Cognitive Prognosis of Acquired Brain Injury Patients Using Machine Learning Techniques

Joan Serrà, Josep Ll. Arcos
IIIA-CSIC, Artificial Intelligence Research Institute,
Spanish National Research Council,
Bellaterra, Spain.
Email: {jserra,arcos}@iiia.csic.es

Alejandro Garcia-Rudolph, Alberto García-Molina,
Teresa Roig Rovira, Josep M. Tormos
Institut Guttmann Neurorehabilitation Hospital, Badalona, Spain.
Email: alejandropablogarcia@gmail.com,
{agarciam,troig,jmtormos}@guttmann.com

***Abstract*—The cognitive prognosis of acquired brain injury (ABI) patients is a valuable tool for an improved and personalized treatment. In this paper, we explore the task of automatic cognitive prognosis of ABI patients via machine learning techniques. Based on a set of pre-treatment assessments, distinct classifiers are trained to predict whether the patient will improve in one or any of three cognitive areas: attention, memory, and executive functioning. Results show that variables such as the age at the moment of the injury, the patient's etiology, or the neuropsychological evaluation scores obtained before the treatment are relevant for prognosis and easily yield statistically significant accuracies. Additionally, the prognostic relevance of these and other variables is studied by means of standard feature selection methodologies. The outputs of the present paper add to the discussion on current cognitive rehabilitation practices and push towards the exploitation of existing technologies for improving medical evaluations and treatments.**

***Keywords*—machine learning; brain injury; prognosis; classifiers; cognitive rehabilitation; neuropsychological evaluations.**

I. Introduction

Acquired brain injury (ABI) is a leading cause of death and disability worldwide [24]. It is considered one of the most common neurological disorders and, for survivors, it is widely regarded as a very debilitating condition [13]. ABI patients experience multiple impairments, specially physical (e.g., mobility, vision, sleep) and/or cognitive (e.g., attention, memory, executive function, language) impairments.

Cognitive impairments are particularly problematic, since they can limit daily activities and restrict participation in community, employment, recreation, and social relationships [11]. In particular, disturbances in basic cognitive functions such as attention and memory may cause or exacerbate additional disturbances in executive functioning, communication, and other relatively more complex cognitive functions. Attention is defined as a set of multifaceted processes including abilities to select relevant stimuli, manipulate and contain mental images and modulate responses to the environment [23]. Memory is the process of encoding, storing and retrieving information. Executive functions are those functions that allow us to function effectively and adaptively succeed within our social contexts [23]. Cognitive measures are among the most important predictors of patients' return to work and independent living, even among those with good medical recoveries [6].

The design of therapies for improving or potentially recovering the cognitive abilities of ABI patients is still an open issue [4]. Determining the appropriate method and timing of treatment for an individual with ABI depends on a number of factors, including severity of injury, stage in recovery, and premorbid, comorbid and environmental conditions, unique to each individual [15]. Although there is substantial evidence for cognitive rehabilitation treatments, additional research is required to guide the development of better clinical practices, particularly with respect to selecting the most effective treatment for a particular patient [4], [15].

A valuable tool for a more effective treatment of ABI patients is prognosis [18], i.e., anticipating the treatment's outcome from the usual course of the disease and/or the peculiarities of each individual case. Outcome prediction and early identification of reliable prognostic factors is of paramount importance to direct treatments, shape general policies, identify critical subjects, adapt treatment protocols to specific individuals, perform a more exhaustive monitoring of selected patients, and much else [18]. However, predictive modeling is particularly difficult when considering the numerous complex clinical elements that occur after ABI and their interplay.

In this paper, we exploit machine learning techniques [9], [10], [16] to predict the expected cognitive outcome of ABI patients using pre-treatment diagnosis data (prognosis). A number of studies employ machine learning techniques for the automatic prognosis of ABI patients [2], [3], [17], [19], [21]. Decision trees are the most common choice [2], [3], [17], [19], but also neural networks [17], [21] or different regression models [2], [17], [21] are used. Overall, these studies focus on determining survival, predicting gross outcome, and/or identifying predictive factors of a patient's condition after traumatic brain injury (TBI; usually acute TBI). In addition, to the best of our knowledge, no studies focus on long-term cognitive rehabilitation and, in particular, on the neuropsychological evaluations that are commonly used for assessing improvements at the cognitive level [23].

