Breast cancer classification applying artificial metaplasticity algorithm

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

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

A R T I C L E   I N F O

Keywords:
Neural network
Backpropagation algorithm
Pattern recognition
Metaplasticity
Breast cancer

A B S T R A C T

A novel improvement in neural network training for pattern classification is presented in this paper. The proposed training algorithm is inspired by the biological metaplasticity property of neurons and Shannon’s information theory. This algorithm is applicable to artificial neural networks (ANNs) in general, although here it is applied to a multilayer perceptron (MLP). During the training phase, the artificial metaplasticity multilayer perceptron (AMMLP) algorithm assigns higher values for updating the weights in the less frequent activations than in the more frequent ones. AMMLP achieves a more efficient training and improves MLP performance. The well-known and readily available Wisconsin Breast Cancer Database (WBCD) has been used to test the algorithm. Performance of the AMMLP was tested through classification accuracy, sensitivity and specificity analysis, and confusion matrix analysis. The results obtained by AMMLP are compared with the backpropagation algorithm (BPA) and other recent classification techniques applied to the same database. The best result obtained so far with the AMMLP algorithm is 99.63%.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

The concept of biological metaplasticity was defined in 1996 by Abraham [1] and now widely applied in the fields of biology, neuroscience, physiology, neurology and others [1,2]. The prefix “meta” comes from Greek and means “beyond” or “above”. In neuroscience and other fields “metaplasticity” indicates 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, that depends on prior synaptic activity [3]. Metaplasticity is due, at least in part, to variations in the level of postsynaptic depolarization that induce 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 alteration are functions of postsynaptic depolarization during synaptic activation. Long-term potentiation (LTP) is generated following low levels of postsynaptic depolarization, whereas long-term depression (LTD) is produced by stronger depolarizations (see Figs. 1 and 2).

The induction of synaptic changes in the levels of neural activity is explained in Fig. 1 [4]. Metaplasticity can be represented as variations in curve elongation with respect to the level of activity and implies a shift of the LTP threshold according to the weight strength of the synapse [5]. Figs. 1 and 2 graphically illustrate this idea. Understanding metaplasticity may yield new insights into how the modification of synapses is regulated and how information is stored by synapses in the brain [3].

The main objective of the proposed work is to model and test the biological property of metaplasticity on a multilayer perceptron (MLP) trained with BPA. This interpretation has been modeled in the neural network (NN) training phase. The well-known Wisconsin Breast Cancer Database (WBCD) [6] was used to test the proposed artificial metaplasticity with the MLP algorithm (AMMLP). The AMMLP algorithm was then compared with classical backpropagation and other algorithms, recently proposed by other researchers, that were successfully applied to the same database.

Section 2 presents a brief introduction to concepts related to the mechanism of metaplasticity. Section 3 introduces our hypothesis on the relationship between metaplasticity and Shannon’s information theory, which will lay the foundation for understanding the proposed model. In Section 4, general mathematical theory is applied to support the proposed implementation of artificial metaplasticity (AMP) in ANNs with error minimization-based learning. Section 5 describes the implementation of the AMP algorithm in the MLP neural network, trained with the BPA. Section 6 presents a real application of the AMMLP algorithm to the WBCD. Section 7 shows the experimental results. Section 8 compares our results to other methods from the literature and presents a brief discussion. Finally, Section 9 presents the summarized conclusions.

2. Synaptic plasticity

Synaptic plasticity refers to the modulation of the efficacy of information transmission between neurons and is related to the regulation of the number of ionic channels in the synapse.
2 defines which curve must be used is the value of the synaptic weight, receptors, the activation of intracellular signaling cascades and these modifications include the redistribution of postsynaptic function, either enhancing or depressing neuronal transmission.

Fig. 1. Changes in synaptic strength due to postsynaptic activity in biological neurons. If postsynaptic activity is high, the curve will move to the right, reinforcing the LTP. This graphic shows a family of curves in which each curve indicates the variation in weight, $\Delta w$, respective of the neuron's activation. The parameter that defines which curve must be used is the value of the synaptic weight, $w$. For higher values of the weight, the curve elongates further to the right.

