

# Artificial metaplasticity prediction model for cognitive rehabilitation outcome in acquired brain injury patients

Alexis Marcano-Cedeño<sup>1</sup>, Paloma Chausa<sup>2</sup>, Alejandro García<sup>3</sup>, César Cáceres<sup>4</sup>, Josep M. Tormos<sup>5</sup>, Enrique J. Gómez<sup>6</sup>

## ABSTRACT

**Objective:** The main purpose of this research is the novel use of artificial metaplasticity on multilayer perceptron (AMMLP) as a data mining tool for prediction the outcome of patients with acquired brain injury (ABI) after cognitive rehabilitation. The final goal aims at increasing knowledge in the field of rehabilitation theory based on cognitive affectation.

**Methods and materials:** The data set used in this study contains records belonging to 123 ABI patients with moderate to severe cognitive affectation (according to Glasgow Coma Scale) that underwent rehabilitation at Institut Guttmann Neurorehabilitation Hospital (IG) using the tele-rehabilitation platform PREVIRNEC<sup>®</sup>. The variables included in the analysis comprise the neuropsychological initial evaluation of the patient (cognitive affectation profile), the results of the rehabilitation tasks performed by the patient in PREVIRNEC<sup>®</sup> and the outcome of the patient after a 3–5 months treatment. To achieve the treatment outcome prediction, we apply and compare three different data mining techniques: the AMMLP model, a backpropagation neural network (BPNN) and a C4.5 decision tree.

**Results:** The prediction performance of the models was measured by ten-fold cross validation and several architectures were tested. The results obtained by the AMMLP model are clearly superior, with an average predictive performance of 91.56%. BPNN and C4.5 models have a prediction average accuracy of 80.18% and 89.91% respectively. The best single AMMLP model provided a specificity of 92.38%, a sensitivity of 91.76% and a prediction accuracy of 92.07%.

**Conclusions:** The proposed prediction model presented in this study allows to increase the knowledge about the contributing factors of an ABI patient recovery and to estimate treatment efficacy in individual patients. The ability to predict treatment outcomes may provide new insights toward improving effectiveness and creating personalized therapeutic interventions based on clinical evidence.

## 1. Introduction

An acquired brain injury (ABI) is a brain damage caused by events after birth, rather than part of a genetic or congenital disorder. It usually affects cognitive, physical, emotional, social or independent functioning and can result from either traumatic brain injury (TBI) (e.g. physical trauma due to accidents, falls, assaults, neurosurgery, etc.) or nontraumatic injury derived from either

an internal or external source (e.g. stroke, brain tumors, infection, poisoning, hypoxia, ischemia, encephalopathy or substance abuse).

At present, ABI is a major public health concern and a leading cause of disability worldwide. Severe disability resulting from an ABI has an estimated incidence of 2 per 100,000 population per year and moderate disability 4 per 100,000 population per year. The population most at risk suffering ABI are young people aged between 15 and 24 years with a male predominance, with traffic accidents as the main cause. In USA, it is estimated that 5.3 million Americans, more than 2 per of the country, are currently disabled due to acquired brain injury. About 1.5 million Americans suffer a acquired brain injury every year [1]. In Europe, that incidence is estimated between 175 and 200 per 100,000 population per year [2]. The World Health Organization (WHO) predicts that by the year 2020, ABI will be among the ten most common ailments. These

injuries dramatically change the life of patients and their families [3].

ABI patients have to go through a rehabilitation process that is usually focused on specific areas such as cognitive, motor, sensorial, etc. Rehabilitation of cognitive functions in patients with ABI has the objective of increasing the autonomy and quality of life of the patients (minimizing and/or compensating for functional alterations in this kind of patients) and their families [4]. The treatment consists of hierarchically organized treatment tasks and exercises which require repetitive use of the impaired cognitive system in a created, progressively more demanding sequence [5]. The rapid growth on ABI case numbers and the importance of cognitive functions in daily activities, both demand efficient programs of cognitive rehabilitation.

The recent introduction of automated systems for cognitive rehabilitation of patients with ABI generates large amounts of data [6,7]. The analysis of these data through information analysis and data mining techniques, allows us to obtain new knowledge to evaluate and improve the effectiveness of the rehabilitation process. Along these lines, the aim of this work is to apply data mining techniques to predict treatment outcome in patients with ABI. In particular, we use a learning algorithm based on the concept of biological metaplasticity as the training algorithm of a multilayer perceptron. The artificial metaplasticity on multilayer perceptron model is compared with two machine learning methods: a backpropagation neural network and a C4.5 decision tree.

The data used in this study has been obtained from the PREVIRNEC<sup>®</sup> platform. PREVIRNEC<sup>®</sup> is a cognitive tele-rehabilitation platform for neuropsychological clinical interventions, developed over an architecture based on web technologies. It's conceived as a tool to enhance cognitive rehabilitation, strengthening the relationship between the neuropsychologist and the patient, allowing personalization of treatment and monitoring the performance of rehabilitation tasks. PREVIRNEC<sup>®</sup> has been developed by the Universitat Rovira i Virgili and the Technical University of Madrid (Spain), together with the Institut Guttmann Neurorehabilitation Hospital neuropsychology and research departments [7].

The remainder of this paper is organized as follows: Section 2 refers a brief state of the art in data mining techniques applied to predict clinical outcomes of patients with ABI. Section 3 introduces artificial metaplasticity (AMP) and the applied mathematical theory to support its proposed implementation in artificial neural networks (ANNs) with error minimization-based learning. It also covers the description of the data set, the backpropagation neural network and the C4.5 decision tree used in this study. In Section 4 the mining process and the experimental results are presented. Section 5 shows a brief discussion and, finally, Section 6 presents the summarized conclusions.