Our study shows that pre-treatment diagnosis data is predictive of ABI patients' response to cognitive rehabilitation. Specifically, we show that initial neuropsychological evaluations, and also their combination with generic information such as the patient's age, studies, or the cause of the injury, have a considerable power for predicting treatment responses. Moreover, our results suggest that such predictive power relies on the data itself, as similar accuracies are obtained by a number of machine learning algorithms based on different principles. An additional goal of our study is to see whether our specific results add to current knowledge of relevant risk

TABLE I. SUMMARY OF DIAGNOSIS DATA (SEE TEXT). PRE-EVALUATION SCORES CORRESPOND TO: NO IMPAIRMENT (0), MILD IMPAIRMENT (1), MODERATE IMPAIRMENT (2), SEVERE IMPAIRMENT (3), AND VERY SEVERE IMPAIRMENT (4). THE LETTER υ DENOTES MEAN $\pm$ STANDARD DEVIATION.

Demographic data		Pre-evaluation tests ([0,4])		Pre-evaluation diagnosis ([0,4])	
Gender	{'male', 'female'}	1	Digit span forward WAIS	Spec. 1	Categorization
Studies	{'no studies', 'primary',	2	Trail marking test, part A	Spec. 2	Divided attention
	'secondary', 'degree'}	3	Stroop word	Spec. 3	Flexibility
Age at injury	$[17, 76]$; $\upsilon = 40.6 \pm 14.5$	4	Stroop color	Spec. 4	Inhibition
Age treatment	$[17, 76]$; $\upsilon = 41.2 \pm 14.5$	5	Stroop word-color	Spec. 5	Planning
Delay treatment	$[0, 31]$; $\upsilon = 1.1 \pm 2.8$	6	Digit symbol WAIS	Spec. 6	Sequencing
Treatment weeks	$[1, 77]$; $\upsilon = 17.5 \pm 12.8$	7	Block design WAIS	Spec. 7	Selective attention
Sessions per week	$[1, 5]$; $\upsilon = 3.0 \pm 1.3$	8	Digit span backward WAIS	Spec. 8	Sustained attention
Etiology (specific)	{'TBI', 'multiple sclerosis',	9	Letter-number sequencing WAIS	Spec. 9	Working memory
	'hemorrhagic stroke',	10	RAVLT short-term memory	Spec. 10	Verbal memory
	'ischemic-thrombotic stroke',	11	RAVLT long-term memory	Spec. 11	Visual memory
	'ischemic-embolic stroke',	12	RAVLT recognition	Gen. 1	Attention
	'ischemic undetermined stroke',	13	Trail marking test, part B	Gen. 2	Executive functions
	'other non-TBI', 'other'}	14	WCST categories	Gen. 3	Memory
Etiology (general)	{'stroke', 'TBI', 'other'}	15	WCST perseverative errors		
		16	Stroop interference		
		17	PMR maximally produce words		

factors or help in assessing critical values of the considered pre-treatment data.

The remainder of the paper is organized as follows. We first present our methodology (Sec. II), including a description of the considered data (Sec. II-A), our feature pre-processing steps (Sec. II-B), the machine learning tools we use (Sec. II-C), and the followed evaluation strategy (Sec. II-D). We next show the obtained results and discuss them to some detail (Sec. III). A brief summary section concludes the paper (Sec. IV).

II. MATERIALS AND METHODS

A. Diagnosis data

The considered data comes from PREVIRNEC©, a web-based tele-rehabilitation platform conceived as a tool to enhance cognitive rehabilitation [22]. Every participant considered in this analysis underwent a pre-treatment evaluation involving the three main cognitive functions (*attention*, *memory*, and *executive functions*) by means of a standard tests battery detailed below. After treatment, participants were again evaluated using the same tests battery to measure improvement/non-improvement in the respective functions. A pool of 503 patients is considered: all of them were assessed in *attention* (299 improved), 496 in *memory* (317 improved), and 501 in *executive functions* (334 improved). We additionally consider a further category, *any*, where we assess whether there is an improvement in, at least, one cognitive function, and no worsening in any of the others (503 assessments, 368 improved). In total, we face four binary classification [16] problems (two classes: improvement/non-improvement in *attention*, *memory*, *executive functions*, and *any*). As input variables we dispose of (Table I):

- Demographic data: *gender*, level of *studies*, and the patient's age at the time of the injury (denoted by *age at injury*).
- Clinical data: this includes a *general etiology* description and a more *specific* one. It also includes a neuropsychological assessment battery consisting of 17 *tests* across the three main cognitive functions [23]. The obtained *test* scores are combined into 11 *specific diagnosis* scores, representing the respective sub-functions of attention (sustained, selective, divided), memory (visual, verbal, working), and executive functions (inhibition, planning, flexibility, sequencing, categorization). *Specific diagnosis* scores are further summarized into 3 *general diagnosis* scores, corresponding to the 3 main cognitive functions (*attention*, *executive functions*, and *memory*).
- Treatment data: the patient's age at the time of starting the treatment (denoted by *age treatment*), the delay between injury and treatment (in years, named *delay treatment*), the treatment duration (in weeks, named *treatment weeks*), and the number of *sessions per week*.

