Proceedings of the
7th Graz Brain-Computer Interface Conference 2017 DOI: 10.3217/978-3-85125-533-1-51

Can feature selection be used to detect physiological components in P300 based BCI for Amyotrophic lateral sclerosis patients?

L. Bianchi[1], M. Cosmi[1], C. Liti[1], V. Piccialli[1]

[1]DICII, University of Rome Tor Vergata, Rome, Italy

E-mail: chiara.liti@uniroma2.it

ABSTRACT: The detection of brain state changes can dramatically improve the comprehension of cerebral functioning. To reach this aim, machine learning based automatic tools may be extremely useful to correctly classify different brain responses. The performance of these instruments depends on the features and the classification algorithm employed, but also from a good data preprocessing able to improve the poor signal-to-noise ratio [4] of the EEG signal. In this work, we combine data preprocessing with a feature selection based on the filter ReliefF and the linear SVM classifier LibLinear in order to analyse the data deriving from a P300 speller paradigm on patients with Amyotrophic lateral sclerosis (ALS). The purpose of this study is twofold: on the one hand we want to maximize the predictor's performance, but most importantly, we aim at showing how the features ranking can be used to support scientific hypotheses or diagnoses.

INTRODUCTION

In neuroscience, a fundamental theme is the study of brain functioning, for different scopes, such as neurorehabilitation, diagnosis support and brain activity monitoring in general. The detection of brain state changes plays a fundamental role because it can dramatically improve the comprehension of cerebral functioning. Evoked potentials, for example, which are the electrical responses recorded from the brain after specific stimulations, are widely used by researchers and clinicians to support scientific hypotheses [7] or to make diagnoses [8]. Recently, in [1] a feature ranking approach combined with SVM classifier was applied over the EEG signal of nine healthy subjects. The subset of features identified for each subject was physiologically correct. Indeed the filter was able to detect physiological components elicited during the protocol either in space or in latency. In this work we try to use a similar approach over Amyotrophic lateral sclerosis (ALS) patients, even thought the poor signal-to-noise ratio that characterise this kind of electroencephalographic (EEG) signal makes this task more difficult to perform.
The rest of this paper is structured as follows: *Material and Methods* provides a description of the dataset used and of the strategy defined. *Results* shows the outcomes both in terms of predictor's performance and physiological components detection. *Discussion* contains a comparison of the results obtained with the standard strategy and the one proposed in this work. Finally, *Conclusion* contains a brief summary of this work.

MATERIALS AND METHODS

The analysed dataset is the one proposed in [5] that can be downloaded from the BNCI Horizon 2020 database [6] (Dataset 8: P300 speller with ALS patients (008-2014)). The dataset consists of eight patients affected by ALS, and each patient was shown a 6 by 6 matrix containing alphanumeric characters. The user's task was to focus attention on characters in a word that was prescribed by the investigator (i.e., one character at a time). All rows and columns of this matrix were successively and randomly intensified at a rate of 4 Hz. Two out of 12 intensifications contained the desired character (i.e., one particular row and one particular column). Scalp EEG signals were recorded from eight channels according to 10-10 standard (Fz, Cz, Pz, Oz, P3, P4, PO7 and PO8). The EEG signal was digitized at 256 Hz and band-pass filtered between 0.1 and 30 Hz. Participants were required to copy-spell seven predefined words of five characters each (runs), by controlling a P300 matrix speller. Rows and columns on the interface were pseudo-randomly intensified for 125ms, with an inter stimulus interval (ISI) of 125 ms, yielding a 250 ms lag between the appearance of two stimuli (stimulus onset asynchrony, SOA). For each character selection (trial) all rows and columns were intensified 10 times (stimuli repetitions) thus each single item on the interface was intensified 20 times and the total number of flashes was 120. For each channel, 240 samples after stimuli onsets were selected for the analyses. The dataset consisted hence of 4200 instances. Drawing inspiration from [5], we split the dataset by using the first three words as training, and the last four as testing set. We considered four different versions of the dataset:

Single Trial we considered the original dataset, that is 4200 instances with 1921 attributes (240 samples $\times$ 8 channels plus the attribute that represent the row or the column intensified for each trial);

Decimated EEG data were then resampled in the time domain by replacing each sequence of 12 samples with their mean value, yielding 17×8 samples per epoch (eight being the number of channels), which

 DOI: 10.3217/978-3-85125-533-1-51

were concatenated in a feature vector of size 137

Decimated 5-averaged EEG data were resampled in the time domain as in the previous dataset version, but also in the instances domain, so that five consecutive instances of the same stimulation class were replaced by one instance of their average; this dataset version was then formed by 840 instances and 137 attributes;

Decimated 10-averaged This version is similar to the previous one, except that averages in the instances domain were computed every 10 consecutive instances of the same stimulation class; this dataset version was then formed by 420 instances and 137 attributes;

The three decimated datasets were obtained by using standard techniques for increasing the signal to noise ratio and hence should represent an improvement over the Single Trial dataset. However, it should be noted that, for online applications, the time necessary to perform a classification increases proportionally to the number of the averaged instances. Following the results described in [1], we use the filter ReliefF for feature selection, see [3]. ReliefF is a robust feature selection filter that can deal with incomplete and noisy data. This method randomly selects an instance R_i, then searches for k of its nearest neighbours from the same class called nearest hits H_j, and also k nearest neighbours from each of the different classes, called nearest misses $M_j(C)$. It updates the quality estimation $W[A]$ for all attributes A according to their values for R_i, hits H_j and misses M_j. Due to the noise of the data we have decided to weight nearest neighbours by their distance.

We tested different SVM based classifiers and the most efficient one was LibLinear described in [2], since it resulted the best on all the datasets, and it is also fast in terms of time for building the model. Liblinear returns an hyperplane $w^T x + b$, that discriminates among the two classes. It is important to stress that we did not use the standard classification function of LibLinear (that is $y(x) = sign(w^T x+b)$), but we exploited the information on the protocol that there is exactly one target element every six instances. Therefore, we assigned the target class to the maximum over the six flashes of $w^T x + b$ both for rows and columns. Whenever this assignment does not correspond to the real target, it results in both a false positive and a false negative.

We compare our results with the standard strategy in BCI (used also in [5]), that is SWLDA. SWLDA uses a stepwise method to perform a multilinear regression of the response values. We used the Matlab implementation with its default setting and with the decision function exploiting the knowledge on the protocol. Therefore, also for this method, we assigned the target class to the maximum over the six flashes of $w^T x+b$ both for rows and columns.

RESULTS

The following tables show, for each patient and for each kind of dataset, the accuracy, the Cohen's $Kappa$, and the true positives rate on the test set (that we recall is the last $4/7$ of the whole dataset), obtained both with SWLDA (default setting, and with the decision function described above) and with our strategy, that is LibLinear combined with the feature selection given by ReliefF.

A01	SWLDA			OUR		
dataset	acc	k	TP	acc	k	TP
orig	0.8145	0.3323	0.4436	0.811	0.320	0.434
dec	0.854	0.475	0.563	0.854	0.475	0.563
5-avg	0.925	0.730	0.775	0.917	0.700	0.750
10-avg	0.958	0.850	0.875	0.958	0.850	0.875

Table 1: Results for Patient A01 on all the datasets

A02	SWLDA			OUR		
dataset	acc	k	TP	acc	k	TP
orig	0.8187	0.3474	0.4561	0.815	0.335	0.446
dec	0.828	0.382	0.485	0.837	0.412	0.510
5-avg	0.938	0.775	0.813	0.933	0.760	0.800
10-avg	0.967	0.880	0.900	0.967	0.880	0.900