## 2. Predictive data mining in ABI rehabilitation

Data mining is the extraction of implicit, previously unknown and potentially useful information from data [8]. The term data mining has become a synonym for "Knowledge Discovery in Databases" [9], that emphasizes the data analysis process rather than the use of specific analysis methods. Data mining is therefore a part of a knowledge discovery process that follows: a number of steps: the selection, preprocessing and transformation of data, the analysis phase and the interpretation of the patterns extracted. Two types of data mining models are distinguished [8,10]: predictive and descriptive. The predictive model makes prediction about unknown data values by using the known values. The descriptive model identifies the patterns or relationships in data and explores the properties of the examined data [10].

Data mining has been applied with success in different fields of knowledge and in the last few years, it has been increasingly used in medical literature [10,11]. One of the objectives of data mining in clinical medicine is to create models that can use specific patient information to predict the outcome of interest and to support clinical decision-making or form the basis of hypotheses for future experiments. Data mining methods may be applied to the construction of decision models for procedures such as prognosis, diagnosis and treatment planning [9,12].

Different statistical methodologies and data mining techniques have been applied to predict clinical outcomes of rehabilitation of patients with ABI. The best-known methods are decision trees (DT) and ANNs, both of which are widely used in the ABI literature [13,14].

Several studies compared ANN directly to another predictive models such as multiple regression, adaboost, logistic regression and support vector machine (SVM). Andrews et al., conducted a comparative study between decision trees and logistic regression to predict improvement in patients with ABI [13]. In Yi et al. [14], compared different learning techniques (decision tree, adaboost, support vector machine and artificial neural networks) to be used as a decision support tool based on rules for the treatment of patients with TBI. Brown et al., executed a study to consider all clinical elements available concerning a survivor of TBI admitted for inpatient rehabilitation, and identify those factors that predict disability, need for supervision, and productive activity one year after injury [15]. Rovlias and Kotsou, applied a classification tree to predict the evolution of a patient with severe head trauma [16]. With similar objectives and methodology, in [17], Segal et al., compared the accuracy of models based on neural networks with multiple regression and classification trees to predict the course of 1644 patients with TBI. Pang et al. in [18], proposed a hybrid model in combining different classification techniques (discriminant analysis, logistic regression, decision trees, Bayesian networks and neural networks) to study the correlation between predictors at the time of admission and outcome at 6 months in patients with ABI. In Rughani et al. [19], applied an ANN for predicting survival following traumatic brain injury and compared its predictive ability with regression models and clinicians.

The research work presented in this paper is focus on the cognitive rehabilitation of patients with acquired brain injury and included information about the affectation profile and the rehabilitation process that the patient has followed. Our major contribution is the novel use of AMMLP as a data mining tool for prediction the outcome of patients with ABI after cognitive rehabilitation. To evaluate the predictive validity of the AMMLP model, we carry out a comparative study between this model, a backpropagation neural network and a C4.5 decision tree.

## 3. Methods and materials

### 3.1. Artificial metaplasticity

The concept of biological metaplasticity was defined in 1996 by W.C. Abraham [20] and is now widely applied in the fields of biology, neuroscience, physiology, neurology and others [20,21]. 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* [22].

Recently, researchers as Marcano-Cedeño [23,24], Andina [25] and Ropero-Peláez [26], have introduced and modeled the biological property of the metaplasticity in the field of the ANNs obtaining excellent results.



Different models and simulations of AMP have resulted in conclusions relevant to not only the cybernetics field, but also to biology and medicine state-of-the-art [23,26,27]. 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 [28]. 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 [25], where Andina et al. propose general AMP concepts for ANNs, and demonstrate them over radar detection data and Marcano-Cedeño et al. applied the same algorithm for breast cancer classification [23]. In this paper it has been implemented as a data mining tool for predicting the improvement of patients with ABI after cognitive rehabilitation.

### 3.2. Artificial metaplasticity algorithm

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 from un-frequent patterns (low synaptic activity) that, according to Shannon's Theorem, implicitly carry more information.

#### 3.2.1. 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):  $x \in R^n$ , where  $R^n$  is the  $n$ -dimensional space, i.e.  $x = (x_1, x_2, \dots, x_n)$ ,  $x_i \in R^1$ ,  $i = 1, 2, \dots, n$ ; and its corresponding  $j$  outputs given by vector  $y = (y_1, y_2, \dots, y_m)$ ,  $y_j \in (0, 1)$ ,  $j = 1, 2, \dots, m$  [29]. Let us consider now the random variable of input vectors  $X = (X_1, X_2, \dots, X_n)$  with probability density function (pdf)  $f_X(x) = f_X(x_1, x_2, \dots, x_n)$ . The strategy of MLP learning is to minimize an expected error,  $E_M$ , defined by the following expression:

$$E_M = \varepsilon\{E(x)\} \quad (1)$$

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  $f_Y(y)$  and  $Y_d$  the desired output vector, and  $F(X)$  is the nonlinear function performed by the MLP. The symbol  $\varepsilon$  represents the mathematical expectation value, that is,

$$E_M = \int_{R^n} E(x) f_X(x) dx \quad (2)$$

#### 3.2.2. AMP in gradient descent algorithm

Backpropagation algorithm (BPA) for training a MLPs follows Widrow gradient descent algorithm over an estimation of this expected error in each training iteration,  $t \in N$ , for determining the necessary modification in the ANN weight matrix  $W(t)$  in each bias and weight value in the MLP [29]. 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:

$$e(x) \triangleq E(x) f_X(x); \quad E_M = \int_{R^n} \frac{e(x) f_X^*(x)}{f_X^*(x)}; \quad dx = \varepsilon^* \left\{ \frac{e(x)}{f_X^*(x)} \right\} \quad (3)$$

