Data mining applied to the cognitive rehabilitation of patients with acquired brain injury

A. Marcano-Cedeño , Paloma Chausa , Alejandro García , César Cáceres , Josep M. Tormos , Enrique J. Gómez

A B S T R A C T

Acquired brain injury (ABI) is one of the leading causes of death and disability in the world and is associated with high health care costs as a result of the acute treatment and long term rehabilitation involved. Different algorithms and methods have been proposed to predict the effectiveness of rehabilitation programs. In general, research has focused on predicting the overall improvement of patients with ABI. The purpose of this study is the novel application of data mining (DM) techniques to predict the outcomes of cognitive rehabilitation in patients with ABI. We generate three predictive models that allow us to obtain new knowledge to evaluate and improve the effectiveness of the cognitive rehabilitation process. Decision tree (DT), multilayer perceptron (MLP) and general regression neural network (GRNN) have been used to construct the prediction models. 10-fold cross validation was carried out in order to test the algorithms, using the Institut Guttmann Neurorehabilitation Hospital (IG) patients database. Performance of the models was tested through specificity, sensitivity and accuracy analysis and confusion matrix analysis. The experimental results obtained by DT are clearly superior with a prediction average accuracy of 90.38%, while MLP and GRRN obtained a 78.7% and 75.96%, respectively. This study allows to increase the knowledge about the contributing factors of an ABI patient recovery and to estimate treatment efficacy in individual patients.

1. Introduction

Acquired brain injury (ABI) is one of the leading causes of death and disability in the world. In Europe, brain injuries from traumatic and non-traumatic causes are responsible for more years of disability than any other cause (The Lancet Neurology, 2010). Because most of these patients are young people, their remaining functional limitations and psychosocial problems contribute significantly to health care related costs and loss of productivity.

After sustaining an ABI, patients have impairments consisting of not only physical, but also cognitive, social, and behavioral limitations. The most frequently occurring cognitive sequelae after an ABI pertain to mental process slowness, attention deficits, memory impairments, and executive problems. The injury dramatically changes the life of patients and their families (Pérez et al., 2010). The rapid growth on ABI case numbers and the importance of cognitive functions in daily activities, both demand efficient programs of cognitive rehabilitation.

Recovery from ABI can be facilitated with cognitive rehabilitation. Cognitive rehabilitation aims to compensate, or restore when possible, lost brain functions, improving the quality of life of the patients (Fundaci Institut Guttmann, 2008; Sohlberg & Mateer, 2001).

One of the problems of the rehabilitation process is its time length, that in many cases is inadequate for a complete and effective rehabilitation. To improve and expand the cognitive rehabilitation process, automated systems for cognitive rehabilitation of patients with ABI have been recently introduced (Solana, Cáceres, Gómez, Ferrer-Celma, & Ferre-Bergada, 2011; Tormos, García-Molina, García-Rudolph, & Roig, 2009). These systems generate large amounts of data. The analysis of these data, using data mining techniques, allows us to obtain new knowledge to evaluate and improve the effectiveness of the rehabilitation process. Also using information analysis and data mining techniques, we can create predictive models and decision support systems for the treatment of patients with ABI.

The data used in this study were obtained from the PREVIRNEC© platform. PREVIRNEC© is a cognitive tele-rehabilitation platform, developed over a web-based architecture based on web technologies and it's conceived as a tool to enhance cognitive rehabilitation, strengthening the relationship between the neuropsychologist and the patient, extending the treatment duration and frequency, allowing personalization of treatment and monitoring the performance of rehabilitation tasks. PPREVIRNEC© has been developed

during the past six years by the Universitat Rovira i Virgili and Technical University of Madrid (Spain), together with the Institut Guttmann Neurorehabilitation Hospital, IG (Spain) neuropsychology and research departments (Solana et al., 2011). This platform has been included in the hospital clinical protocols since 2005 and at the moment of this analysis PREVIRNEC© database stores 1120 patients, with a total of 183047 rehabilitation tasks executions.

Different statistical methodologies and predictive data mining methods have been applied to predict clinical outcomes of rehabilitation of patients with ABI (Rughani et al., 2010; Ji, Smith, Huynh, & Najarian, 2009; Pang et al., 2007; Segal et al., 2006; Brown et al., 2005; Rovlias & Kotsou, 2004; Andrews et al., 2002). Most of these studies are focused in determining survival, predicting disability or the recovery of patients, and looking for the factors that are better at predicting the patient's condition after suffering an ABI.