Table 2: Results for Patient A02 on all the datasets

A03	SWLDA			OUR		
dataset	acc	k	TP	acc	k	TP
orig	0.8371	0.4135	0.5113	0.836	0.411	0.509
dec	0.869	0.529	0.608	0.873	0.544	0.620
5-avg	0.917	0.700	0.750	0.942	0.790	0.825
10-avg	0.925	0.730	0.775	0.942	0.790	0.825

Table 3: Results for Patient A03 on all the datasets

A04	SWLDA			OUR		
dataset	acc	k	TP	acc	k	TP
orig	0.824	0.365	0.4712	0.835	0.408	0.506
dec	0.856	0.481	0.568	0.843	0.433	0.528
5-avg	0.854	0.475	0.563	0.867	0.520	0.600
10-avg	0.892	0.610	0.675	0.950	0.820	0.850

Table 4: Results for Patient A04 on all the datasets

A05	SWLDA			OUR		
dataset	acc	k	TP	acc	k	TP
orig	0.849	0.456	0.5464	0.832	0.395	0.496
dec	0.850	0.460	0.550	0.863	0.505	0.588
5-avg	0.933	0.760	0.800	0.942	0.790	0.825
10-avg	0.975	0.910	0.925	0.967	0.880	0.900

Table 5: Results for Patient A05 on all the datasets

A06	SWLDA			OUR		
dataset	acc	k	TP	acc	k	TP
orig	0.850	0.459	0.5489	0.835	0.405	0.504
dec	0.881	0.571	0.643	0.878	0.562	0.635
5-avg	0.958	0.850	0.875	0.946	0.805	0.838
10-avg	0.967	0.880	0.900	0.975	0.910	0.925

Table 6: Results for Patient A06 on all the datasets

A07	SWLDA			OUR		
dataset	acc	k	TP	acc	k	TP
orig	0.860	0.495	0.5789	0.830	0.389	0.491
dec	0.873	0.541	0.618	0.863	0.508	0.590
5-avg	0.933	0.760	0.800	0.963	0.865	0.888
10-avg	0.967	0.880	0.900	0.967	0.880	0.900

Table 7: Results for Patient A07 on all the datasets

 DOI: 10.3217/978-3-85125-533-1-51

A08	SWLDA			OUR		
dataset	acc	k	TP	acc	k	TP
orig	0.911	0.678	0.7318	0.906	0.663	0.719
dec	0.933	0.757	0.798	0.928	0.7396	0.783
5-avg	0.983	0.940	0.950	0.979	0.925	0.938
10-avg	0.983	0.940	0.950	0.983	0.940	0.950

Table 8: Results for Patient A08 on all the datasets

In order to show the physiological significance of the feature selection that we use, in Figure 1 (patient A02) and Figure 2 (patient A07) are shown the target signals (orange) vs the non target signal (blue) on all the electrodes for patient A02 and patient A07. N200 VEP component can be observed in Fig. 1 on Oz, Po7 and Po8 whereas the P300 component can be observed on both patients in the 400-500 ms range and on Fz and Cz electrodes. Despite the fact that these averages were obtained from a relevant number of trials (300 targets vs 1500 non targets), they appear to be quite different from those known from the literature and from healthy subjects for two main reasons: first of all the EEG signals from these patients have a lower signal to noise ratio, and secondly the responses overlap after each stimulation as they are elicited (every 250ms) before the physiological response in extinguished (usually after no less than 800ms), thus causing some interference.