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

$$\hat{E}_M \triangleq \frac{1}{P} \sum_{k=1}^P \frac{e(x_k^*)}{f_X^*(x_k^*)} \quad (4)$$

where  $x_k^*$ ,  $k = 1, 2, \dots, P$ , are independent sample vectors whose pdf is the weighting function  $f_X^*(x)$ . Note that many functions may fit to the definition of  $f_X^*(x)$ , in particular:

$$[f_X^*(x)]_{opt} = \frac{1}{E_M} e(x) \quad (5)$$

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  $f_X^*(x)$  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  $f_X^*(x)$  applied and tested in this paper is:

$$f_X^*(x) = \frac{A}{\sqrt{(2\pi)^N} \cdot e^{B \sum_{i=1}^N x_i^2}} \triangleq \frac{1}{w_X^*(x)} \quad (6)$$

where  $w_X^*(x)$  is defined as  $1/f_X^*(x)$ ,  $N$  is the number of neurons in the MLP input layer, and parameters  $A$  and  $B \in R^+$  are algorithm optimization values which depend on the specific application of the AMLP algorithm. Values for  $A$  and  $B$  have been empirically determined. Eq. (6) is a gaussian distribution, so it has been assumed that  $X$  pdf is Gaussian (if it is not the case, the real  $X$  pdf should be used instead). Then,  $w_X^*(x)$  has high values for un-frequent  $x$  values and close to 1 for the frequent ones and can therefore be straightforwardly applied in weights updating procedure to model the biological metaplasticity during learning.

#### 3.2.3. AMP in MLP training: AMMLP

In the case of an MLP trained with BPA applied to  $L$  classes,  $H_l$ ,  $l = 0, 1, \dots, L-1$ , previous studies have shown that the output for each class is the MLP inherent estimation of a *posteriori* probability of the class [30], based on Bayes Theorem, we then have:

$$y_l \cong P\left(\frac{H_l}{x}\right) = \frac{f_X(x/H_l) \cdot P(H_l)}{f_X(x)} \quad (7)$$

This enables a direct implementation of metaplasticity. For each class, by assuming the proposed AMP model described in Section 3.2.2 can be made  $f_X^*(x) = f_X(x)$  and from Eq. (7) and (4)

$$e\left(\frac{x}{H_l}\right) \triangleq E(x) f_X\left(\frac{x}{H_l}\right) \quad (8)$$

$$\hat{E}_{M_l} = \frac{1}{M_l} \sum_{k=1}^{M_l} \frac{E(x_k) f_X(x_k/H_l)}{f_X(x_k)} \cong \frac{1}{M_l} \sum_{k=1}^{M_l} E(x_k) \frac{y_l}{P(H_l)}$$

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

$$\frac{y_l}{P(H_l)} \cong \frac{1}{f^*(x)} \quad (9)$$



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 does 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*) [25], 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

$$E^*[W(t)] = \frac{E[W(t)]}{f_X^*(x)} \quad (10)$$

And apply the BPA [29] to the weighted error  $E^*(W)$  for weights reinforcement in each iteration  $t \in \mathbb{N}$ . If  $s, j, i \in N$  are the MLP layer, node and input counters respectively, for each  $W(t)$  component,  $w_{ij}^{(s)}(t) \in \mathbb{R}$ , and being  $\eta \in \mathbb{R}^+$  the *learning rate*, then the weight reinforcement in each iteration is given by:

$$w_{ij}^{(s)}(t+1) = w_{ij}^{(s)}(t) - \eta \frac{\partial E^*[W(t)]}{\partial w_{ij}^{(s)}} = w_{ij}^{(s)}(t) - \eta \frac{1}{f_X^*} \frac{\partial E[W(t)]}{\partial w_{ij}^{(s)}} \quad (11)$$

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:

$$\Delta^* w = w^*(x) \Delta w \quad (12)$$

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  $x$ .

### 3.3. Algorithms review

#### 3.3.1. Backpropagation neural network

Backpropagation neural network (BPNN) is the most widely used search technique for training neural networks. Information in ANN is stored in the connection weights which can be thought of as the memory of the system. The purpose of BP training is to change iteratively the weights between the neurons in a direction that minimizes the error  $E$ , defined as the squared difference between the desired and the actual outcomes of the output nodes, summed over training patterns (training data set) and the output neurons.

The algorithm uses a sample-by-sample updating rule for adjusting connection weights in the network. In one algorithm iteration, a training sample is presented to the network. The signal is then fed in a forward manner through the network until the network output is obtained. The error between the actual and desired network outputs is calculated and used to adjust the connection weights. Basically, the adjustment procedure, derived from a gradient descent method, is used to reduce the error magnitude. The procedure is firstly applied to the connection weights in the output layer, followed by the connection weights in the hidden layer next to output layer. This adjustment is continued backward through to network until connection weights in the first hidden layer are

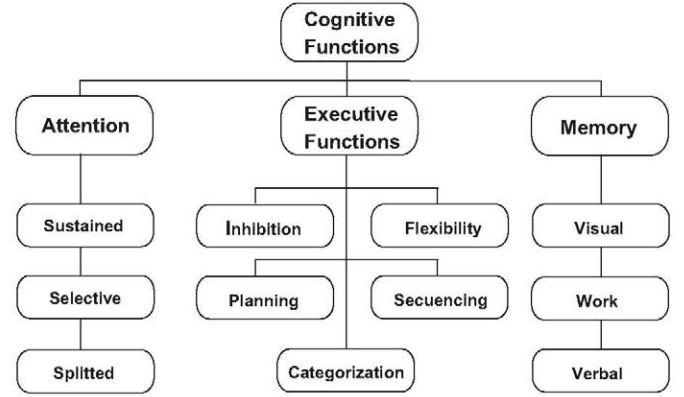


Fig. 1. Classification the cognitive functions and their respective subfunctions according to Institut Guttmann Neurorehabilitation Hospital.

reached. The iteration is completed after all connection weights in the network have been adjusted [23,31].