The purpose of this study is the novel application of data mining to predict the outcomes of the cognitive rehabilitation of patients with ABI. Three algorithms were used in this study: decision tree (DT), multilayer perceptron (MLP) and a general regression neural network (GRNN). PREVIRNEC© database and IG's Electronic Health Records (EHR) (Institut Guttmann Neurorehabilitation Hospital, 1997) has been used to test the algorithms. For assessing the algorithm's accuracy of prediction, we used the most common performance measures: specificity, sensitivity, accuracy and confusion matrix. The results obtained were validated using the 10-fold cross-validation method.

The remainder of this paper is organized as follows. Section 2 presents a brief introduction to data mining, the algorithms used in this research and a detailed description of the database. Section 3 shows the experimental results obtained. Section 4 presents a discussion of these results. Finally, Section 5 describes the summarized conclusions.

2. Materials and methods

2.1. Review of data mining techniques

2.1.1. Knowledge discovery in databases and Data mining

Today there is still some confusion about the terms *Knowledge Discovery in Databases* (KDD) and *Data Mining (DM)*. Often these two terms are used interchangeably. The term KDD is used to denote the overall process of turning low-level data into high-level knowledge. KDD is defined as: *the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data* (Fayyad et al., 1996). On the other hand, data mining is commonly defined as the extraction of patterns or models from observed data. Therefore DM is at the core of the knowledge discovery process, although this step usually takes only a small part (estimated at 15–25%) of the overall KDD effort (Brachman & Anand, 1996). The overall KDD process contemplates five phases: data preparation, data preprocessing, data mining, evaluation and interpretation, and implementation (Fayyad et al., 1996).

Data mining classification technique is split in two phases: the first one is the construction of a classification model that consists in training the algorithm using a data set in order to build the predictive classification model. The second phase is the evaluation of the model classification efficiency, using a testing data set (Yen, Chen, & Chen, 2011; Köksa, Batmaz, & Caner, 2011; Seng, & Chen, 2010).

2.1.2. Decision trees (J48)

Decision tree (DT) provides powerful techniques for classification and prediction, that are widely used in data mining. The most commonly used DT algorithms include Quinlans ID3, C4.5, C5 (Quinlan, 1993), Breimans classification and regression tree (CART) (Breiman, Friedman, Olshen, & Stone, 1984) and Chi-squared Automatic Interaction Detector (CHAID) (Hartigan, 1975). As the name implies, this technique recursively separates observations in branches to construct a tree for the purpose of improving the prediction accuracy. In doing so, it uses mathematical algorithms to identify a variable and corresponding threshold for the variable that splits the input observation into two or more subgroups (Yeh, Hou, & Chang, 2012). The most commonly mathematical algorithm used for splitting includes Entropy based information gain (used in ID3, C4.5, C5), Gini index (used in CART), and chi-squared test (used in CHAID). This step is repeated at each leaf node until the complete tree is constructed. The objective of the splitting algorithm is to find a variable-threshold pair that maximizes the homogeneity (order) of the resulting two or more subgroups of samples (Delen, Fuller, McCann, & Ray, 2009).

Based on the favorable prediction results we have obtained from the preliminary runs, in this study we use the J48 algorithm as our decision tree method. The J48 algorithm is an implementation of the C4.5 algorithm (Witten & Frank, 2005) included in the WEKA software platform (weka, 2011). In order to tune the J48 algorithm to optimize its performance, we varied the confidence factor (default value of confidence factor is 0.5), which is a value that is used by the algorithm to prune developed trees (pruning of a DT is conducted to avoid over-fitting the model on the records used for modelling). A lower confidence factor results in more pruning (Witten & Frank, 2005), and a minimum number of objects for a leaf of 2.

2.1.3. Multilayer perceptron (MLP)

Multilayer perceptron (MLP) are the most commonly used feedforward neural networks due to their fast operation, ease of implementation, and smaller training set requirements (Haykin, 1994). The MLP consists of three sequential layers: input, hidden and output layers. The hidden layer processes and transmits the input information to the output layer. A MLP model with insufficient or excessive number of neurons in the hidden layer may cause problems of bad generalization and over-fitting. There is no analytical method for determining the number of neurons in the hidden layer, so it is usually chose empirically (Haykin, 1994; Marcano-Cedeño, Quintanilla-Domínguez, & Andina, 2011).

Each neuron j in the hidden layer sums its input signals x_i impinging onto it after multiplying them by their respective connection weights w_{ji}. The output of each neuron is described as follows:

$$y_i = f\left(\sum w_{ji} x_i\right) \tag{1}$$

where f is an activation function using the weighted summations of the inputs. An activation function can be a simple threshold, sigmoidal, or hyperbolic tangent function (Marcano-Cedeño et al., 2011; Güler, Gökçil, & Gülbandilar, 2009; Orhan, Hekim, & Ozer 2011). In this study, a sigmoidal transfer function was used as the activation function.