Fig. 2. Metaplasticity consists of a shift in the LTP threshold, according to the initial weight of the synapse. The two figures above graphically depict this idea. For higher initial values of synaptic weight, the curve is elongated so that the LTP threshold value corresponds to higher values of postsynaptic activity. Note that postsynaptic activity is measured in volts, whereas synaptic activity is measured in activations/time, that is, frequency units.

The mechanisms responsible for synaptic plasticity involve both molecular and structural modifications that affect synaptic function, either enhancing or depressing neuronal transmission. These modifications include the redistribution of postsynaptic receptors, the activation of intracellular signaling cascades and the formation/retraction of dendrites [7]. The first model of synaptic plasticity was postulated by Hebb and is commonly known as the Hebb rule [8].

2.1. Metaplasticity vs intrinsic plasticity

Metaplasticity generally prevents null or saturated synaptic weights. However, these extreme situations cannot be completely prevented. Intrinsic plasticity regulates the position (rightward shift) of the neuron's activation function according to previous levels of activity [10]. Metaplasticity uses intrinsic plasticity to exclude nullification or saturation of synaptic weight. Metaplasticity contributes to neuronal homeostasis, maintaining individual neuron activity levels and, ensuring that neither nullification nor saturation occurs [9].

3. Metaplasticity and Shannon's information theory

As is well-known within the ANN field, in 1949 Hebb postulated that during the learning phase, synaptic connections between biological neurons are strengthened due to the correlated activity of presynaptic and postsynaptic neurons [8]. This plasticity property of synaptic connections is modeled in many ANNs as a change in the connection weights of the artificial neurons or nodes. Therefore, synaptic plasticity of biological neural networks has been simulated in artificial networks by changing the weight values of the simulated neuronal connections. These weights are the most relevant parameters in ANN learning and performance. Modeling these new discovered properties of biological neurons that follow metaplasticity rules provides a large potential for improving ANN learning. In addition, the results of these simulations may also corroborate the biological hypothesis of neuronal metaplasticity. Utilizing the potential of this new modeling approach, artificial metaplasticity (AMP) models have been devised and tested. A model that closely followed biological metaplasticity and intrinsic plasticity was successfully tested in [10], reinforcing the calcium dysregulation hypothesis for Alzheimer’s disease. However, of all AMP models tested by the authors, the most efficient model (as a function of learning time and performance) is the approach that connects metaplasticity and Shannon’s information theory, which establishes that less frequent patterns carry more information than frequent patterns [11]. This model defines artificial metaplasticity as a learning procedure that produces greater modifications in the synaptic weights with less frequent patterns than frequent patterns, as a way of extracting more information from the former than from the latter. As biological metaplasticity, AMP then favors synaptic strengthening for low-level synaptic activity, while the opposite occurs for high level activity. The model is applicable to general ANNs, as stated in [12], where Andina et al. propose general AMP concepts for ANNs, and demonstrate them over Radar detection data. In this paper it has been implemented and tested for a MLP over WBCD.

4. Backpropagation algorithm and AMP

The AMP implementation applied tries to improve results in learning convergence and performance by capturing information associated with significant rare events. It is based on the idea of modifying the ANN learning procedure such that un-frequent patterns which can contribute heavily to the performance are considered with greater relevance during learning without changing the convergence of the error minimization algorithm. It has been proposed on the hypothesis that biological metaplasticity property maybe significantly due to an adaptation of nature to extract more information.

Please cite this article as: A. Marcano-Cedeño, et al., Breast cancer classification applying artificial metaplasticity algorithm, Neurocomputing (2010), doi:10.1016/j.neucom.2010.07.019
from un-frequent patterns (low synaptic activity) that, according to Shannon’s Theorem, implicitly carry more information.