#### 3.3.2. C4.5 decision tree

A C4.5 decision tree is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility [9]. DT are commonly used in operations research, specifically in decision analysis, to help identify a strategy most likely to reach a goal. Another use of decision trees is as a descriptive means for calculating conditional probabilities. In data mining and machine learning, a decision tree is a predictive model; that is, a mapping from observations about an item to conclusions about its target value. More descriptive names for such tree models are classification tree (discrete outcome) or regression tree (continuous outcome). In these tree structures, leaves represent classifications and branches represent conjunctions of features that lead to those classifications. The machine learning technique for inducing a C4.5 from data is called decision tree learning, or (colloquially) decision tree. DT can classify variables according to a certain rule or classify data based on some data characteristics [32]. The goal of all decision tree algorithms is to maximize the "distance" between groups when splitting. Since evaluation measures of the "distance" are different, this allows to differentiate the decision tree algorithms. After the decision tree model is established, we can compute the error rate to pruning the decision tree. Pruning is a behavior that improves the prediction and classification ability of decision trees. In this research we have used a DT with three levels of pruning to compare the results obtained by AMMLP and BPNN.

### 3.4. Data description

The Institut Guttmann Neurorehabilitation Hospital (IG)[33] is specialized in the comprehensive rehabilitation and medical/surgical treatment of peoples suffering from spinal cord injury, acquired brain injury or other serious physical disabilities of neurological origin. One of the areas of expertise of IG is the rehabilitation of cognitive functions in patients with ABI. IG classifies the cognitive functions in the following domains: attention, executive function and memory (See Fig. 1).

As a first step of the IG cognitive rehabilitation process, the patients are administered a neuropsychological (NPS) assessment battery at admission for evaluating their cognitive functions. The NPS is the instrument that allows us to make a proper assessment of the patient's cognitive profile, while functional and psychosocial scales reflect the impact of deficits in daily life. The NPS assessment battery used by the IG includes 27 items covering the major cognitive domains (attention, memory and



**Table 1**

Cognitive functions, subfunctions, neuropsychological tests and items/attributes selected for this research.

Functions	Subfunctions	Neuropsychological test	Item	
Attention	Sustained	CPT	CPT_OMISSIONS CPT_COMMISSIONS CPT-TR TMT-A STROOP-WORD	
		WAIS-III	STROOP-COLOR STROOP-WORD-COLOR	
		STROOP	DIGIT SYMBOL CODING	
		WAIS-III	IMAGES	
		PIENC	STROOP_INTERFERENCE	
	Selective	STROOP	LETTER-NUMBER SEQUENCING	
		WAIS-III	TMT-B	
		TMT		
	Splitted	WAIS-III		
		TMT		
Memory	Work	WAIS-III	BACKWARD DIGIT SPAN	
	Verbal	WAIS-III	LETTER-NUMBER SEQUENCING	
		Immediately		
	Short period	WAIS-III	FORWARD DIGIT SPAN	
		RAVLT	RAVLT075	
	Visual			
		Long period	RAVLT	RAVLT015
	Recognition	RAVLT	RAVLT015R	
	Executuve functions	Planning	WAIS-III	BLOCK DESIGN
			WAIS-III	LETTER-NUMBER SEQUENCING
Inhibition		STROOP	STROOP_INTERFERENCE	
		WAIS-III	LETTER-NUMBER SEQUENCING	
Flexibility		WCST	WCST-PERSEVERATIVE ERRORS	
		PRM	PRM	
Sequencing		WAIS-III	BLOCK DESIGN	
		TMT	TMT-B	
Categorization		WCST	WCST_CATEGORIES	

executive functions) measured using standardized cognitive tests (see Table 1). This first evaluation allows us to know the initial cognitive profile of patient's affectation. After NPS initial evaluation all patients initiate a three to five months rehabilitation program adequated to their degree of affectation by means of PREVIRNEC<sup>®</sup> platform program. This computerized platform enables patients to perform different tasks during their rehabilitation sessions.

Once finished rehabilitation program, the patients are administered the NPS assessment battery again. Differences between pre and post treatment NPS test scores are used to measure patient's improvements in the domains of attention, memory and executive functions, and their respective subfunctions (see Table 1). These changes are assigned to one of the following labels: improvement, no improvement, normal, NSI (No significant improvement).

One hundred and twenty-three IG's ABI patients suffering moderate to severe cognitive affectation that underwent rehabilitation treatment with PREVIRNEC<sup>®</sup> platform from November 2007 to November 2009 were included in this analysis.

PREVIRNEC<sup>®</sup> is composed of a set of rehabilitation tasks for training different cognitive functions: attention, memory, executive functions and language. At the time of this analysis PREVIRNEC<sup>®</sup> included two hundred and fifteen rehabilitation tasks. Once a specific task is executed, the result (0–100 real number) is registered in the platform. Only the results of the task “memory” have been included to simplify the analysis. The task “memory”, designed to train the cognitive subfunction verbal-visual memory, has been selected because of its high number of executions (up to 3396). The analysis also includes the normalized values of some items selected from the neuropsychological assessment battery (the 4 related to cognitive subfunction verbal-visual memory, see Table 1), some demographic data (age and education level, see Table 2), and the outcome of the patient once the treatment is concluded. A more detailed description of the features included in the analysis is presented below:

- **Age group:** each participant age is categorized in 3 groups: Group 1 (from 17 to 30 years old), Group 2 (from 31 to 56 years old) and Group 3 (> 56). Prediction code “AGE”.
- **Educational level:** each participant education background level is categorized in 3 groups: Group 1 (Elementary School), Group 2 (High School) and Group 3 (University). Prediction code “EDU-LEVEL”.
- **Wechsler adult intelligence scale, (WAIS-III):** is used to measure adult and adolescent intelligence. In this research, the SPAN item of the WAIS-III has been included and its value has been normalized according to the International Classification of Functioning, Disability and Health (ICF) [33]: 0-normal, 1-mild affectation, 2-moderate affectation, 3-severe affectation, 4-acute affectation.
- **The rey auditory-verbal learning test, (RAVLT):** it is widely used in the neuropsychological assessment for evaluating verbal learning and memory. It's easy to understand, and appropriate for both children and adults (ages 7–89). In this research three items of this test have been included: MRAVL075, MRAVL015, MRAVL015TR and their values have been also normalized according to the International Classification of Functioning, Disability and Health (ICF) [34]: 0-normal, 1-mild affectation, 2-moderate affectation, 3-severe affectation, 4-acute affectation
- **Result:** is the result obtained by the patients in the execution of the rehabilitation tasks in PREVIRNEC<sup>®</sup> platform, with a range of

**Table 2**

Statistical distribution of demographic variables used in this research.