B. Feature pre-processing

The previous data comprises qualitative as well as quantitative variables (nouns/text and numbers, respectively). As quantitative variables are majority, we first convert qualitative variables into quantitative features. In particular, binary variables are directly coded as binary features and m-level qualitative variables are coded as vectors of m binary features [10]. For instance, with our data, *gender* = {'male', 'female'} becomes *gender* = {0, 1} and *studies* = {'no studies', 'primary', 'secondary', 'degree'} becomes *studies* = {{1,0,0,0}, {0,1,0,0}, {0,0,1,0}, {0,0,0,1}}. For dates, we use only the integer corresponding to the year, and thus, e.g., *date* = '1998/04/27' becomes *date* = 1998. Fields with integer or real values are kept as they are.

Before training our classifiers we normalize all features to a common range, considering a low and a high percentile for each feature. Specifically, we re-scale individual features so that their values at the 5 and 95 percentiles correspond to 0 and 1, respectively. These percentile values are kept for applying the same normalization at the testing stage. To avoid ties in feature vectors we add a small jitter η after normalization, $\eta = 10^{-6}\xi$, where ξ is a Gaussian random number generator with zero mean and unit variance.

A few missing values are present in our data. A fraction of patients have not performed *pre-evaluation tests 14* ($\approx$5%), *15* ($\approx$5%), and *16* ($\approx$0.4%). In addition, we do not dispose of the *age at injury* and, consequently, of the *delay treatment* for approximately 4% of the patients. In all these missing

value cases we opt for distribution-based imputation [20]. For *pre-evaluation tests*, we impute distribution means, directly computed from the other patients that have completed the corresponding evaluation test. For *age at injury* and *delay treatment* we proceed similarly.

C. Machine learning tools

To show that the predictive power of the considered features is generic and not biased towards a specific classification scheme, we employ basic algorithms exploiting four different machine learning principles [10], [16]: decision tree learning, instance-based learning, probabilistic learning, and support vector machines. The implementations we use come from the scikits-learn package (version 0.10: http://scikit-learn.org) and, unless stated otherwise, their default parameters are taken. In total we use six implementations [10], [16]:

- Tree: Classification and regression tree (CART) classifier. We use the Gini coefficient as the measure of node impurity and an arbitrarily set minimum number of 7 instances per leaf.
- KNN: k-Nearest neighbor classifier. We use the Euclidean distance and an arbitrary value of $k = 9$.
- NB: Naive Bayes classifier. We loosely employ a Gaussian function to estimate the likelihood of all of features.
- SVM: Support vector machine. We consider a linear kernel (SVM_L), a polynomial kernel of degree 2 (SVM_P), and a radial basis function kernel (SVM_R).

Apart from classification performance, we also apply some alternative/complementary techniques to assess the importance of individual features and groups of them. For this part of the analysis we use balanced data samples (same number of instances per class) from the full data set and resort to the Weka package [9] (version 3.6.6: http://www.cs.waikato.ac.nz/ml/weka/) for algorithm implementations, also taking the default parameters, if not stated otherwise. Depending on the required assessment, we look at the binary splits of the tree-based classifier [16], the feature weights assigned by SVM_L [8], or the feature rankings produced by χ^2 feature evaluation [9]. Additionally, we consider class-conditioned feature distributions [10].

D. Evaluation measure and statistical significance

We measure binary classification performance with the out-of-sample percentage of correctly classified instances. We perform a 20 times 10-fold cross-validation on balanced data (same number of instances per class in train and test sets; two classes) and take the average accuracy [10]. For space reasons we omit confusion matrices and class/label-dependent accuracies (in the big majority of cases we obtained rather even confusion matrices and thus very similar accuracies for improvement or non-improvement classes).

To assess the merit of the classifiers' predictions we run the same experiment with a randomized data set with shuffled feature values. This way, we maintain the same distribution for each feature and keep the original dimensionality of the problem. The accuracies for this random baseline, always around 50%, can then be used to assess the statistical significance of the increment provided by the original features under the same classification algorithm. For determining statistical significance we employ the Wilcoxon signed-rank test [12] on the 200 individual accuracy values obtained for each fold. The Wilcoxon signed-rank test is a non-parametric statistical hypothesis test used when comparing two related samples (or two repeated measurements on a single sample) in order to assess whether their population mean ranks differ. We use a two-tailed p-value of 0.01 but apply the Bonferroni adjustment to compensate for multiple tests [1]. Considering 6 classifiers, 4 cognitive functions, and a number of data trials below 50 we have a final p^*-value of $p^* = 0.01/(6 \cdot 4 \cdot 50) = 8.33 \cdot 10^{-6}$. Notice that the combined use of a non-parametric test for related samples together with the aforementioned Bonferroni adjustment represents a strict and conservative criteria for determining statistical significance (cf. [1], [5]). Therefore, it enforces a high standard for reporting that a set of accuracies for a given classifier, test label, and data trial is better than another.