4.1. Mathematical definitions

Let us define an input vector for a MLP with \( n \) inputs (bias inputs are assumed to exist and be of fixed value set to 1): \( x \in \mathbb{R}^n \), where \( \mathbb{R}^n \) is the \( n \)-dimensional space, i.e., \( x = \{x_1, x_2, ..., x_n\}, x_i \in \mathbb{R}^i, i = 1, 2, ..., n \); and its corresponding \( y \) outputs given by vector \( y = \{y_1, y_2, ..., y_L\}, y_i \in \{0, 1\}, i = 1, 2, ..., L \). Let us consider now the random variable of input vectors \( X = \{X_1, X_2, ..., X_n\} \) with probability density function (pdf) \( f_\theta(x) = f_\theta(x_1, x_2, ..., x_n) \). The strategy of MLP learning is to minimize an expected error, \( E_M \), defined by the following expression:

\[
E_M = E[E(x)]
\]

where \( E(x) \) is the expression of an error function between the real and the desired network output, being respectively \( Y = F(X) \), with pdf \( f_\theta(y) \) and \( Y_d \), the desired output vector, and \( F(X) \) is the nonlinear function performed by the MLP. The symbol \( E \) represents the mathematical expectation value, that is,

\[
E_M = \int \bar{E}(x)f_\theta(x)\,dx
\]

4.2. AMP in gradient descent algorithm

BPA training algorithm applied in MLPs follows Widrow gradient descent algorithm over an estimation of this expected error in each training iteration, \( t \in \mathbb{N} \), for determining the necessary modification in the ANN weight matrix \( W(t) \) in each bias and weight value in the MLP [13]. The algorithm objective is to reduce the output classification error in subsequent training epochs, stopping the training phase if the error is low enough to satisfy the design requirements.

To introduce AMP in the gradient descent algorithm, Eq. (2) has been manipulated in the following way:

\[
e(x) = E(x)f_\theta(x), \quad E_M = \int \bar{E}(x)f_\theta(x)\,dx = \bar{E}E(x)f_\theta(x)
\]

where a new probability density function (pdf) \( f_\theta(x) \) has been introduced, requiring that \( f_\theta(x) \neq 0 \) wherever \( e(x) \neq 0 \), \( \forall x \in \mathbb{R}^n \) and new mathematical expectation, \( \bar{E} \), defined in Eq. (3) represents that the minimization of \( E_M \) can also be achieve from statistical inference theory applied to Eq. (3), by estimating over the weighted function \( e(x)f_\theta(x) \) instead of \( e(x) \), under \( f_\theta(x) \) pdf, through the following estimator:

\[
\hat{E}_M = \frac{1}{M} \sum_{k=1}^{M} \bar{E}e_{x_k}f_\theta(x_k)
\]

where \( x_k^t, k = 1, 2, ..., P \), are independent sample vectors whose pdf is the weighting function \( f_\theta(x) \). Note that many functions may fix to the definition of \( f_\theta(x) \), in particular:

\[
[f_\theta(x)]_{pdf} = \frac{1}{E_M}e(x)
\]

that can be proved by taking Eq. (5) into Eq. (4); only one simple sample vector (P = 1) is then required for exactly estimating \( E_M \) without error. The optimal solution for \( f_\theta(x) \) given by Eq. (5) is not realistic, because \( E_M \) is not known a priori (it has to be estimated by Eq. (4)). But, a suboptimal solution can be used. For example, the suboptimal solution for \( f_\theta(x) \) applied and tested in this paper is

\[
\tilde{f}_\theta(x) = \frac{A}{\sqrt{2\pi}^n}\exp\sum_{i=1}^{n}x_i^2 \approx \frac{1}{W_\theta(x)}
\]

where \( W_\theta(x) \) is defined as \( 1/f_\theta(x) \), \( N \) is the number of neurons in the MLP input layer, and parameters \( A \) and \( B \) are algorithm optimization values which depend on the specific application of the AMMLP algorithm. Values for \( A \) and \( B \) have been empirically determined.

4.3. AMP in MLP training: AMMLP

In the case of an MLP trained with BPA applied to \( L \) classes, \( H_l, l = 0, 1, ..., L - 1 \), previous studies have shown that the output for each class is the MLP inherent estimation of a posteriori probability of the class [14], based on Bayes Theorem, we then have

\[
y_l \geq P(H_l|x) = \frac{f_\theta(x|H_l) \cdot P(H_l)}{f_\theta(x)}
\]

This enables a direct implementation of metaplasticity. For each class, by assuming the proposed AMP model described in Section 4.2 \( f_\theta(x) \) can be inferred and from Eqs. (7) and (4) we obtain

\[
e(x|H_l) \approx \bar{E(x)}f_\theta(x|H_l)
\]

where \( k = 1, 2, ..., M \) are the independent sample vectors of class \( l \) in the training set. Then, from Eqs. (8) and (4)

\[
y_l \geq \frac{1}{P(H_l)}
\]