Variables	n (%)
Age	
Age 1 (from 17 to 30 years old)	53 (43.09)
Age 2 (from 31 to 55 years old)	42 (34.15)
Age 3 (> 56 years old)	28 (22.76)
Education	
Edu 1 (elementary school)	60 (48.78)
Edu 2 (high school)	40 (32.52)
Edu (university)	23 (18.70)



**Table 3**  
Summary of the attributes and classes used in this research.

Attribute number	Code attribute	Value of attribute
1	AGE	1–3
2	EDU-LEVEL	1–3
3	SPAN	1–4
4	RAVL075	1–4
5	RAVL015	1–4
6	RAVL15R	1–4
7	RESULT	0–100
8	IMPROVEMENT	Yes/No

[0,100]. In this study, these results belong to the task “memory”. Prediction code “RESULT”.

- **Improvement:** for each cognitive function, improvement values are obtained from neuropsychological assessments as the difference of pre and post values. Improvement takes (Yes/ No: no improvement and NSI), Prediction code “IMPROVEMENT”.

#### Data sets

The data set consists of 3396 samples taken from the task “memory” of PREVIRNEC<sup>®</sup> database. The database contains 1702 (50.12%) improvement samples and 1694 (49.88%) no improvement samples (which includes “no improvement” and “NSI”). Each record in the database have eight attributes and is associated with its class label, which is either improvement or not improvement (see Table 3).

### 3.5. Methods and algorithms implemented

#### 3.5.1. The artificial metaplasticity on multilayer perceptron algorithm in MATLAB

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 0.001 (error decreases but cannot reach to 0). Training was performed with different numbers of epochs 1000, 2000 and 3000 and learning rate (L.R.) 0.2, 0.5, 1.0 respectively.

The AMMLP algorithm was developed in MATLAB (MATLAB version 7.6.0.324, R2008a) and on a 3.4 GHz Pentium IV computer with 2 GB of RAM.

#### 3.5.2. Backpropagation neural network and J48 provided by WEKA

Waikato environment for knowledge analysis (WEKA), Version 3.4.3 was the data mining platform for the BPNN and DT algorithms executions. WEKA was developed at the University of Waikato in New Zealand [35], it is written in Java, is available on the Internet [35], and comprises a variety of data-mining algorithms. The J48 classifier algorithm is the WEKA implementation of the C4.5 decision tree [12].

In the Weka J48 classifier, lowering the confidence factor decreases the amount of post-pruning (the parameter altered to test the effectiveness of post-pruning in Weka is the confidence factor). For this research we used three different confidence factors (0.25, 0.3 and 0.5) and we obtained three different decision trees: DT1, DT2 and DT3. For the BPNN, we used the same network structure that was applied to the AMMLP.

### 3.6. Performance evaluation methods

To measure the performance of the models used in this study, the evaluation has been divided into two parts: the first was to determine models accuracy, which is related to the prediction

accuracy, analysis of sensitivity and specificity, and the confusion matrix (this measures are built from a confusion matrix which records correctly and incorrectly recognition such as the true positive (TP), false positive (FP), false negative (FN) and the true negative (TN) in binary classification) [23]. 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 models.

#### 3.6.1. Receiver operating characteristic 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 [23]. 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 [23]. 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.

#### 3.6.2. The area under the receiver operating characteristic curve

The area under the ROC curve (AUCROC) is used as a measure of diagnostic capability of the dataset. The AUC value will always satisfy the following inequalities:

$$0 \leq AUC \leq 1$$

It is clear that an AUC close to one indicates a very reliable diagnostic test [23]. The AUC values obtained in this case were of 0.926, 0.846 and 0.899 for AMMLP, BNPP and DT.

## 4. Results

This section presents the results obtained by the AMMLP, BPNN and DT models in the prediction of the outcome of ABI patients after cognitive rehabilitation. The results are analyzed and compared in order to determine the best predictive model.

To evaluate the performance of the prediction models, 10-fold cross-validation was selected, which is known to provide a good estimation of the generalization error of a classifier. With a 10-fold cross-validation procedure, the data set is split into 10 non-overlapping subsets of equal size. Each subset is divided in such a way that contains approximately the same number of examples from each class [9]. A classifier is trained 10 times, each time using a version of the data in which one of the subsets is omitted (testing data). Each trained classifier is then tested on the data from the subset which was not used during training. The results are averaged over the 10 classifiers to obtain an overall error estimate [9]. In this study the data set was split in 9 subsets with 340 records and 1 subset with 336.

#### Architecture model selection

In this study, different structures and parameters were tried in order to find the optimal choices for each method. The top three results of each model are presented in this subsection. In the case of AMMLP we used network parameters recently applied to other studies as the initial metaplasticity parameters A and B (see [25,36] and, specifically, [23]). Table 4 shows the three best architectures for the AMMLP model.

**Table 4**

Best network structures and metaplasticity parameters for AMMLP algorithm.