III. Results and discussion

We start looking at the predictive power of individual concepts (Table II). We see that *gender* never achieves a single statistically significant accuracy. Hence, when taken alone, it can be regarded as irrelevant for prognosis. The level of *studies* presents some controversy. Although this concept is frequently used as a proxy for cognitive reserve [14], we obtain few statistically significant accuracies, and these are usually below 55%. Thus, we cannot strongly confirm its use as a proxy. In future work we plan a deeper study of this issue.

According to our results, two highly prognostic concepts are *age at injury* and *age treatment*, with statistically significant accuracies above 55% most of the time (notice that they are highly correlated: $\rho = 0.98$, $p < 10^{-6}$). A tree split analysis (Sec. II-C) for these two individual concepts reveals two thresholds at which the possibilities for *any* improvement relatively diminish: below 37 years old we find more recoveries than non-recoveries and above 57 years old we find more non-recoveries than recoveries. The time elapsed between the injury and the beginning of the treatment (*delay treatment*) is not much predictive of the patient's improvement after treatment. However, we could assume some prognostic value in the case of *attention*, where all classifiers report statistically significant accuracies (Table II, top left). Indeed, with the class-conditioned distributions for *attention* we see a slight tendency towards non-improvement for delays larger than 1 or 2 years. In the future, we will consider days or weeks instead of years as units, so that a better refinement is possible.

Clearly, the most informative concepts are *specific* and *general etiologies* and *pre-evaluation tests* and *diagnoses* (Table II). In particular, we see that they all reach statistically significant accuracies beyond 55% most of the time, independently of the classifier used (Table II, middle rows). A further inspection with the χ^2 feature ranker considering all *etiologies* deems the *general etiologies* 'stroke' and 'other' as very relevant, together with the *specific etiology* 'other'. The SVM_L weights also point to 'strokes' as a relevant *general etiology* for prognosis, and specially to 'schemic-thrombotic stroke' and 'undetermined stroke' *specific etiologies*. These two are usually associated with improvement.

TABLE II. CLASSIFICATION ACCURACIES FOR DIFFERENT CONCEPTS, FUNCTIONS, AND CLASSIFIERS. FOR EASE OF VISUALIZATION, ONLY STATISTICALLY SIGNIFICANT ACCURACIES ARE SHOWN (BASELINE RANDOM ACCURACY IS CLOSE TO 50%, SEE SEC. II-D). THE LARGEST ACCURACIES FOR EACH COGNITIVE FUNCTION AND CLASSIFIER ARE SHOWN IN BOLD. THE FIRST ROWS OF EACH TABLE CORRESPOND TO SINGLE CONCEPTS AND THE LAST ONES TO A COMBINATION OF CONCEPTS.

Data	*Attention*						*Memory*					
	Tree	KNN	NB	SVM_L	SVM_P	SVM_R	Tree	KNN	NB	SVM_L	SVM_P	SVM_R
Gender												
Studies			51.8		53.8	53.5		50.8	50.2		54.7	55.0
Age at injury			58.2	58.1	56.7	58.0	53.4	53.1	57.1	56.8	56.6	56.5
Age treatment			**58.5**	58.5	56.8				56.0	57.6	57.1	56.7
Delay treatment	53.7	53.6	54.0	56.3	53.9	55.7			50.9			
Treatment weeks	54.5			57.3		57.8				51.4		
Sessions per week									51.0	50.2		
Etiology (specific)		51.9	52.6	55.4	55.1	55.9		54.1	54.3	60.7	60.1	59.9
Etiology (general)			56.5	57.1	57.1	56.5		55.2	59.6	59.7	59.7	59.5
Pre-evaluation (tests)	53.5		53.4		55.8	54.6	61.4	64.1	67.8	65.7	**66.1**	66.3
Pre-evaluation (specific diagnoses)	52.8	55.4	55.2	55.7	56.0	55.1	61.9	**65.2**	66.5	65.5	64.8	66.4
Pre-evaluation (general diagnoses)			54.4		54.6	55.5	62.1	64.7	**68.3**	**66.9**	62.9	66.9
Etiology (specific+general)				55.7	56.8	56.0	53.2	54.9	55.7	60.7	59.4	59.2
Pre-evaluation (all diagnoses)	52.8	56.5	54.8	55.1	55.3		61.8	65.3	67.1	65.3	64.2	65.6
Pre-evaluation (tests+diagnoses)	54.7	54.7	54.5	54.9	56.1	55.8	**62.5**	64.9	67.0	65.8	66.0	66.7
Etiology (gen.) + Pre-eval (diag.)	53.9	55.7	56.4	58.3	**58.8**	58.0	62.3	65.1	66.9	66.1	65.8	66.1
Etiol(g)+Pre-eval(d)+AgeInj+Delay	**55.3**	57.0	56.7	59.8	58.6	58.5	60.2	65.0	67.4	66.0	65.2	66.1
Informative mixture (see text)	55.1	**57.9**	52.6	**60.8**	58.1	**61.0**	60.6	64.0	60.6	66.5	64.8	**67.7**