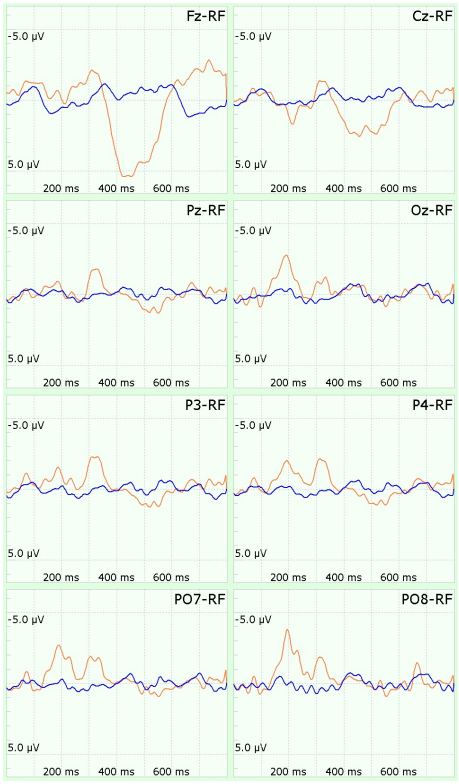

Figure 1: Target (Orange) vs Non-Target (Blue) for Patient A02.

We have selected these two patients because they represent opposite classifiers performance: A02 is among the worst while A07 is one of the best, as shown by the results in Tables 2 and 7.

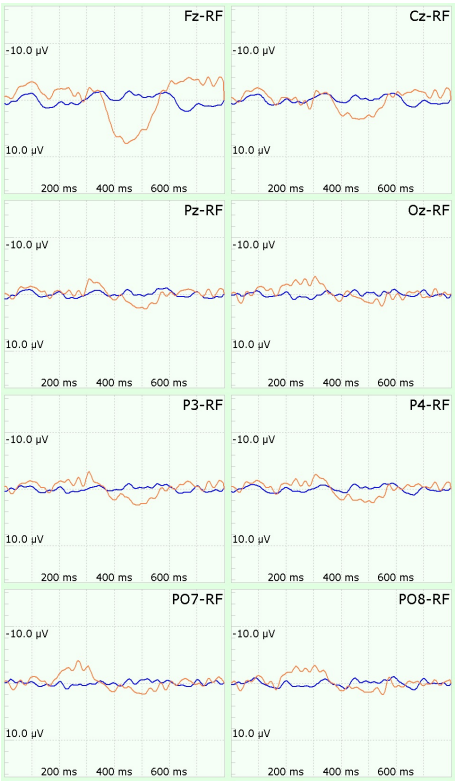

Figure 2: Target (Orange) vs Non-Target (Blue) for Patient A07.

In order to investigate whether the features selected by ReliefF are coherent with the physiological signals represented in Figures 1 and 2, we compared topographic maps (Fig.3, A02; Fig. 4, A07) computed at certain time interval, and according to three different methods:

- on the left we draw the weights chosen by SWLDA for each feature;
- in the center we have the ERPs
- on the right the score computed by ReliefF

In all maps plotted values are relative to 7 distinct time intervals, and averaged across 12 consecutive time samples. SWLDA and ReliefF weights were computed after averaging 10 consecutive instances of the same stimulation class and using 3/7 of the whole dataset. So, 180 instances with 1921 attributes (1920 from the signals and one from the label) were used for training them. Note that each of the 1920 obtained weights are bound to a feature and then to a sample and an electrode. Weights are then reduced after averaging 12 consecutive of them, in order to further increase the signal to noise ratio. Note that this approach preserve time and space information so that each weight is still bound to an electrode and a time (in this case an interval of 12 samples).
From all the possible maps, one for each time interval, a subset of 7 (the most significant ones) is represented in Figures 3 and 4.