The sum of squared differences between the desired and actual values of the output neurons E is defined as follows (Marcano-Cedeño et al., 2011; Güler et al., 2009):

$$E = \frac{1}{2} \sum (y_{dj} - y_j)^2 \tag{2}$$

where y_{dj} is the desired value of output neuron j and y_j is the actual output of that neuron. Each w_{ji} weight is adjusted to minimize the value E depending on the training algorithm adopted. In this context, the backpropagation method (BP) is widely used as a primary part of an artificial neural network model. However, since the BP has some constraints such as slow convergence (Haykin, 1994; Marcano-Cedeño et al., 2011; Güler et al., 2009) or not being

able to find the global minimum of the error function, a number of variations for the BP were proposed. Therefore, in this study we used the BP supported by the Levenberg–Marquardt algorithm (Haykin, 1994). We also used different parameters of hidden layer, learning rate and momentum with the purpose of obtaining the best network structure for the MLP.

2.1.4. General regression neural network (GRNN)

The General regression neural network (GRNN) consists of four layers: input, pattern, summation and output layers. The number of input units in the first layer is equal to the total number of parameters. This first layer is fully connected to the second, pattern layer, where each unit represents a training pattern and its output is a measure of the distance of the input from the stored patterns. Each pattern layer unit is connected to the two neurons in the summation layer: S-summation neuron and D-summation neuron. The S-summation neuron computes the sum of the weighted outputs of the pattern layer while the D-summation neuron calculates the unweighted outputs of the pattern neurons. The connection weight between the ith neuron in the pattern layer and the S-summation neuron is y_i the target output value corresponding to the ith input pattern (Specht, 1991). For D-summation neuron, the connection weight is unity. The output layer merely divides the output of each S-summation neuron by that of each D-summation neuron, yielding the predicted value to an unknown input vector X as

$$y_i(X) = \frac{\sum_{i=1}^{n} y_i EXP[-D(X, X_i)]}{\sum_{i=1}^{n} EXP[-D(X, X_i)]} \tag{3}$$

where n indicates the number of training patterns and the Gaussian D function in Eq. (3) is defined as

$$D(X, X_i) = \sum_{j=1}^{p} \left(\frac{X_j - X_i j}{\zeta} \right)^2 \tag{4}$$

where p indicates the number of elements of an input vector. The x_j and x_{ij} represent the jth element of x and x_i, respectively. The ζ is generally referred to as the spread factor, whose optimal value is often determined experimentally (Specht, 1991). A large spread corresponds to a smooth approximation function. Too large a spread means a lot of neurons will be required to fit a fast changing function. Too small a spread means many neurons will be required to fit a smooth function, and the network may not generalize well. In this study, different spreads were tested to find the best value for the given problem. The GRNN does not require an iterative training procedure as in the BP method (Burcu & Tülay, 2008). GRNN is successfully used in pattern recognition due to the advantages on fast learning and convergence to the optimal regression surface as the number of samples becomes very large (Burcu & Tülay, 2008).

2.2. Data description

The Institute Guttmann is a hospital specialized in the comprehensive rehabilitation and medical/surgical treatment of persons suffering from spinal cord injury, ABI or other serious physical disabilities of neurological origin. One of the areas of expertise of IG is cognitive rehabilitation of patients with ABI. There is no standard classification of cognitive functions. The IG classifies cognitive functions as follows: attention, executive function, memory, and language. Disturbances of attention, memory and executive functioning are the most common neurocognitive consequences of TBI at all levels of severity (Upadhyay, 2008), therefore in this study we only work with them (See Fig. 1).

As a first step of the cognitive rehabilitation process, the specialists of IG perform a neuropsychological (NPS) assessment of the patient, for evaluating their cognitive functions and determining the degree of affectation in each cognitive function. The NPS is

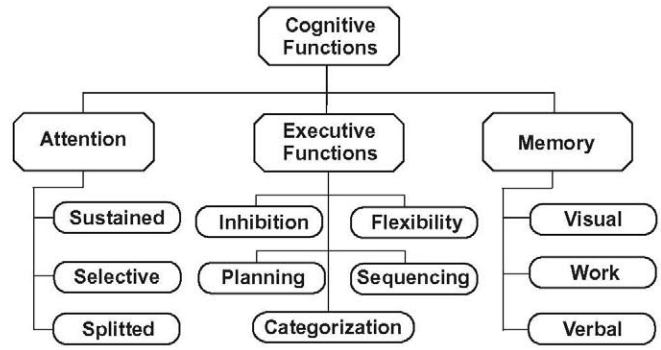

Fig. 1. Classification of the cognitive functions and their respective subfunctions according to Institute Guttmann Neurorehabilitation Hospital.