Data	*Executive Functions*						*Any*					
	Tree	KNN	NB	SVM_L	SVM_P	SVM_R	Tree	KNN	NB	SVM_L	SVM_P	SVM_R
Gender												
Studies					50.4				50.4	55.6	56.0	
Age at injury			57.4	57.2	56.8	56.8			57.7	58.0	57.7	58.6
Age treatment			56.9	57.7	56.9	57.9		52.7	58.1	57.9	58.5	58.9
Delay treatment				55.8		54.5			52.7	55.2		52.7
Treatment weeks		52.9				53.5	51.7	54.1		51.2		
Sessions per week												
Etiology (specific)			53.0	56.3	56.7	57.2		55.5	56.2	59.8	58.3	58.6
Etiology (general)			55.8	57.0	56.3			55.2	58.7	57.1	58.9	58.5
Pre-evaluation (tests)		56.0	60.4	57.3	58.8	58.5	57.5	61.1	62.3	61.4	62.2	63.0
Pre-evaluation (specific diagnoses)	55.9	56.4	60.8	58.2	59.6	59.3	58.3	60.9	62.1	62.9	62.0	63.5
Pre-evaluation (general diagnoses)	56.8	57.9	61.3	59.7	58.3	59.4	58.1	60.9	62.5	64.3	61.8	62.7
Ethiology (specific+general)		53.9	53.0	54.7	57.0		53.2	55.5	55.3	59.0	58.8	59.6
Pre-evaluation (all diagnoses)	56.9	56.5	61.5	57.6	59.5	58.9	**60.0**	60.3	62.6	63.0	62.3	63.6
Pre-evaluation (tests+diagnoses)	55.5	57.7	61.2	55.9	58.7	58.0	57.2	60.6	62.8	59.8	62.5	62.4
Etiology (gen.) + Pre-eval (diag.)	57.0	57.2	**61.6**	59.3	**60.6**	**60.4**	58.9	61.3	**64.0**	63.1	62.7	63.4
Etiol(g)+Pre-eval(d)+AgeInj+Delay	**57.3**	**59.2**	61.2	60.2	59.8	59.9	59.9	**61.5**	62.9	62.9	63.1	63.1
Informative mixture (see text)	55.7	56.6	60.1	**60.8**	60.2	60.2	57.8	60.8	59.0	**64.5**	**64.2**	**64.8**

The best accuracies though are achieved by *pre-evaluation tests* and *diagnoses*, which generally score around or above 60%. The only cognitive function that could be an exception is *attention*. For this, perhaps etiology is more predictive. In general, accuracies for individual *pre-evaluations* are around 55% for *attention*, 66% for *memory*, 59% for *executive functions*, and 61% for *any*. Noticeably, we find that the predictive power of *pre-evaluation* scores (*tests* and *diagnoses*) comes from very high or very low values. In fact, looking at the decision trees and the class-conditioned distributions we see that, in general, score values below 2 (moderate impairment) tend to indicate improvement, whereas score values above 3 (severe impairment) tend to indicate non-improvement.

Overall, the combinations of *general* and *specific etiologies*, or *general* and *specific pre-evaluation* scores, do not increase accuracy significantly (Table II, middle rows). This was somehow expected, as the information of the *general pre-evaluation diagnosis* is derived from the *specific pre-evaluation diagnosis* which, in turn, is derived from the *pre-evaluation test* scores (see Sec. II-A). A similar reasoning can be made with *general* and *specific etiologies*. However, by combining etiologies or pre-evaluations themselves, we generally see that a larger number of classifiers reach statistically significant accuracies. Thus, we could say that their predictive power is somehow reinforced.

When we do see an accuracy increment is when mixing these two concepts (*etiologies* and *pre-evaluation* scores), or when further considering other slightly predictive concepts such as *age at injury* or *delay treatment* (Table II, bottom rows). The best results are probably achieved by an arbitrary combination of different concepts, excluding some non-significant and some correlated ones. The "informative mixture" result in Table II corresponds to combining all *diagnosis pre-evaluations* with *general etiology*, *age at injury*, *age treatment*, *treatment weeks*, and *studies*. Notice that the fact that some concepts are non-significant/correlated when taken individually does not imply that they are useless for classification when combined with other features [7]. The

TABLE III. TEN MOST INFORMATIVE FEATURES ACCORDING TO THE CHOSEN FEATURE RELEVANCE ANALYSIS METHODS (FROM A TOTAL OF 52 FEATURES). FEATURES CHOSEN BY AT LEAST TWO METHODS ARE HIGHLIGHTED IN BOLD. WE SEE THAT METHODS AGREE IN MANY OF THEM. REPEATEDLY CHOSEN FEATURES THAT DO NOT DIRECTLY MATCH THE GENERAL COGNITIVE FUNCTION ARE HIGHLIGHTED WITH THE * SYMBOL.