 DOI: 10.3217/978-3-85125-533-1-51

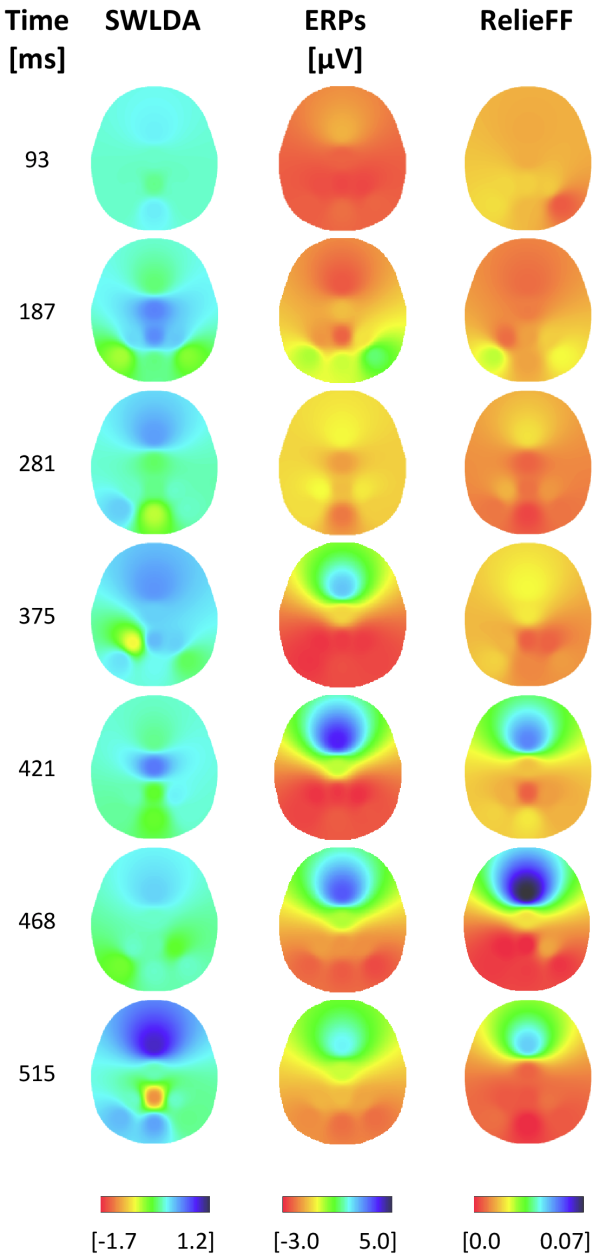

Figure 3: Topographic maps relative to patient A02 computed according to 3 different methods and 7 different time intervals.

From both Figures 3 and 4 it can be seen that maps computed from features selected from ReliefF are very similar to those obtained from the ERPs whereas those computed from SWLDA weights are quite different. This clearly suggests that features selected with our approach are more related to physiological signals than those selected from SWLDA.