the instrument that allows the specialists 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 in the IG includes 27 items covering the major cognitive domains (attention, memory and executive functions) measured using standardized cognitive tests (see Table 1). This first evaluation allowed us to know the initial cognitive profile of the patient's affectation. All NPS assessment battery items are normalized to a 0–4 scale, where 0 = no affectation, 1 = mild affectation, 2 = moderate affectation, 3 = severe, and 4 = acute affectation. This NPS initial evaluation of the patient will allow the IG professional to plan a three to five months personalized rehabilitation program using the PREVIRNEC© platform. Once treatment is completed, all patients are administered the same NPS assessment tests. Differences between pre and post NPS scores were used to measure particular patient's outcome in the domains of attention, memory, executive functions and in their respective subfunctions. These changes are assigned to one of the following labels: improvement, no improvement, normal, NSI (No significant improvement).

Table 1
Cognitive functions, subfunctions, neuropsychological tests and items designed and classified by Institute Guttmann Neurorehabilitation Hospital (IG).

Funtions	Subfuntions		Neuropsychological test	Item
Attention	Sustained		CPT	OMISSIONS
			COMISSIONS	
			CPTTR	
			WAIS-III	TMT-A
	Selective		Test de Stroop	A_WORD
				A_COLOR
			A_WORD-COLOR	
			WAIS-III	VP WAISIII
	Splitted		Test de Stroop	FESTROOP
			TMT	FETMTB
Memory	Work		WAIS-III	MDIGITS
			WAIS-III	MLLETRES
		Immediately	WAIS-III	ASPAN
	Verbal	Short period	RAVLT	RAVLT075
	Visual	Long period	RAVLT	RAVLT015
		Recognition	RAVLT	RAVLT015R
Executive functions	Planning		WAIS-III	VCCUBS
			PMR	FEPMR
	Inhibition		Test de Stroop	FESTROOP
			WAIS-III	MLLETRES
	Flexibility		WCST	FEWCSTE
			PRM	FEPMR
	Sequencing		WAIS-III	VCCUBS
			TMT	FETMTB
	Categorization		WCST	FEWCSTC
			PRM	FEPMR

In this research we included two hundred and fifty patients with moderate to severe cognitive affectation (according to glasgow coma scale) that underwent rehabilitation treatment with the PREVIRNEC© platform of IG. At the moment of this analysis PREVIRNEC© included the results of these patients performing 183047 rehabilitation tasks. Each task is associated with one subfunction (see Table 1). This work seeks to demonstrate the predictive ability of the three methods by means of randomly selected tasks. In this study we focus in the analysis of the memory function, therefore, we only work with the tests that evaluate this function, and specifically we used the task called "memory". For this purpose, of the 27 NPS assessment tests, we selected the six tests that are used to assess the cognitive function memory (MDIGITS, MLLETRES, SPAN, MRAVL075, MRAVL015, and MRAVL015TR). The total number of tasks executions during the rehabilitation treatment was 10191. The patients' average age is 36,56 6.5 years old, range 18–68 years old. The sample consists of 185 men and 65 women. We also considered some demographic data such as age and educational level because of their relevance in ABI patients' rehabilitation. A brief description of the data used is presented below:

Demographic data

- *Age Group*: each participant is categorized in 3 age groups: Group 1 (from 17 to 30 years old), Group 2 (from 31 to 55 years old) and Group 3 (from >56 years old). This attribute is called with the named of prediction code "AGE".
- *Educational Level*: each participant is categorized in 3 education background level groups: Group 1 (Elementary School), Group 2 (High School) and Group 3 (University). This attribute is called with the named code "EDU-LEVEL".

Neuropsychological data

- *Wechsler Adult Intelligence Scale, (WAIS-III)*: it is used in the neuropsychological assessment to measure adult and adolescent intelligence. In this research, three items of the WAIS-III (MDIGITS, MLLETRES, SPAN) were included. The results of this test were normalized in the range of [0, 4] described above.
- *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 we have included three items of the RALVT test: MRAVL075, MRAVL015, MRAVL015TR. The results of each test were normalized in the range of [0, 4] described above.