Attention	Feature relevance analysis method		
	Tree	SVM_L	χ^2
1	**Pre-eval. test 7 (exec. func.)***	Pre-eval. test 16 (exec. func.)	**Pre-eval. diag. spec. 6 (exec. func.)***
2	**Pre-eval. diag. spec. 6 (exec. func.)***	**Treatment weeks**	**Pre-eval. diag. spec. 2 (attention)**
3	**Age treatment**	**Pre-eval. test 7 (exec. func.)***	Pre-eval. test 13 (attention)
4	**Treatment weeks**	Pre-eval. test 15 (exec. func.)	**Pre-eval. diag. spec. 8 (attention)**
5	Pre-eval. test 8 (memory)	**Age treatment**	**Age treatment**
6	**Pre-eval. diag. gen. 1 (attention)**	Pre-eval. test 3 (attention)	**Delay treatment**
7	Pre-eval. test 12 (memory)	Age at injury	**Pre-eval. test 7 (exec. func.)***
8	**Pre-eval. test 10 (memory)**	**Delay treatment**	**Pre-eval. diag. gen. 1 (attention)**
9	**Pre-eval. diag. spec. 2 (attention)**	Pre-eval. test 2 (attention)	Pre-eval. test 10 (memory)
10	Studies 1 (primary)	**Pre-eval. diag. spec. 8 (attention)**	Etiology spec. 8 (TCE)
Memory	Feature relevance analysis method		
	Tree	SVM_L	χ^2
1	**Pre-eval. diag. gen. 3 (memory)**	**Pre-eval. test 17 (exec. func.)***	**Pre-eval. diag. gen. 3 (memory)**
2	**Pre-eval. diag. spec. 9 (memory)**	Pre-eval. test 13 (attention)	**Pre-eval. test 2 (attention)***
3	Sessions per week	**Pre-eval. test 5 (attention)***	Pre-eval. diag. spec. 11 (memory)
4	Pre-eval. diag. spec. 6 (exec. func.)	Etiology spec. 3 (exec. func.)	Pre-eval. diag. spec. 10 (memory)
5	**Pre-eval. test 2 (attention)***	Pre-eval. diag. spec. 1 (attention)	**Pre-eval. diag. spec. 9 (memory)**
6	**Pre-eval. test 5 (attention)***	**Pre-eval. test 6 (attention)***	**Pre-eval. test 5 (attention)***
7	**Pre-eval. test 17 (exec. func.)***	Pre-eval. diag. spec. 8 (attention)	Pre-eval. test 12 (memory)
8	Studies 1 (primary)	Pre-eval. test 16 (attention)	**Pre-eval. test 4 (attention)***
9	Age treatment	**Pre-eval. test 4 (attention)***	Pre-eval. test 3 (attention)
10	Treatment weeks	Gender	**Pre-eval. test 6 (attention)***
Executive functions	Feature relevance analysis method		
	Tree	SVM_L	χ^2
1	**Pre-eval. diag. spec. 10 (memory)***	**Etiology spec. 1 (multiple sclerosis)**	**Pre-eval. diag. spec. 5 (exec. func.)**
2	**Pre-eval. diag. spec. 5 (exec. func.)**	**Delay treatment**	**Pre-eval. diag. gen. 2 (exec. func.)**
3	Treatment weeks	**Pre-eval. test 12 (memory)***	Pre-eval. diag. spec. 4 (exec. func.)
4	**Pre-eval. diag. gen. 2 (exec. func.)**	Pre-eval. test 5 (attention)	Pre-eval. diag. spec. 6 (exec. func.)
5	Age at injury	Etiology spec. 5 (undetermined stroke)	Pre-eval. test 9 (exec. func.)
6	**Etiology spec. 1 (multiple sclerosis)**	**Pre-eval. test 8 (memory)***	Pre-eval. diag. spec. 11 (memory)
7	**Pre-eval. test 12 (memory)***	Pre-eval. test 15 (exec. func.)	**Pre-eval. diag. spec. 10 (memory)***
8	**Pre-eval. test 1 (memory)***	Etiology spec. 6 (other non-TBI)	**Pre-eval. test 8 (memory)***
9	**Pre-eval. test 17 (exec. func.)**	Pre-eval. test 4 (attention)	**Pre-eval. test 1 (memory)***
10	**Delay treatment**	Pre-eval. test 2 (attention)	**Pre-eval. test 17 (exec. func.)**
Any	Feature relevance analysis method		
	Tree	SVM_L	χ^2
1	**Pre-eval. diag. spec. 10 (memory)**	**Pre-eval. diag. spec. 8 (attention)**	**Pre-eval. diag. gen. 3 (memory)**
2	**Etiology gen. 2** (other)	Etiology spec. 3 (ischemic-thrombotic stroke)	Pre-eval. diag. spec. 11 (memory)
3	**Pre-eval. diag. gen. 3 (memory)**	**Pre-eval. test 5 (attention)**	**Pre-eval. diag. spec. 10 (memory)**
4	Age treatment	Pre-eval. test 11 (memory)	Pre-eval. test 12 (memory)
5	**Pre-eval. test 2 (attention)**	Pre-eval. test 15 (exec. func.)	**Pre-eval. test 2 (attention)**
6	**Pre-eval. diag. spec. 5 (exec. func.)**	Pre-eval. diag. spec. 1 (exec. func.)	**Pre-eval. test 5 (attention)**
7	Age at injury	Pre-eval. test 2 (attention)	Pre-eval. diag. spec. 4 (exec. func.)
8	**Pre-eval. diag. spec. 8 (attention)**	Pre-eval. test 3 (attention)	Pre-eval. diag. spec. 8 (attention)
9	Pre-eval. diag. spec. 6 (exec. func.)	Pre-eval. test 9 (exec. func.)	Pre-eval. diag. spec. 9 (memory)
10	Pre-eval. diag. gen. 1 (attention)	**Etiology gen. 2 (other)**	**Pre-eval. diag. spec. 5 (exec. func.)**