Eq. (7) takes advantage of the inherent a posteriori probability estimation for each input class of MLP outputs, so it is used to quantify a pattern’s frequency. Note that if this is not the case, as it happens in first steps of BPA training algorithm, the training may not converge. In this first steps, the outputs of the MLP do not provide yet any valid estimation of the a posteriori probabilities, but rather random values corresponding to initial guess of the MLP weights, \( W \). It is then better in these first steps of training, either to apply ordinary BPA training or to use another valid weighting function till BPA starts to minimize the error objective. Also, many suboptimal functions may yield good results. For example, in the following experiments, a typical approximation premise that assumes a Gaussian distribution for the inputs has been implemented, proposing the function for weight updating (known as a weighting function) [12], given by Eq. (6).

To analytically introduce AMP in an arbitrary MLP training, all that has to be done is to introduce the weighting function in the error function between the real and the desired network output, as a function of the weights matrix \( W(t) \) in each training iteration, \( t \), that is

\[
E^*(W(t)) = \frac{E(W(t))}{f_\theta(x)}
\]

And apply the BPA [13] to the weighted error \( E^*(W) \) for weights reinforcement in each iteration \( t \in \mathbb{N} \). If \( s, j, n \in \mathbb{N} \) are the MLP layer, node and input counters, respectively, for each \( W(t) \) component, \( w_{ij}^n(t) \in \mathbb{R} \), and being \( \eta \in \mathbb{R}^+ \) the learning rate, then the weight reinforcement in each iteration is given by

\[
w_{ij}^{n+1}(t+1) = w_{ij}^{n}(t) - \eta \frac{\partial E^*(W(t))}{\partial w_{ij}^{n}} = w_{ij}^{n}(t) - \frac{1}{f_\theta(x)} \frac{\partial E(W(t))}{\partial w_{ij}^{n}}
\]

So, as the pdf weighting function proposed is the distribution of the input patterns that does not depend on the network parameters,
the AMMLP algorithm can then be summarized as a weighting operation for updating each weight in each MLP learning iteration as
\[ \Delta w = w^*(x)\Delta w \]  
being \( \Delta w = w(t+1) - w(t) \) the weight updating value obtained by usual BPA and \( w^*(x) \) the realization of the described weighting function \( w^*(x) \) for each input training pattern \( x \).

5. Application of AMMLP algorithm to breast cancer classification

The proposed artificial metaplasticity algorithm (AMMLP) is here applied to a critical medical problem, breast cancer classification. Because of its relevance, this problem is the subject of intensive research. The well-known Wisconsin Breast Cancer Database (WBCD) has been used to test the AMMLP.

5.1. Overview of the Wisconsin Breast Cancer Database

Breast cancer is a malignant tumor that develops from breast cells. Although research has identified some of the risk factors (i.e., aging, genetic risk factors, family history, menstrual periods, not having children and obesity) that increase a woman’s chance of developing breast cancer, the inherent cause of most breast cancers remains unknown. Further, how many of these risk factors cause cells to become cancerous remains a mystery. However, researchers and scientists are continually improving our understanding of how certain changes in DNA can trigger normal breast cells to become cancerous [15].

The correct pattern classification of breast cancer is an important worldwide medical problem. Cancer is 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 and medical community. Breast cancer etiologies remain unclear and no single dominant cause has emerged. Prevention is still a mystery and the best way to improve patient survival is through early detection. If the cancerous cells are detected before they spread to other organs, the survival rate is greater than 97%. A major class of problems in medical science involves disease diagnosis based on various tests performed on patients. For this reason, the use of classifier systems in medical diagnosis is gradually increasing. There is no doubt that classification systems, through minimizing possible errors likely produced due to tiredness or lack of experience, can provide more detailed medical data that can be checked in a shorter period of time [16].

This study analyzed the Wisconsin Breast Cancer Database [6]. The WBCD data set consists of 699 samples 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. Integer values from 1 to 10 are assigned to the assessments, 1 being the closest to benign and 10 the most anaplastic. Each sample is also associated with its class label, which is either benign or malignant. This data set contains 16 entries with missing attribute values that were excluded from analysis in this study. The database contains 444 (65.0%) benign samples and 239 (35.0%) malignant samples.