Tasks executions data

- *Result*: is the result obtained by the participant in the execution of the tasks of cognitive rehabilitation, with a range of [0, 100]. In particular, these results belong to one of the tasks called "memory". This attribute is called with the named code "RESULT".
- *Improvement*: for each cognitive function, improvement values are obtained from neuropsychological assessments as the difference of pre and post values. The four possible values: improvement, no improvement normal, NSI (No significant improvement). This is transformed into a boolean variable (Yes/No). In this study the patients labeled as "normal" were omitted. In order to balance the data set, patients labeled "NSI" were added to patients labeled as "No improvement". This attribute is called with the named code "IMPROVEMENT".

2.2.1. Data sets

The data set consists of 10191 samples from the task "memory" of the IG database. The database contains 5476 (53.73%) improvement samples and 4715 (46.26%) no improvement samples. Each sample is also associated with its class label,which is either "improvement" or "no improvement". Table 2 shows the attributes and classes used in this research.

Table 2
Summary of the attributes and classes used in this research.

Attribute number	Code attribute	Values of attribute
1	AGE	1–3
2	EDU-LEVEL	1–3
3	MDIGITS	0–4
4	MLLETRES	0–4
5	SPAN	0–4
6	MRAVL075	0–4
7	MRAVL015	0–4
8	MRAVL015TR	0–4
9	RESULT	0–100
	Classes	
10	IMPROVEMENT	Yes/No

Table 3
DT with different confidence factors.

Model	Confidence factor	Minimum number objects	Accuracy (%)
DT1	0.25	2	91.18
DT2	0.4	2	90.63
DT3	0.6	2	90.37

Table 4
Results obtained by multilayer perceptron with different network structures and parameters.

Model	I^2	HL^2	O^2	LR^2	Momentum	Epoch	Accuracy (%)
MLP1	9	5	1	0.1	0.1	1000	79.57
MLP2	9	10	1	0.9	0.2	3000	78.42
MLP3	9	10	1	0.5	0.2	2000	79.57

Table 5
Results obtained by GRNN using different spread.

Model	Spread	Accuracy (%)
GRNN1	1	76.55
GRNN2	0.5	76.98
GRNN3	0.06	77.19

2.2.2. Measures of quality

The measures of quality of models are built from a confusion matrix which records correctly and incorrectly recognition of the true positives (TP), false positives (FP), false negatives (FN) and the true negatives (TN) in binary classification. In order to extend the usage of confusion matrix, we define the TP, FP, FN and TN in this paper as follows:

- True positive (TP): the TP denotes the number of patients that improve and are classified as patients that improve.
- False positive (FP): the FP denotes the number of patients that do not improve and are classified as patients that improve.
- False negative (FN): the FN denotes the number of patients that improve and are classified as patients that do not improve.
- True negative (TN): the TN denotes the number of patients that don't improve and are classified as patients that do not improve.

The sensitivity, specificity and accuracy prediction of classification are defined by:

Table 6
Results for 10-fold cross-validation for all folds and all model. Bold values highlight the best results obtained in this research.

Fold No.	Multilayer perceptron (MLP)					Decision tree (DT)					General regression neural network (GRNN)				
	Confusion	Matrix	Accuracy (%)	Sensitivity (%)	Specificity (%)	Confusion	Matrix	Accuracy (%)	Sensitivity (%)	Specificity (%)	Confusion	Matrix	Accuracy (%)	Sensitivity (%)	Specificity (%)
1	7168	2003	79.26	78.16	89.21	8357	814	91.18	91.12	91.67	6775	2396	73.55	73.87	70.59
	110	910				85	935				300	720			
2	6980	2192	77.31	76.10	88.22	8295	877	90.43	90.45	90.29	6812	2360	74.20	74.27	73.60
	120	899				98	921				269	750			
3	7160	2012	79.23	78.06	89.79	8281	891	90.27	90.28	90.09	6920	2123	75.95	75.45	80.47
	104	915				101	918				199	820			
4	7003	2169	77.51	78.21	87.93	8215	957	89.40	89.57	88.81	7008	2164	76.88	76.40	81.16
	123	896				114	905				192	827			
5	7195	1977	79.43	78.45	88.71	8265	907	90.27	90.11	91.27	6911	2261	75.76	75.35	79.49
	115	904				89	930				209	810			
6	7180	1992	79.57	78.28	91.16	8310	892	90.81	90.60	92.73	6975	2197	76.46	76.05	80.18
	90	929				74	945				202	817			
7	7025	2147	77.59	76.59	86.56	8350	822	91.11	91.03	91.76	7020	2152	76.98	76.53	80.96
	137	882				84	935				194	825			
8	7175	1997	79.30	78.23	89.01	8255	917	90.05	90.00	90.48	6899	2273	75.56	75.22	82.13
	112	907				97	922				218	801			
9	7155	2017	78.98	78.01	87.73	8261	911	89.99	90.06	89.30	7055	2117	77.19	76.92	79.69
	125	894				109	910				207	812			
10	7170	2002	79.48	78.17	91.26	8279	893	90.30	90.26	90.58	7022	2150	77.07	76.55	81.64
	89	930				96	923				187	832			
Average			78.77	77.83	88.26			**90.38**	**90.35**	**90.62**			75.96	75.66	78.99
St. Dev.[a]			0.9288	0.7980	1.6795			**0.4612**	**0.4738**	**1.1853**			1.2467	1.0169	3.7923