accuracies achieved by combining concepts are practically always above 60% in all cognitive functions. The highest ones correspond to 61% for *attention*, 67% for *memory*, 61% for *executive functions*, and 64% for *any*. Combining all concepts did not yield to any notable improvement over the other combinations shown in Table II.

We finally perform a brief feature relevance analysis. From the pool of all available features, we run the chosen feature relevance analysis methods (Sec. II-C) and show the 10 best ranked features for each one (Table III). As expected, we see that *pre-evaluation* scores (both *tests* and *diagnoses*) are the majority among the most relevant features. Additionally, we see that *age treatment* and *delay treatment* appear frequently among the 10 best features.

For every cognitive function we find some 'obviously selected' *pre-evaluations*. For instance, *general diagnosis 1*, which evaluates attention, is selected in *attention*, or *specific diagnosis 9*, which corresponds to working memory, is selected for *memory*. There are a number of these rather obvious correspondences. However, we see some *pre-evaluations* that do not directly match the cognitive function they help to predict. The full account of such *pre-evaluations* can be gathered from Table III. We now enumerate some of them: regarding *attention*, we find *test 7* (block design WAIS, executive functions), *test 10* (RAVLT short-term memory, memory), and *specific diagnosis 6* (sequencing, executive functions); regarding *memory*, we find *test 2* (trail marking test part A, attention), *test 5* (stroop word-color, attention), and *test*

6 (digit symbol WAIS, attention); finally, regarding *executive functions*, we find *test 1* (digit span forward WAIS, attention), *test 8* (digit span backward WAIS, memory), and *test 12* (RAVLT recognition, memory).

From our point of view, all these emerging associations between tests/diagnoses and cognitive functions only highlight the interconnectedness of our brain and, therefore, the large dependencies that exist between different cognitive functions. If these associations acquire support or show some persistence in future investigations, one could potentially think of additionally considering them for the assessment of those cognitive functions whose improvement they help to predict. As shown, machine learning techniques can bring valuable guidance in discovering such hidden connections.

IV. Conclusion

In this paper, we provide an application of machine learning techniques to assess acquired brain injury data. In particular, we focus on automatic cognitive prognosis, i.e., the task of predicting whether the patient will improve in a number of cognitive functions using only pre-treatment data. Our contribution shows that variables such as age at injury, etiology, or neuropsychological evaluation scores are relevant for prognosis, yielding statistically significant prediction accuracies. Importantly, the obtained results are independent of the classification scheme, what stresses the predictive power of the considered variables. Finally, machine learning techniques also prove capable of discovering hidden and emerging relations involving pre-evaluation tests and the studied cognitive functions. Overall, these are largely unexplored areas with a high and valuable potential.

In future work we plan to include treatment data to our analysis. This way, combining treatment with diagnosis data, we may be able to advance the outcome of a new patient from a pool of previous patients. In particular, considering both diagnosis data and the performance of the treatment activities that have been already carried out, similar machine learning techniques could be trained to assess whether the patient is correctly responding to treatment or whether such treatment needs to be revised.