Model	Network structure			Metplasticity parameters		Mean squared error	L.R.	Epoch	Training time (s)	Prediction accuracy (%)	
	I <sup>1</sup>	HL <sup>2</sup>	O <sup>3</sup>	A	B					Training	Testing
AMMLP1	4	4	1	37	0.02	0.001	1	2000	77.51	97.58	92.07
AMMLP2	4	5	1	39	0.5	0.001	0.5	1000	88.79	95.85	91.36
AMMLP3	4	8	1	38	0.4	0.001	0.2	3000	75.22	94.45	90.72

Where I<sup>1</sup>: Input, HL<sup>2</sup>: Hidden Layer, O<sup>3</sup>: Output.**Table 5**

Network structure and parameters for BPNN model.

Model	Network structure			Mean squared error	L.R.	Epoch	Training time (s)	Prediction accuracy (%)	
	I <sup>1</sup>	HL <sup>2</sup>	O <sup>3</sup>					Training	Testing
BPNN1	4	4	1	0.001	1	2000	140.48	85.58	81.09
BPNN2	4	5	1	0.001	0.5	1000	14.56	84.85	80.89
BPNN3	4	8	1	0.001	0.2	3000	150.98	84.18	79.89

Where I<sup>1</sup>: Input, HL<sup>2</sup>: Hidden Layer, O<sup>3</sup>: Output**Table 6**

DT with different confidence factors.

Model	Confidence factor (%)	Minimum number objects (%)	Training time (s)	Prediction accuracy (%)	
				Training	Testing
DT1	0.3	2	90.35	91.98	89.10
DT2	0.5	2	91.25	91.69	89.25
DT3	0.25	2	90.45	92.45	90.33

In order to compare the performance of the AMMLP model with the BPNN one, we used the same network structure for both models (see Table 5).

To determine the best C4.5 architecture, we generate different models varying the confidence factor. Table 6 shows the three best configurations obtained.

Tables 7–9 summarized the prediction accuracy get by the three models evaluated. The input data for each learning algorithm is a collection of records which structure is described in Table 2. Each record corresponds to one execution of the rehabilitation task “memory” and is characterized by a tuple (x, y), where x is the attribute set and y is the class label. The attribute

set (x, y) is composed by 4 elements: AGE, EDU-LEVEL, RESULT and one of the four items related to the cognitive subfunction verbal/visual memory of the NPS assessment battery administered at the beginning of the rehabilitation process (ASPAN, RAVLT075, RAVLT015, MRAVL015TR). The class label is the patient improvement (YES/NO).

The resulting ROC curves and the AUC value for the three methods used in this research are presented in Fig. 2.

## 5. Discussion

The prediction performance of the models was measured by ten-fold cross-validation and several architectures were tested for each of them. As shown in Tables 7–9, the results obtained by the AMMLP are clearly superior to those obtained by the BPNN and C4.5 in terms of specificity, sensitivity and prediction accuracy. Prediction average accuracy of AMMLP, that measures the amount of correctly classified samples, rises to 91.56%. BPNN and C4.5 models have a prediction average accuracy of 80.18% and 89.91% respectively. The results obtained by AMMLP in the ROC curve and in the AUC ROC were also superior to those obtained by BPNN and DT. The excellent performance of the AMMLP model allows us to predict the improvement of a patient with ABI from the cognitive affectation profile and

**Table 7**

AMMLP prediction accuracy for the three best model architectures.

Model	ASPAN (%)	RAVLT075 (%)	RAVLT015 (%)	MRAVL015TR (%)	Model average (%)
AMMLP1	92.07 ± 0.7	91.78 ± 0.9	90.54 ± 0.6	91.88 ± 0.5	91.56 ± 0.6
AMMLP2	91.18 ± 0.9	90.50 ± 0.5	89.78 ± 0.8	90.15 ± 0.8	90.42 ± 0.7
AMMLP3	90.08 ± 0.8	90.47 ± 0.7	89.83 ± 0.9	90.26 ± 0.8	90.71 ± 0.8

**Table 8**

BPNN prediction accuracy for the three best model architectures.

Model	ASPAN (%)	RAVLT075 (%)	RAVLT015 (%)	MRAVL015TR (%)	Model average (%)
BPNNP1	81.09 ± 1.1	80.25 ± 0.7	79.35 ± 0.7	80.02 ± 0.8	80.18 ± 0.6
BPNNP2	80.40 ± 1.3	78.50 ± 0.9	67.70 ± 0.5	76.30 ± 0.6	75.70 ± 0.6
BPNNP3	80.86 ± 1.4	77.75 ± 0.8	75.26 ± 0.8	78.40 ± 0.8	78.07 ± 1.0

**Table 9**

Prediction accuracy for C4.5 with different confidence factors.

Model	ASPAN (%)	RAVLT075 (%)	RAVLT015 (%)	MRAVL015TR (%)	Model average (%)
DT1	88.24 ± 0.8	88.67 ± 0.8	80.47 ± 0.8	88.48 ± 0.8	87.97 ± 0.8
DT2	80.24 ± 0.5	89.75 ± 0.4	89.32 ± 0.6	88.60 ± 0.5	89.23 ± 0.5
DT3	90.33 ± 0.9	90.80 ± 0.8	89.47 ± 0.6	89.05 ± 0.6	89.91 ± 0.7