Confusion matrix shows the classification or the cases in the test dataset in contusion matrix, the columns denote the actual cases and the rows denote the predicted. Accuracy = (TP + TN)/(TP + FP + TN + FN); sensitivity = TP/(TP + FN); specificity = TN/(TN + FP).

[a] St. Dev.: standard deviation.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} (\%) \quad (5)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}} (\%) \quad (6)$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} (\%) \quad (7)$$

2.3. Model evaluation

For test results to be more valuable, a k-fold cross-validation is used among the researchers. It minimizes the bias associated with the random sampling of the training (Fayyad et al., 1996). In this method, whole data are randomly divided into k mutually exclusive and approximately equal size subsets. The classification algorithm is trained and tested k times. In each case, one of the folds is taken as test data and the remaining folds are added to form training data. Thus k different test results exist for each training-test configuration (Polat, Şahan, & Günes, 2006). The average of these results provides the test accuracy of the algorithm (Fayyad et al., 1996). A 10-fold cross-validation is used in all of our experiments. Therefore, the 10191 samples are separated into 10 subsets randomly (9 subsets with 1019 records and 1 subset with 1020) and then each subset is taken as test data in turn.

3. Results

In this section we present the obtained results in this research. All the models used in this study were trained and tested with the same data and validated using 10-fold cross-validation. First of all, we selected the best network structure for each model as explained in the following section.

3.1. Network selection

To select the best configuration for each model used in this study, we tested different network structures and parameters. We present only the top three results of each model.

Table 3 shows the three best architectures for the DT model. DT1 was selected because its accuracy was the best, choosing a confidence factor of 0.25.

Table 4 shows the three best architectures for the MLP model. In this case we selected MLP1, because it has the lowest computational cost.

To determine the best GRNN architecture, we generate different models varying the Spreads factor. Table 5 shows the three best configurations obtained. GRNN3 was selected because its accuracy was the best, choosing a spreads factor of 0.06.

3.2. Performance evaluation

In this study, the models were evaluated based on the accuracy measures discussed above (classification accuracy, sensitivity and specificity). The results were achieved using 10-fold cross-validation for each model, and are based on the average results obtained from the test data set (the 10th fold) for each fold. The results obtained are showed in Table 6.

As seen in Table 6, the results obtained by DT model are superior to the ones obtained by MLP and GRNN. Also, if we compare the standard deviation of the models, the decision tree seems to be the most stable of the studied prediction models.

4. Discussion

The reported results (Table 6), validated by means of the 10-fold cross-validation method, show that the DT model produces a higher level of prediction accuracy than the MLP and GRNN models do. Prediction average of the DT model was 90.38% accurate with sensitivity and specificity rates of 90.35% and 90.62%, respectively. MLP model obtained a prediction average accuracy of 78.77% with a sensitivity rate of 77.83% and a specificity rate of 88.26%. GRNN model achieved 75.93%, 75.66% and 78.99% of prediction average accuracy, sensitivity and specificity rates respectively. Similarly, if we compare the standard deviation, it can be seen that DT is again superior in terms of stability to the other prediction models.

We need to make a note that the results obtained by the GRNN model could improve substantially with a higher database, because it is well known that this algorithm require high volume of training patterns for obtaining a good performance. A derived consequence of enlarging the database would be the decrease of the backpropagation algorithm convergence rate used to train the MLP model. Another technique to improve results might be the use of feature selection algorithms to obtain a more accurate cognitive affectation profile.