Acknowledgments

We thank all the patients and staff from Institut Guttmann who cooperated in data collection. This work has been partially funded by TIN-2012-38450-C03-03 from the Spanish Government (all authors), JAEDOC069/2010 from Consejo Superior de Investigaciones Científicas (J.S.), and 2009-SGR-1434 from Generalitat de Catalunya (J.S. and J.Ll.A.).

References

[1] H. Abdi. "Bonferroni and Sidak corrections for multiple comparisons". In N. J. Salkind, editor, Encyclopedia of Measurement and Statistics, pages 103–107. SAGE Publications, Thousand Oaks, USA, 2007.

[2] P. J. D. Andrews 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, 97:326–336.

[3] A. W. Brown et al. "Clinical elements that predict outcome after traumatic brain injury: a prospective multicenter recursive partitioning (decision-tree) analysis". Journal of Neurotrauma, 2005, 22(10):1040–1051.

[4] K. D. Cicerone et al. "Evidence based cognitive rehabilitation: updated review of the literature from 2003 through 2008". Archives of Physical Medicine and Rehabilitation, 2011, 92(4):519–530.

[5] J. Demsar. "Statistical comparison of classifiers over multiple data sets". Journal of Machine Learning Research, 2006, 7:1–30.

[6] P. J. Eslinger, G. Zappalà, F. Chakara, and A. M. Barrett. "Cognitive impairments". In N. D. Zasler, D. I. Katz, and R. D. Zafonte, editors, Brain injury medicine: principles and practice, chapter 59, pages 990–1001. Demos Medical Publishing, New York, USA, 2nd edition, 2011.

[7] I. Guyon and A. Elisseeff. "An introduction to variable and feature selection". Journal of Machine Learning Research, 2003, 3:1157–1182.

[8] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik. "Gene selection for cancer classification using support vector machines". Machine Learning, 2002, 46(1-3):389–422.

[9] M. Hall et al. "The WEKA data mining software: an update". ACM SIGKDD Explorations Newsletter, 2009, 11(1):10–18.

[10] T. Hastie, R. Tibshirani, and J. Friedman. The elements of statistical learning. Springer, Berlin, Germany, 2nd edition, 2009.

[11] J. M. Hoffman et al. "Understanding pain after traumatic brain injury: impact on community participation". American Journal of Physical Medicine and Rehabilitation, 2007, 86(12):962–969.

[12] M. Hollander and D. A. Wolfe. Nonparametric statistical methods. Wiley, New York, USA, 2nd edition, 1999.

[13] J. M. Jones, J. Jetten, S. A. Haslam, and W. H. Williams. "Deciding to disclose: the importance of maintaining social relationships for well-being after acquired brain injury". In J. Jetten, C. Haslam, and S. A. Haslam, editors, The social cure: identity, health and well-being, chapter 14, pages 255–272. Psychology Press, New York, USA, 2012.

[14] R. N. Jones et al. "Conceptual and measurement challenges in research on cognitive reserve". Journal of the Int. Neuropsychological Society, 2011, 17:593–601.

[15] R. Koehler, E. E. Wilhelm, and I. Shoulson. Cognitive rehabilitation therapy for traumatic brain injury: evaluating the evidence. The National Academic Press, Washington, USA, 2011.

[16] T. M. Mitchell. Machine Learning. McGraw-Hill, New York, USA, 1997.

[17] B. C. Pang et al. "Hybrid outcome prediction model for severe traumatic brain injury". Journal of Neurotrauma, 2007, 24(1):136–146.

[18] P. Perel et al. "Predicting outcome after traumatic brain injury: practical prognostic models based on large cohort of international patients". BMJ, 2008, 336(7641):425–429.

[19] A. Rovlias and S. Kotsou. "Classification and regression tree for prediction of outcome after severe head injury using simple clinical and laboratory variables". Journal of Neurotrauma, 2004, 21(7):886–893.

[20] M. Saar-Tsechansky and F. Provost. "Handling missing values when applying classification models". Journal of Machine Learning Research, 2007, 8:1625–1657.

[21] M. E. Segal 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.

[22] J. Solana et al. "PREVIRNEC, a new platform for cognitive tele-rehabilitation". In Proc. of the Int. Conf. on Advanced Cognitive Technologies and Applications (COGNITIVE), 2011, pp. 59–62.

[23] E. Strauss, E. M. S. Sherman, and O. Spreen. "A compendium of neuropsychological tests: administration, norms and commentary". Oxford University Press, Oxford, UK, 3rd ed. edition, 2006.

[24] World Health Organization. Neurological disorders: public health challenges. WHO Press, Geneva, Switzerland, 2006.