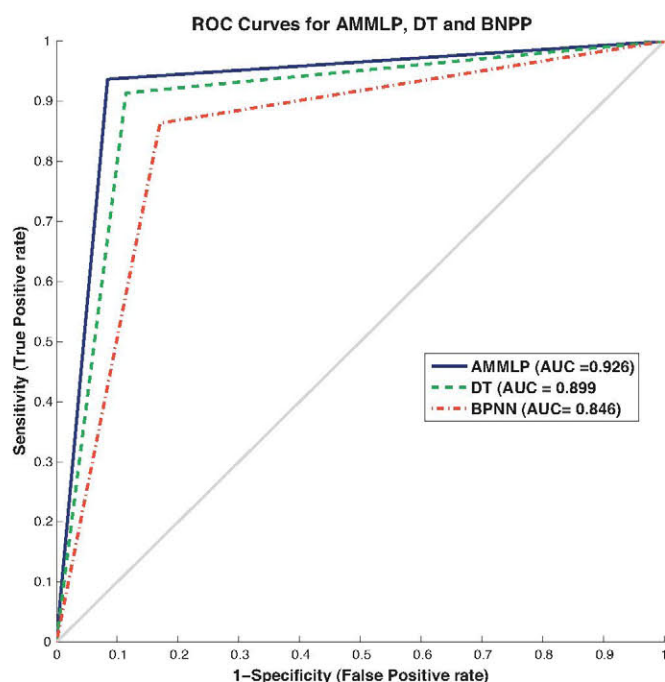


Fig. 2. ROC curves for the methods used in this research.

Table 10

Specificity, sensitivity and prediction accuracy obtained by AMMLP, BNPP and C4.5 models for the tuple (AGE, EDU-LEVEL, RESULT, ASPAN, IMPROVEMENT).

Model	Prediction (%)		
	Specificity	Sensitivity	Accuracy
AMMLPs	92.38 ± 0.9	91.76 ± 0.6	92.07 ± 0.7
BNPPs	88.43 ± 1.4	73.64 ± 1.1	81.09 ± 1.2
DT	90.75 ± 0.8	89.89 ± 0.9	90.33 ± 0.9

the information of the rehabilitation tasks performed. The best single AMMLP model provided a specificity of 92.38%, a sensitivity of 91.76% and a prediction accuracy of 92.07%, and corresponds to the input data containing the item ASPAN of the neuropsychological assessment battery. Considering the results obtained by the three prediction models, the variable ASPAN is the best predictor of outcome of all the neuropsychological battery items included in this study (see Table 10).

## 6. Conclusion and future works

ABI constitutes a major and increasing social and healthcare concern of a great diagnostic and therapeutic complexity. Its high recurrence and survival rate after the initial critical phases, makes it a prevalent problem that needs to be addressed. Cognitive rehabilitation improves disorders related to memory, attention, language, etc. and increases the patient's autonomy. In order to optimize the rehabilitation process, treatments must be intensive, personalized to the patient's condition and evidence-based; and require constant monitoring.

Currently, there is a lack of knowledge regarding patients' affection profiles and the combination of therapeutic tasks that optimize the treatment's effectiveness. This research work tries to deal with this situation by analyzing the outcome of ABI patients as a function of the cognitive affection profile, obtained from the neuropsychological initial evaluation of the patient, and the rehabilitation program he has followed.

In this work we have presented a novel use of the AMMLP model as a data mining tool for prediction the outcome of ABI patients

after cognitive rehabilitation. The AMMLP model have been compared with: a backpropagation neural network (BNPP) and a C4.5 decision tree. The Institut Guttmann Neurorehabilitation Hospital (IG) database has been used to test the models. This database contains data collected from PREVIRNEC<sup>®</sup> platform, a cognitive tele-rehabilitation platform integrated to the clinical practice of IG and other 14 rehabilitation centers. In light of the results obtained, the AMMLP prediction model outperforms the BNPP and the C4.5 models with an average prediction accuracy that rises to 91.56%.

The results obtained and the knowledge process defined, allow us to better understand the cognitive rehabilitation process in patients with ABI, to generate hypothesis about the rehabilitation program that better fits the patient's affection and thus, to aid clinicians in the design of more effective treatment methods.

We are currently extending the analysis to include new data mining techniques and patients' data. As a next step, we also plan to introduce information about other cognitive functions, as the items of the neuropsychological assessment battery and the results of the rehabilitation tasks related to attention and executive functions. With these extensions the method will provide clinicians and researchers with a deeper analysis on ABI patients' outcome.

## Acknowledgments

The authors would like to thank all the Institut Guttmann Neurorehabilitation Hospital staff and patients for all support and cooperation in this research. This research work was partially funded by Ministerio de Ciencia e Innovación (IPT-300000-2010-30 and P0890525).

## References

- [1] Traumatic Brain Injury. IOP Publishing Centers for Disease Control and Prevention Web. <http://www.cdc.gov/TraumaticBrainInjury/> (accessed 26.06.12).
- [2] The Lancet Neurology. Traumatic brain injury: time to end the silence. *The Lancet Neurology* 2010;9:331.
- [3] Pérez R, Costa U, Torrent M, Solana J, Ojiso E, Cáceres C, et al. Upper limb portable motion analysis system based on inertial technology for neurorehabilitation purposes. *Sensors* 2010;10:10733–51.
- [4] Fundaci Institut Guttmann. *Tecnologías Aplicadas al Proceso Neurorrehabilitador: estrategias para valorar su eficacia*. Badalona: Fundacio Institut Guttmann; 2008.
- [5] Sohlberg MM, Mateer CA. *Cognitive rehabilitation. An integrative neuropsychological approach*. New York: The Guilford Press; 2001.
- [6] Tormos JM, García-Molina A, García-Rudolph A, Roig T. Information and communication technology in learning development and rehabilitation. *International Journal of Integrated Care* 2009;9(5). ISSN: 1568-4156.
- [7] Solana J, Cáceres C, Gómez EJ, Ferrer-Celma S, Ferrer-Bergada M, García-Lopez P, et al. PREVIRNEC a new platform for cognitive telerehabilitation. In: *Proc. of the third international conference on advanced cognitive technologies and applications, COGNITIVE 2011*. 2011. p. 59–62. ISBN: 978-1-61208-155-7.
- [8] Meise S, Mattfeld D. Synergies of operations research and data mining. *European Journal of Operational Research* 2010;206:1–10.
- [9] Frawley WJ, Piatetsky-Shapiro G, Matheus CJ. *Knowledge discovery in databases: an overview*. Knowledge discovery in databases. CA: AAAI/MIT Press; 1991.
- [10] Bellazzi R, Zupan B. Predictive data mining in clinical medicine: current issues and guidelines. *International Journal of Medical Informatics* 2008;77(2): 81–97.
- [11] Bae E, Bailey J, Dong G. A clustering comparison measure using density profiles and its application to the discovery of alternate clusterings. *Data Mining and Knowledge Discovery* 2010;21(3):427–71.
- [12] Witten IH, Frank E. *Data mining practical machine learning tools and techniques*. San Francisco, USA; 2005.
- [13] Andrews PJ, Sleeman DH, Statham PF, McQuatt A, Corbule V, Jones PA, et al. Predicting recovery in patients suffering from traumatic brain injury by using admission variables and physiological data: a comparison between decision tree analysis and logistic regression. *Journal of Neurosurgery* 2002;97:326–36.
- [14] Ji SY, Smith R, Huynh T, Najarian K. A comparative analysis of multi-level computer-assisted decision making systems for traumatic injuries. *Medical Informatics and Decision Making* 2009;9(2):1–17.
- [15] Brown AW, Molec JF, McClelland RL, Diehl NN, Englander J, Cifu DX. Clinical elements that predict outcome after traumatic brain injury: a prospective