5. Conclusion

In this paper three prediction models, decision tree (DT), multilayer perceptron (MLP) and general regression neural network (GRNN) are developed and compared. These models were applied to a real clinical head injury data set provided by the Institute Guttmann Neurorehabilitation Hospital and several architectures were tested in order to obtain the best structure and performance for each of them. The reported results, especially by DT model (90.38% prediction average accuracy), indicate that it is feasible to estimate the outcome of ABI patients as a function of the cognitive affectation profile, obtained from the neuropsychological initial evaluation of the patient, and the rehabilitation process data collected by the PREVIRNEC© platform. Next steps would be to include more data derived from the clinical and demographic history of the patient, and also from the rehabilitation tasks performed as part of his cognitive rehabilitation treatment. Other future research would investigate different data mining approaches in order to achieve higher prediction rates.

The findings from the present study led to an increase of knowledge in the field of rehabilitation theory. This will serve to the specialists to generate and validate clinical hypotheses as a previous step to the creation of personalized therapeutic interventions based on clinical evidence.

Acknowledgements

The authors thank all the staff and patients from the Institut Guttmann Neurorehabilitation Hospital for all their 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

Andrews, P. J., Sleeman, D. H., Statham, P. F., McQuatt, A., Corruble, V., Jones, P. A., et al. (2002). 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, 97*, 326–336. http://dx.doi.org/10.3171/jns.2002.97.2.0326.

Brachman, R., & Anand, T. (1996). The process of knowledge discovery in databases: A human-centered approach. In U. Fayyad, G. Piatetsky-Shapiro, Smyth P. Amith, & R. Uthurusamy (Eds.), *Advances in Knowledge Discovery and Data Mining*. Cambridge: MIT Press.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees* (1st ed.). California: Wadsworth International Group.

Brown, A. W., Malec, J. F., McClelland, R. L., Diehl, N. N., Englander, J., & Cifu, D. X. (2005). Clinical elements that predict outcome after traumatic brain injury: A prospective multicenter recursive partitioning (decision-tree) analysis. *Journal*

of Neurotrauma, 22(10), 1040–1051. http://dx.doi.org/10.1089/neu.2005.22.1040.

Burcu, E., & Tülay, Y. (2008). Improving classification performance of sonar targets by applying general regression neural network with PCA. *Expert Systems with Applications, 5*, 472–475. http://dx.doi.org/10.1016/j.eswa.2007.07.021.

Delen, D., Fuller, C., McCann, C., & Ray, D. (2009). Analysis of healthcare coverage: A data mining approach. *Expert Systems with Applications, 36*(2), 995–1003. http://dx.doi.org/10.1016/j.eswa.2007.10.041.

Frawley, W. J., Paitetsky-Shapiro, G., & Matheus, C. J. (1996). *Knowledge discovery in databases: An overview. Knowledge Discovery in Databases*. California: AAAI/MIT Press.

Frawley, W. J., Paitetsky-Shapiro, G., & Matheus, C. J. (1996). From data mining to knowledge discovery: An overview. In U. M. Fayyad et al. (Eds.), *Advances in knowledge discovery and data mining* (pp. 611–620). AAAI Press/The MIT Press.

Fundaci Institut Guttmann. (2008). *Tecnologías aplicadas al proceso neurorrehabilitador: Estrategias para valorar su eficacia* (1st ed.). Barcelona: Fundacin Institut Guttmann.

Güler, I., Gökçil, Z., & Gülbandilar, E. (2009). Evaluating of traumatic brain injuries using artificial neural. *Expert Systems with Applications, 36*(7), 10424–10427. http://dx.doi.org/10.1016/j.eswa.2009.01.036.

Hartigan, J. A. (1975). *Clustering algorithms*. New York: John Wiley & Sons.

Haykin, S. (1994). *Neural networks, a comprehensive foundation*. New York: Macmillan College Publishing Company Inc.

Institut Guttmann Neurorehabilitation Hospital. (1997). Citing internet sources. URL: http://www.guttmann.com.

Ji, S. Y., Smith, R., Huynh, T., & Najarian, K. (2009). A comparative analysis of multi-level computer-assisted decision making systems for traumatic injuries. *Medical Informatics and Decision Making, 9*(2). http://dx.doi.org/10.1186/1472-6947-9-2.

Köksa, G., Batmaz, I., & Caner, M. (2011). A review of data mining applications for quality improvement in manufacturing industry. *Expert Systems with Applications, 38*(10), 13448–13467. http://dx.doi.org/10.1016/j.eswa.2011.04.063.

Marcano-Cedeño, A., Quintanilla-Domínguez, J., & Andina, D. (2011). WBCD breast cancer database classification applying artificial metaplasticity neural network. *Expert Systems with Applications, 38*(8), 9573–9579. http://dx.doi.org/10.1016/j.eswa.2011.01.167.