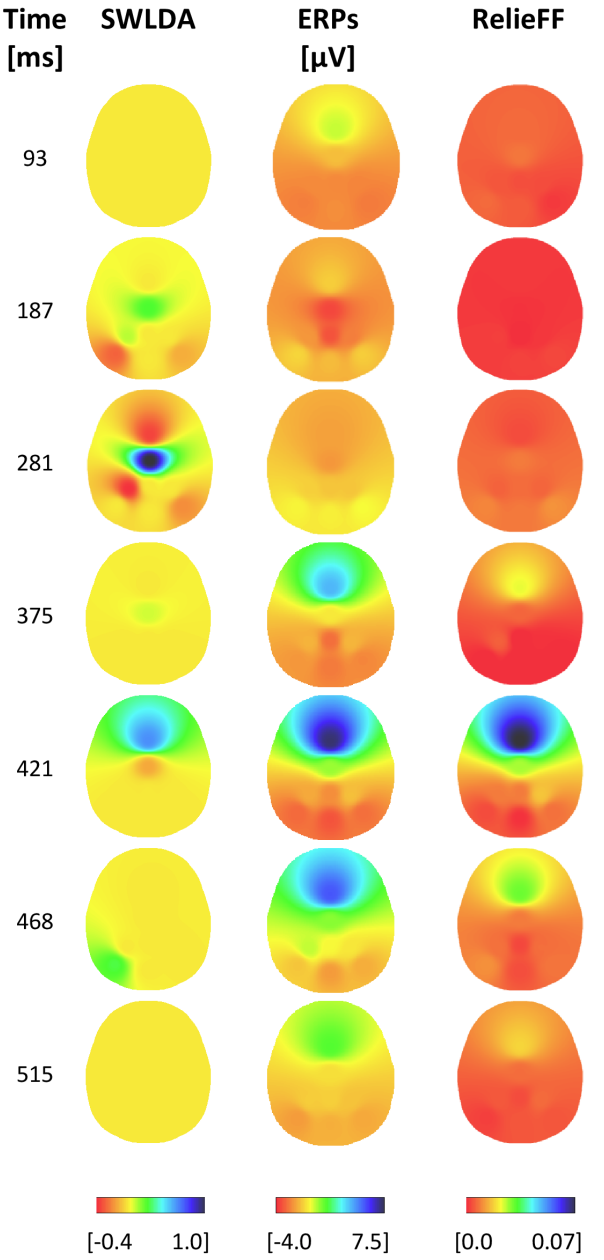

Figure 4: Topographic maps relative patient A07 computed according to 3 different methods and 7 different time intervals.

DISCUSSION

The reported results show how our approach based on ReliefF and LibLinear represents a valid alternative to the widely used SWLDA, since it provides results comparable to SWLDA in terms of accuracy and Cohen?s Kappa, but selecting features that are physiologically relevant while the features selected by SWLDA show scarce correlation with the ERPs. This kind of information might furnish relevant insights to identify which brain areas and when are involved during certain cerebral activities, thus improving the comprehension of brain functioning and furnishing a valuable instrument for supporting scientific hypotheses or diagnoses. It is also clear how the prepro-

 DOI: 10.3217/978-3-85125-533-1-51

cessing of the data (especially the averaging of the signal) is effective at improving the performance of the classifier, as shown by the increasing accuracy (and Cohen's Kappa) obtained over the four datasets.

CONCLUSION

The detection of brain state changes translates into classification problems with a huge number of features that make difficult to distinguish those relevant ones for diagnostic use. Therefore, distinguishing significant characteristics not only would improve the predictor's performance, but would also provide a better understanding of the underlying cerebral process that generated the data. From classification point of view, the obtained results show how our approach represents a valid alternative to the standard SWLDA approach. More significantly, as for the feature selection, the performance obtained with our strategy outperforms SWLDA, since it turns out that also in the case of ALS patients, this feature selection filter is particularly robust, and returns a subset of selected feature that is physiologically compatible. Figures 3 and 4 show how ReliefF was able to detect physiological components elicited during the protocol either in space (e.g. Cz, Pz, ...) or in latency (e.g. P300).

REFERENCES

[1]Bianchi L., Liti C. and Piccialli V.. Features Reduction for P300 Spellers. In Proceedings of the Sixth International Brain-Computer In- terface Meeting, May 30-June 3 2016, Asilomar Conference Center, Pacific Grove, California, USA.

[2] Fan R.-E. , Chang K.-W., Hsieh C.-J., Wang X.-R., and Lin C.-J.. LIBLINEAR: A library for large linear classification. Journal of Machine Learning Research 2008;9:1871-1874.

[3] Kononenko, I. Estimating attributes: analysis and extensions of Relief. In De Raedt, L. and Bergadano, F., editors, Machine Learning: ECML-94, 1994, pages 171-s182. Springer Verlag.

[4] Lotte F., Congedo M., Lècuyer A., Lamarche F. and Arnaldi B. A review of classification algorithms for EEG-based brain?computer interfaces Journal of Neural Engineering, 2007; 4(2).

[5] Riccio A, Simione L, Schettini F, Pizzimenti A, Inghilleri M, Olivetti Belardinelli M, Mattia D, Cincotti F. Attention and P300-based BCI performance in people with amyotrophic lateral sclerosis. Frontiers in Human Neuroscience 2013;7:732.

[6] BNCI Horizon 2020. The Future of Brain/Neural Computer Interaction: Horizon 2020 `http://bnci-horizon-2020.eu/database/data-sets`

[7] L. Bianchi, at all. Which physiological components are more suitable for visual ERP based brain computer interface? A preliminary MEG/EEG study. Brain Topography, 23:180-185, 2010

[8] DECODER Project, European ICT program FP7-ICT-2009-04; Grant Agreement 247919