity: For sensitivity and specificity analysis, we use the following expressions.

\[
\text{Sensitivity} = \frac{TP}{TP + FN} \times 100 \%
\]

\[
\text{Specificity} = \frac{TN}{FP + TN} \times 100 \%
\]

- Confusion matrix: A confusion matrix contains information about actual and predicted classifications performed by a classifier. Performance of such classifiers is commonly evaluated using the data in the matrix. Table 3 shows the confusion matrix for a two class classifier:

As can be observed, AMMLP is superior to classical Backpropagation MLP training in all cases.

The performance of the two classifiers in detection of breast cancer is presented in Table 5.

7.1.2. Performance results ROC

- Receiver Operating Characteristic (ROC) curve: The Receiver Operating Characteristic (ROC) curve is a two dimensional measure of classification performance, widely used in biomedical research to assess the performance of diagnostic tests (Witten & Frank, 1999; Zhou & Harezlak, 2002). An ROC curve is a plot of sensitivity vs. (1-specificity), or equivalently the true positive fraction vs. the false positive fraction, computed from the application of a series of thresholds to the system output. ROC graphs plot false positive (1-specificity) rates on the x-axis and true positive (sensitivity) rates on the y-axis. A simple, easy to implement approach to generating ROC curves is to collect the probabilities for all the various tests, along with the true class labels of the corresponding instances, and generate a single ranked list based on this data (Vega-Corona, Alvarez, & Andina, 2003; Witten & Frank, 1999). If the ROC curve rises rapidly towards the upper right-hand corner of the graph, or if the value of area under the curve is large, the test can be said to perform well. An area close to 1.0 indicates that the test is reliable, while an area close to 0.5 shows that the test is unreliable. In this case we used the ROC curve to show the AMMLP’s superiority over the ABP. The obtained ROC curve with our proposed model is presented in the Fig. 3.

- The Area Under the ROC Curve (AUC): Another method of evaluating classifier performance was used in calculating the ROC curve, the area under ROC curve (AUC). An AUC is a measure of test accuracy. The ROC curve describes two-dimensional visualization of ROC curve set of classifier performance. For the purpose of comparing two sets of classifiers, it is sometimes suitable to reduce ROC performance to a single scalar value representing expected performance (Hopley & Schalkwyk, 2001). The easiest method is to calculate the area under the ROC curve which is part of the area of the unit square. Consequently the value of AUC will always satisfy the following inequalities:

\[0 \leq \text{AUC} \leq 1\]

It is clear that an AUC close to 1 (area of unit square) indicates very reliable diagnostic test (Bradley, 1997). The AUC can be computed by integrating the area under the ROC curve (summing the areas of trapezoids) or by the Mann- Whitney-Wilcoxon test statistic (Purves, 1992; Bettinger, 2003). In our study, we obtained the AUC values by the trapezoidal rule. If the area is sliced into vertical segments, each segment is a trapezoid. The total AUC is calculated by adding these segment areas together. The results after calculating the AUC in this case were 0.989 for AMMLP and 0.928 for BP. This indicates once again the superiority of AMMLP over BP, in this particular case.

### Table 3
Confusion matrices.

<table>
<thead>
<tr>
<th>Representation of confusion matrix</th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>Negative</td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>

### Table 4
Confusion matrices of classifiers used for classification of breast cancer.

<table>
<thead>
<tr>
<th>Type classifiers</th>
<th>Desired result</th>
<th>Output results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Benign records</td>
<td>Malignant</td>
</tr>
<tr>
<td>AMMLPs</td>
<td>178</td>
<td>0</td>
</tr>
<tr>
<td>Malignant records</td>
<td>2</td>
<td>93</td>
</tr>
<tr>
<td>BPNNS</td>
<td>175</td>
<td>3</td>
</tr>
<tr>
<td>Malignant records</td>
<td>12</td>
<td>83</td>
</tr>
</tbody>
</table>

### Table 5
Classification accuracies of classifiers used for detection of breast cancer.

<table>
<thead>
<tr>
<th>Type classifier</th>
<th>Classification accuracies (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Specificity</td>
</tr>
<tr>
<td>AMMLPs</td>
<td>97.89</td>
</tr>
<tr>
<td>BPNNS</td>
<td>98.31</td>
</tr>
</tbody>
</table>
Fig. 3. ROC curves of the classifier and show indicates one more time the AMMLP superiority over the BP, in this particular case. (a) Show AMMLP ROC and AUC of 0.989 (b) Show BP ROC and AUC of 0.928.