Orhan, U., Hekim, M., & Ozer, M. (2011). EEG signals classification using the K-means clustering and a multilayer perceptron neural network model. *Expert Systems with Applications, 38*(10), 13475–13481. http://dx.doi.org/10.1016/j.eswa.2011.04.149.

Pang, B. C., Kuralmani, V., Joshi, R., Hongli, Y., Lee, K. K., Ang, B. T., et al. (2007). Hybrid outcome prediction model for severe traumatic brain injury. *Journal of Neurotrauma, 24*(1), 136–146. http://dx.doi.org/10.1089/neu.2006.0113.

Pérez, R., Costa, U., Torrent, M., Solana, J., Opisso, E., Cceres, C., et al. (2010). Upper limb portable motion analysis system based on inertial technology for neurorehabilitation purposes. *Sensors, 10*, 10733–10751. http://dx.doi.org/10.3390/s101210733.

Polat, K., Şahan, S., & Günes, S. (2006). A new method to medical diagnosis: Artificial immune recognition system (AIRS) with fuzzy weighted pre-processing and application to ECG arrhythmia. *Expert Systems with Applications, 31*(2), 264–269. http://dx.doi.org/10.1016/j.eswa.2005.09.019.

Quinlan, J. R. (1993). *C4.5: Programs for machine learning* (1st ed.). San Francisco: Morgan Kaufmann Publishers.

Rovlias, A., & Kotsou, S. (2004). Classification and regression tree for prediction of outcome after severe head injury using simple clinical and laboratory variables. *Journal of Neurotrauma, 21*(7), 886–893. http://dx.doi.org/10.1089/0897715041526249.

Rughani, A. I., Dumont, T. M., Lu, Z., Josh Bongar, M. S., Horgan, M. A., Penar, P. L., et al. (2010). Use of an artificial neural network to predict head injury outcome. *Journal of Neurosurgery, 113*(3), 585–590. http://dx.doi.org/10.3171/2009.11.JNS09857.

Segal, M. E., Goodman, P. H., Goldstein, R., Hauck, W., Whyte, J., Graham, J. W., et al. (2006). The accuracy of artificial neural networks in predicting long-term outcome after traumatic brain injury. *Journal of Head Trauma Rehabilitation, 21*(4), 298–314.

Seng, J. L., & Chen, T. C. (2010). An analytic approach to select data mining for business decision. *Expert Systems with Applications, 37*(12), 8042–8057. http://dx.doi.org/10.1016/j.eswa.2010.05.083.

Sohlberg, M. M., & Mateer, C. A. (2001). *Cognitive rehabilitation: An integrative neuropsychological approach* (1st ed.). New York: Guilford Press.

Solana, J., Cáceres, C., Gómez, E.J., Ferrer-Celma, S., Ferre-Bergada, M., & Garcia-Lopez et al. (2011). PREVIRNEC a new platform for cognitive tele-rehabilitation. In *Proceedings of the third international conference on advanced cognitive technologies and applications, COGNITIVE 2011 September 25–30* (pp. 59–62). Rome, Italy.

Specht, D. F. (1991). A general regression neural network. *IEEE Transactions on Neural Networks, 6*, 568–576. http://dx.doi.org/10.1109/72.97934.

The Lancet Neurology. (2010). Traumatic brain injury: time to end the silence. *Lancet Neurology, 9*, 331. doi:10.1016/S1474-4422(10)70069-7.

Tormos, J. M., García-Molina, A., García-Rudolph, A., & Roig, T. (2009). Information and communication technology in learning development and rehabilitation. *International Journal of Integrated Care, 9*(5).

Upadhyay, D. (2008). Cognitive functioning in TBI patients: A review of literature. *Middle-East Journal of Scientific Research, 3*(3), 120–125.

<http://www.cs.waikato.ac.nz/ml/index.html>, 23th January 2011.

Witten, I. H., & Frank, E. (2005). *Data Mining Practical Machine Learning Tools and Techniques* (2nd ed.). San Francisco: Morgan Kaufmann Publishers.

Yeh, Y. L., Hou, T. H., & Chang, W. Y. (2012). An intelligent model for the classification of childrens occupational therapy problems. *Expert Systems with Applications, 39*(5), 5233–5242. http://dx.doi.org/10.1016/j.eswa.2011.11.016.

Yen, D. Y., Chen, C. H., & Chen, Y. W. (2011). A predictive model for cerebrovascular disease using data mining. *Expert Systems with Applications, 38*(7), 8970–8977. http://dx.doi.org/10.1016/j.eswa.2011.01.114.