6. Materials and methods

MLP exhibits a sigmoidal activation function with scalar output usually in the range (0,1). This property is true for all neurons of the network. To comparatively evaluate the performance of the various classifiers presented in this study, all methods 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 were benign. The testing set, composed of the remaining 40% of the data, consisted of 95 malignant samples and 178 benign samples.

Two different experiments were developed in this study. One experiment was developed to obtain the most accurate result and another was designed to obtain the average of 100 simulations. In the second example, 100 AMMLPs with different initial weights, sampled from random values of a normal distribution (mean of 0 and a variance of 1) has been generated. In each experiment, 100 networks were trained to achieve an average result that is independent of the initial random value of the ANN values. Two different criteria were applied to stop training: in one case, training was stopped when the error reached 0.01 (error decreases but cannot reach to 0). In the other one, training was performed with a fixed number of 2000 epochs.

The AMMLP algorithm was developed in MATLAB® (software MATLAB version 7.4, R2007a) and on a 3.4 GHz Pentium IV computer with 2 GB of RAM.

6.1. Network structure selection

Network parameters recently applied to related studies (see [12,17] and, specifically, [18] where preliminary experiments over WBCD are presented) were used to determine the network structure and the initial metaplasticity parameters (see also the order of A and B in Table 2). Two different criteria were applied to optimize network structure and metaplasticity parameters:

1. Metaplasticity parameters: The number of neurons in the hidden layer is fixed to a sufficiently high value to ensure that the ANN has sufficient processing units to perform the classification. The metaplasticity parameters (A, B) were varied empirically until the error converged \( \approx 0.01 \) in the minimum number of iterations.

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

* N = 683 observations, 239 malignant and 444 benign.

---

Please cite this article as: A. Marcano-Cedeño, et al., Breast cancer classification applying artificial metaplasticity algorithm, Neurocomputing (2010), doi:10.1016/j.neucom.2010.07.019
2. Number of neurons in hidden layers: The number of neurons in hidden layers was varied until the neural network achieved a mean square error (MSE) of approximately 0.01 (metaplasticity parameters remained unchanged) with the minimum number of neurons possible without degrading final performance.

Table 2 shows, as an example, the results obtained for different network structures and metaplasticity parameters in the first experiment presented in the next section. Table 3 shows the network structure, metaplasticity parameters, epochs, MSE and numbers of patterns used in the training and testing phases.

6.2. The AMMLP algorithm

1. Network structure used in the experiments:
   - Number of input neurons equal to the number of attributes of the database records plus the bias input.
   - Number of hidden layers: 1.
   - Hidden neurons: 8.
   - Output neurons: 1 (all classifications present two classes).
   - Learning rate: $\eta = 1$.
   - Activation function is sigmoidal with a value between (0,1).
2. Initialize all weights in weight matrix $W$ randomly and start AMMLP training for the weighting function defined in Eq. (6)
3. Training phase:
   - Test training conditions:
     - if epochs $= 2000$ stop training
     - if mean squared error, MSE $= 0.01$ stop training

7. Results

7.1. Performance evaluation methods

This section presents the AMMLP results and compares them with the results of a classical BPA MLP.

Table 2

<table>
<thead>
<tr>
<th>Network structure</th>
<th>Metaplasticity parameters</th>
<th>Mean squared error</th>
<th>Clustering accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>HL</td>
<td>O</td>
<td>A</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>1</td>
<td>39</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>1</td>
<td>41</td>
</tr>
<tr>
<td>9</td>
<td>7</td>
<td>1</td>
<td>39</td>
</tr>
</tbody>
</table>

Table 3

<table>
<thead>
<tr>
<th>Network parameters applying to the WBCD.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Types classifiers</td>
</tr>
<tr>
<td>-------------------</td>
</tr>
<tr>
<td>AMMLP</td>
</tr>
<tr>
<td>BPNNs</td>
</tr>
</tbody>
</table>

* NA: does not apply.