Table 6
Classification accuracies obtained with our method and other classifiers from literature.

<table>
<thead>
<tr>
<th>Author (year)</th>
<th>Method</th>
<th>Classification accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quinlan (1996)</td>
<td>C4.5</td>
<td>94.74</td>
</tr>
<tr>
<td>Hamilton et al. (1996)</td>
<td>RAIC</td>
<td>95.00</td>
</tr>
<tr>
<td>Ster and Dobnikar (1996)</td>
<td>LDA</td>
<td>96.80</td>
</tr>
<tr>
<td>Nauck and Kruse (1999)</td>
<td>NEFCLASS</td>
<td>95.06</td>
</tr>
<tr>
<td>Pena-Reyes and Sipper (1999)</td>
<td>Fuzzy-GA1</td>
<td>97.36</td>
</tr>
<tr>
<td>Setiono (2000)</td>
<td>Neural-rule 2a</td>
<td>98.10</td>
</tr>
<tr>
<td>Albrecht et al. (2002)</td>
<td>LSA machine</td>
<td>98.80</td>
</tr>
<tr>
<td>Abonyi and Szeifert (2003)</td>
<td>SFC</td>
<td>95.57</td>
</tr>
<tr>
<td>Übeyli (2007)</td>
<td>SVM</td>
<td>99.54</td>
</tr>
<tr>
<td>Polat and Güney (2007)</td>
<td>LS-SVM</td>
<td>98.53</td>
</tr>
<tr>
<td>Gujjarro-Berdias et al. (2007)</td>
<td>LLS</td>
<td>96.00</td>
</tr>
<tr>
<td>Akay (2009)</td>
<td>SVM-CFS</td>
<td>99.51</td>
</tr>
<tr>
<td>Karabatak and Cevdet-Ince (2009)</td>
<td>AR + NN</td>
<td>97.40</td>
</tr>
<tr>
<td>Peng et al. (2009)</td>
<td>CPW</td>
<td>99.50</td>
</tr>
<tr>
<td>This study (2009)</td>
<td>AMMLP</td>
<td>99.26</td>
</tr>
</tbody>
</table>

8. Comparison and discussion

For comparison purposes, Table 6 gives the classification accuracies of our method and previous methods applied to the same database. As can be seen from the results, our AMMLP method obtains an excellent classification accuracy.

9. Conclusion

In this study, a Artificial Neural Network for Classification Breast Cancer based on the biological metaplasticity property was presented. The proposed AMMLP algorithm was compared with the classic MLP with Backpropagation, applied to the Wisconsin Breast Cancer Database. The AMMLP classifier shows a great performance obtaining the following results average for 100 networks: 97.89% in specificity, 100% in sensitivity and the total classification accuracy of 99.26%, the ROC curve to show the AMMPL superiority over the classic MLP with Backpropagation and finally the results obtained after calculating the AUC in this case were as follows for AMMLP is 0.989 while the AUC for BP is 0.928, this indicates one more time the AMMLP superiority over the BP, in this particular case. From the above results, we conclude that the AMMLP obtains very promising results in classifying the possible breast cancer. We believe that the proposed system can be very helpful to the physicians for their as a second opinion for their final decision. By using such an efficient tool, they can make very accurate decisions. Our AMMLP, proved to be equal or superior to the state-of-the-art algorithms applied to the Wisconsin Breast Cancer Database, and shows that it can be an interesting alternative.

Acknowledgements

This research has been partially supported by National (MICINN) and Madrid (CAM) Spanish institutions under the following projects: PTFNN (MICINN ref: AGL2006-12689/AGR). The authors wishes to thank to The National Foundation of Science Technology and Innovation (FONACIT) of the Bolivariana Republic of Venezuela and The National Council for Science and Technology (CONACyT), The Secretariat of Public Education (SEP) and Government of Mexico for its contribution in the development of theirs doctoral studies.

References