- multicenter recursive partitioning (decision-tree) analysis. *Journal of Neurotrauma* 2005;22(10):1040–51.
- [16] Rovlias A, Kotsou S. Classification and regression tree for prediction of outcome after severe head injury using simple clinical and laboratory variables. *Journal of Neurotrauma* 2004;21(7):886–93.
- [17] Segal ME, Goodman PH, Goldstein R, Hauck W, Whyte J, Graham JW, et al. The accuracy of artificial neural networks in predicting long-term outcome after traumatic brain injury. *Journal of Head Trauma Rehabilitation* 2006;21(4):298–314.
- [18] Pang BC, Kuralmani V, Joshi R, Hongli Y, Lee KK, Ang BT, et al. Hybrid outcome prediction model for severe traumatic brain injury. *Journal of Neurotrauma* 2007;24(1):136–46.
- [19] Rughani AI, Dumont TM, Lu Z, Josh Bongar MS, Horgan MA, Penar PL, et al. Use of an artificial neural network to predict head injury outcome. *Journal of Neurosurgery* 2010;113(3):585–90.
- [20] Abraham WC. In: Kato N, editor. *The hippocampus: functions and clinical relevance*. Amsterdam: Elsevier Science; 1996.
- [21] Kinto E, Del-Moral-Hernandez E, Marcano-Cedeno A, Ropero-Peláez J. A preliminary neural model for movement direction recognition based on biologically plausible plasticity rules. In: *In Proc. of international work-conference on the interplay between natural and artificial computation (IWINAC 2007)* 4528, 2007, p. 628–36.
- [22] Abraham WC. Metaplasticity: tuning synapses and networks for plasticity. *Nature Reviews Neuroscience* 2008;9:387–99.
- [23] Marcano-Cedeño A, Quintanilla-Domínguez J, Andina D. WBCD breast cancer database classification applying artificial metaplasticity neural network. *Expert Systems with Applications* 2011;38(8):9573–9.
- [24] Marcano-Cedeño A, Marín-de-la-Barcena A, Jiménez-Trillo J, Piñuela JA, Andina D. Artificial metaplasticity neural network applied to credit scoring. *International Journal of Neural Systems* 2011;21(4):311–7.
- [25] Andina D, Álvarez-Vellisco A, Jevtić A, Fombellida J. Artificial metaplasticity can improve artificial neural network learning. *Intelligent Automation and Soft Computing* 2009;15(4):683–96.
- [26] Ropero-Peláez J, Piqueira JR. Biological clues for up-to-date artificial neurons. In: Andina D, Pham DT, editors. *Computational intelligence for engineering and manufacturing*. The Netherlands: Springer-Verlag; 2007.
- [27] Monteiro J, Netto MI, Andina D, Peláez JR. Using neural networks to simulate the Alzheimer's Disease. In: *Proc. of world automation congress (WAC)*. 2008, p. 1–6.
- [28] Shannon CE. A mathematical theory of communication. *The Bell System Technical Journal* 1948;27:379–423.
- [29] Andina D, Pham DT, editors. *Computational intelligence for engineering and manufacturing*. The Netherlands: Springer-Verlag; 2007.
- [30] Ruck DW, Rogers SK, Kabrisky M, Oxley ME, Suter BW. The multi-layer perceptron as an approximation to a Bayes optimal discrimination function. *IEEE Transactions on Neural Networks* 1990;1(4):296–8.
- [31] Hasan H, Hasan B. Comparative evaluation of genetic algorithm and back-propagation for training neural networks. *Expert Systems with Applications* 2011;38(4):3703–9.
- [32] Quinlan JR. Induction of decision trees. *Machine Learning* 1986;1:81–106.
- [33] Institut Guttmann Neurorehabilitation Hospital. <http://www.guttmann.com> (accessed 31.09.12).
- [34] International Classification of Functioning, Disability and Health (ICF). <http://www.icf-research-branch.org/icf-core-sets-projects/neurologicalconditions/development-of-icf-core-sets-for-traumatic-brain-injurytbi.html> (accessed 01.11.12).
- [35] Weka. <http://www.cs.waikato.ac.nz/ml/weka/> (accessed 15.05.12).
- [36] Marcano-Cedeño A, Quintanilla-Domínguez J, Andina D. Wood defects classification using artificial metaplasticity neural network. In: *Proc. 35th annual conference of IEEE industrial electronics (IECON'09)*. 2009, p. 3422–7.