To measure the performance of the classifiers of breast cancer diagnosis, the evaluation has been divided into two parts: the first was to determine the algorithm accuracy, which is related to the classification accuracy, analysis of sensitivity and specificity, and the confusion matrix. The second part of the evaluation uses a receiver operating characteristic (ROC) to derive a curve of the results obtained from the first part of the evaluation. The area under the ROC curve (AUC) is calculated to measure the performance of the classifier. These methods are explained in the following sections.

7.1.1. Performance result accuracy

- Classification accuracy is 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 from a patient with breast cancer, as diagnosed by the clinic experts.

- True negative ($TN$): An input is normal and is labeled as a healthy individual by the expert clinicians.

- False positive ($FP$): An input is from a patient with breast cancer, but was labeled as a healthy person by the expert clinicians. This error is of critical cost.

- False negative ($FN$): An input is normal but diagnosed as breast cancer.

- Sensitivity and specificity: The following expressions for sensitivity and specificity analysis has been used:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \times 100$$

$$\text{Specificity} = \frac{TN}{FP + TN} \times 100$$

- Confusion matrix: A confusion matrix contains information about actual and predicted classifications performed by a classifier. Performance of the classifier is commonly evaluated using the data in the matrix.

Table 4 shows the classification results obtained in the best simulation for each classifiers used in this study in a confusion matrix.

Table 4

<table>
<thead>
<tr>
<th>Type classifiers</th>
<th>Desired result</th>
<th>Output results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Benign</td>
<td>Malignant</td>
</tr>
<tr>
<td>AMMLPs</td>
<td>Benign records</td>
<td>177</td>
</tr>
<tr>
<td></td>
<td>Malignant records</td>
<td>1</td>
</tr>
<tr>
<td>BPNNs</td>
<td>Benign records</td>
<td>175</td>
</tr>
<tr>
<td></td>
<td>Malignant records</td>
<td>3</td>
</tr>
</tbody>
</table>

Please cite this article as: A. Marcano-Cedeño, et al., Breast cancer classification applying artificial metaplasticity algorithm, Neurocomputing (2010), doi:10.1016/j.neucom.2010.07.019
As seen in Table 4, AMMLP is superior to the classical backpropagation MLP training in all cases.

The performance of the two classifiers in detection of breast cancer is presented in Table 5. This result was obtained from the best simulation. The average obtained for 100 simulations is shown in Table 6.

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 that is widely used in biomedical research to assess the performance of diagnostic tests [19]. A ROC curve is a plot of sensitivity vs. 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 specificity rates on the x-axis and true positive sensitivity rates on the y-axis. A simple, easy to implement approach for generating ROC curves involves collecting the probabilities for all the various tests, along with the true class labels of the corresponding instances, and generating a single ranked list based on the data [19]. If the ROC curve rises rapidly towards the upper right-hand corner of the graph, or if the area value of the curve is large, the test can be described as working well. An area close to one indicates that the test is reliable, while an area close to one half indicates that the test is unreliable. In this case, the ROC curve to demonstrate the superiority of AMMLP over BPA has been used. The resulting ROC curve of our proposed model is presented in Fig. 3.

- **The area under the ROC curve (AUC)**: The AUC value will always satisfy the following inequalities:

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

It is clear that an AUC close to one indicates a very reliable diagnostic test [19]. The AUC values obtained in this case were 0.989 for AMMLP and 0.928 for BPA. This finding indicates once again the superiority of AMMLP over BPA in this particular case.

8. Comparison and discussion

The results obtained with the AMMLP are here compared to the results of other algorithms in two ways: first, these results are compared to recently proposed algorithms applied to the WBCD database. Second, the results are also compared with other good algorithms that have been developed by other researchers using the same database.

- The AMMLP results were compared to recently proposed algorithms applied to the WBCD database by Conforti and Guido in [20]. Their report proposed an optimization model-based approach for learning the best kernel function to be embedded into the support vector machine (SVM) classifier. They generated an optimal kernel function by formulating and solving a semi-defined programming (SDP) model. They obtained accuracy results of 96.79%. However, their learning algorithm cannot be completed in a reasonable amount of time because the SDP/SVM model is computationally inefficient in the case of very large-scale

---

Table 5
Classification accuracies of classifiers used for detection of breast cancer obtained from the best simulation.

<table>
<thead>
<tr>
<th>Type classifier</th>
<th>Classification accuracies (%)</th>
<th>Specificity</th>
<th>Sensitivity</th>
<th>Total classification accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMMLPs</td>
<td>100</td>
<td>99.43</td>
<td>99.63</td>
<td></td>
</tr>
<tr>
<td>BPNNs</td>
<td>94.73</td>
<td>98.31</td>
<td>97.06</td>
<td></td>
</tr>
</tbody>
</table>

Table 6
Classification accuracies of classifiers used for detection of breast cancer obtained from 100 simulations.

<table>
<thead>
<tr>
<th>Type classifier</th>
<th>Classification accuracies (%)</th>
<th>Specificity</th>
<th>Sensitivity</th>
<th>Total classification accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMMLPs</td>
<td>98.94 ± 0.6</td>
<td>99.43 ± 0.3</td>
<td>99.58 ± 0.3</td>
<td></td>
</tr>
<tr>
<td>BPNNs</td>
<td>94.46 ± 0.9</td>
<td>98.57 ± 0.4</td>
<td>97.79 ± 0.04</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 3. ROC curves of the classifier. These curves indicate the superiority of AMMLP over BPA in this particular case. (a) AMMLP ROC with an AUC of 0.989; (b) BPA ROC with an AUC of 0.928.

Please cite this article as: A. Marcano-Cedeño et al., Breast cancer classification applying artificial metaplasticity algorithm, Neurocomputing (2010), doi:10.1016/j.neucom.2010.07.019
sets. Peng et al. [21] presented a feature selection approach to deal with high dimensionality issues in biomedical data classification. Their approach integrates filter and wrapper methods into a sequential search procedure to improve the feature classification performance. The proposed approach was implemented by (1) adding a pre-selection step to improve the effectiveness in searching the feature subsets with improved classification performances and (2) using receiver operating characteristics (ROC) curves to characterize the classification performance of individual and subset features. They obtained a value of 0.997 for area under the ROC curve. In [22] Akay presented an SVM-based model using a grid search to optimize model parameters and an F-score calculation to select input features. Akay reached a classification accuracy of 99.51%. Übeyli [23] using five classifiers SVM, probabilistic neural network, recurrent neural network, combined neural network and multilayer perceptron neural networks—reported an accuracy of 99.54%.

It is important to emphasize, that the authors of these studies do not indicate whether their results represent the best simulation product or are an average of several simulations. Proposed method not indicate whether their results represent the best simulation of the well-known Wisconsin Breast Cancer Database. It provides better results than BPA and also than other state-of-the-art algorithms applied to the same database. The results indicate that use of the AMMLP algorithm is preferred for classifying breast cancer samples, improving present performance of classifiers helping as a second opinion for physicians when making their final diagnostic decisions.

### Acknowledgements

This research was supported by the Madrid Autonomous Government (CAM) CCG08-UPM/000-1995 project. The authors wish to thank the National Foundation of Science Technology and Innovation (FONACIT) of the Bolivarian Republic of Venezuela and the National Council for Science and Technology (CONACYT), the Secretariat of Public Education (SEP) and the Government of Mexico for their contributions to the development of the authors’ doctoral studies.

### References

[1] W.C. Abraham, Activity-dependent regulation of synaptic plasticity (meta-

73055-2_65.


[7] G. Neves, S.F. Cooke, T.V. Bliss, Synaptic plasticity, memory and the hippo-


4, INSPEC Accession Number: 10411864.


[13] D. Andina, D.T. Pham, Computational Intelligence for Engineering and Man-

### Table 7

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

<table>
<thead>
<tr>
<th>Author (year)</th>
<th>Method</th>
<th>Classification accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quinlan [24]</td>
<td>C4.5</td>
<td>94.74</td>
</tr>
<tr>
<td>Hamiton et al. [25]</td>
<td>RAIC</td>
<td>95.00</td>
</tr>
<tr>
<td>Ster and Dobnikar [26]</td>
<td>LDA</td>
<td>60.80</td>
</tr>
<tr>
<td>Nauck and Kruse [27]</td>
<td>NEFCLASS</td>
<td>95.06</td>
</tr>
<tr>
<td>Pena-Reyes and Sipper [28]</td>
<td>FUZZY-GA1</td>
<td>97.26</td>
</tr>
<tr>
<td>Setiono [29]</td>
<td>NEURO-RULE 2a</td>
<td>98.10</td>
</tr>
<tr>
<td>Albrecht et al. [30]</td>
<td>LSA MACHINE</td>
<td>98.80</td>
</tr>
<tr>
<td>Abonyi and Szeifert [31]</td>
<td>SVM</td>
<td>95.57</td>
</tr>
<tr>
<td>Übeyli [23]</td>
<td>SVM</td>
<td>99.54</td>
</tr>
<tr>
<td>Polat and Günes [32]</td>
<td>LS-SVM</td>
<td>98.53</td>
</tr>
<tr>
<td>Guijarro et al. [33]</td>
<td>LLS</td>
<td>96.00</td>
</tr>
<tr>
<td>Akay [22]</td>
<td>SVM-CFS</td>
<td>99.51</td>
</tr>
<tr>
<td>Karabatak and Cevdet [34]</td>
<td>AR+NN</td>
<td>97.40</td>
</tr>
<tr>
<td>Peng et al. [21]</td>
<td>CFW</td>
<td>0.957*</td>
</tr>
<tr>
<td>Conforiti and Guido [20]</td>
<td>SVM-SDP</td>
<td>96.79</td>
</tr>
<tr>
<td>In this study (2010)</td>
<td>AMMLP</td>
<td>99.63*</td>
</tr>
<tr>
<td>In this study (2010)</td>
<td>AMMLP</td>
<td>99.58b</td>
</tr>
</tbody>
</table>

\* Result obtained in the AUC of ROC.
\* B Best result obtained form one simulation.
\* Averaged obtained for 100 simulations.

approach is based on the biological property of metaplasticity. Proposed AMMLP algorithm performance was compared in classification of the well-known Wisconsin Breast Cancer Database. It provides better results than BPA and also than other state-of-the-art algorithms applied to the same database. The results indicate that use of the AMMLP algorithm is preferred for classifying breast cancer samples, improving present performance of classifiers helping as a second opinion for physicians when making their final diagnostic decisions.

### 9. Conclusion

In this study, the artificial metaplasticity on MLP has been applied to the problem of breast cancer classification. The AMMLP

Please cite this article as: A. Marcano-Cedeño, et al., Breast cancer classification applying artificial metaplasticity algorithm, Neurocomputing (2010), doi:10.1016/j.neucom.2010.07.019
A. Marcano-Cedeño was born in Bolivar city, Venezuela. He received the degree of B.Sc. in Computer Science by Central University of Venezuela (UCV), in 1995. He received the Master degree information Systems in Andres Bello Catholic University in 1998. In 2000 he is an assistant professor in the Pedagogical University Experimental Libertador (UPEL). In 2005 he achieved Master degree in Communication Systems and Networks by the Technical University of Madrid (UPM). Currently he is studying his Ph.D. at the Technical University of Madrid (UPM). His research interests include pattern recognition, image processing and analysis, artificial neural networks and fuzzy set techniques.

J. Quintanilla Domínguez was born in 1979 in Salamanca Guanajuato, Mexico. He received the degree of B.Sc. in Electronics and Telecommunication Engineering and M.Sc. in Electrical Engineering from University of Guanajuato in 2004 and 2007, respectively. He is studying his Ph.D. at the Technical University of Madrid (UPM). His research interests include pattern recognition, image processing and analysis, artificial neural networks and fuzzy set techniques.

D. Andina (Prof. Dr.-Eng, IEEE Senior Member), was born in Madrid, Spain, where he received simultaneously two Master degrees, in Computer Science and on Electronics and Communications by the Technical University of Madrid (UPM), Spain, in 1990. He achieved the Ph.D. degree in 1995 with a thesis on Artificial Neural Networks applications in Signal Processing. He presently works for UPM where he heads the Group for Automation in Signals and Communications (GASC/UPM), a research group interested in Signal Processing and Computational Intelligence Applications; Man-Machine Systems and Cybernetics. He is author or co-author of more than 200 national and international publications, being director of more than 50 R+D projects financed by National and Local Governments, European Commission or private Institutions and Firms. He is also Associate Editorial Member of several International Research Journals and Transactions, and has participated in the organization of more than 50 international Research, Innovation or Technology Transfer events.