

A classification algorithm for electroencephalography signals by self-induced emotional stimuli

Daniela Iacoviello, Andrea Petracca, *Student Member, IEEE*, Matteo Spezialetti, *Student Member, IEEE*, and Giuseppe Placidi, *Member, IEEE*

Abstract - The aim of this paper is to propose a real time classification algorithm for the low-amplitude electroencephalography (EEG) signals, such as those produced by remembering an unpleasant odor, to drive a Brain Computer Interface (BCI). The peculiarity of these EEG signals is that they require ad-hoc signals pre-processing by wavelet decomposition, and the definition of a set of features able to characterize the signals and to discriminate among different conditions. The proposed method is completely parameterized, aiming at a multi-class classification and it might be considered in the framework of machine learning. It is a two stages algorithm. The first stage is off-line and it is devoted to the determination of a suitable set of features and to the training of a classifier. The second stage, the real-time one, is to test the proposed method on new data. In order to avoid redundancy in the set of features, the Principal Components Analysis is adapted to the specific EEG signal characteristics and it is applied; the classification is performed through the Support Vector Machine. Experimental tests on 10 subjects, demonstrating the good performance of the algorithm in terms both of accuracy and efficiency, are also reported and discussed.

Index Terms— brain computer interface classification algorithm, electroencephalography, low-amplitude EEG signals, principal components analysis, support vector machine.

I. INTRODUCTION

SIGNALS classification is a main issue in many different applications and, among them, in BCI. A BCI is a computer based communication system that analyses signals generated by voluntary neural activity of the central nervous system. The subject, thinking about an intention, generates voluntary brain signals which will be translated into commands for an output device. In this way it is available a new channel of output for the brain [1-4]. The neural activity useful for BCI can be measured through EEG; it measures the electrical activity produced by neurons either

D.Iacoviello is with the Department of Computer, Control and Management Engineering Antonio Ruberti, Sapienza University of Rome, Via Ariosto 25, 00185, Rome, Italy (e-mail: iacoviello@dis.uniroma1.it)

A.Petracca is with the Department of Life, Health and Environmental Sciences, University of L'Aquila, Via Vetoio, Coppito 2, 67100, L'Aquila, Italy (e-mail: petracca.andrea84@gmail.com)

M.Spezialetti is with the Department of Life, Health and Environmental Sciences, University of L'Aquila, Via Vetoio, Coppito 2, 67100, L'Aquila, Italy (e-mail: spezialetti.matteo@gmail.com)

G.Placidi is with the Department of Life, Health and Environmental Sciences, University of L'Aquila, Via Vetoio, Coppito 2, 67100, L'Aquila, Italy (e-mail: giuseppe.placidi@univaq.it)

by electrodes placed on the head [5] or through microelectrodes positioned inside the skull [6, 7].

Though the presence of the skull or other tissues is sources of noise and blurring, EEG with external electrodes is characterized by high temporal resolution, it is safe, not expensive, portable, not invasive [8] and for this reason it is often used for BCI.

BCI is currently applied in different fields, ranging from video games to army and as a communication support for people with disabilities. In fact, in the last ten years BCI has revealed its potentialities for severely disabled and locked-in individuals who had very few possibilities to interact with the ambient and with other subjects [9-11].

Generally a BCI for disabled subjects is based on signals produced by sensory-motor rhythm amplitudes [12-14] or induced by external stimuli (visual [15], auditory [16-18], or tactile [19,20]). Recently, attention has been addressed to the recognition of the emotional responses externally elicited by multimedia presentation; in particular, what was investigated in [21] was the association among music-induced emotions and EEG.

There are patients for whom the above mentioned paradigms cannot be used and who need other forms of stimulations to be explored. Recently, in [22] the Authors proposed a new promising paradigm that used the signals generated by the disgust produced by remembering an unpleasant odor. The task in [22] was alternative to the classical stimuli because:

- 1) it was a self-induced stimulus voluntarily generated by the subject and not externally elicited;
- 2) it was free from disturbs that could be caused by any external stimulation, being a self-induced stimulation.

However, the signals produced by a self-induced task have lower amplitude compared to those produced by external stimulations because:

- 1) the analyzed subject could lose concentration during the task;
- 2) the signals caused by remembering a past situation are always lower than those produced while living the situation directly [23].

Moreover, the signals produced by a self-induced task are not well time-resolved. This makes the classification of these signals very hard.

In literature the analysis of EEG signal is a well-established

field of research and the different techniques proposed mainly depend on the kind of signal to be analyzed and the information to be retrieved. For decades [24] the EEG signals have been interesting for their information content, stimulating the adaptation of signal analysis techniques to these particular physiological data [25,26].

Three main aspects must be faced when dealing with signal classification: the transformation to be applied to the signal, the choice of the features to be evaluated over the transformed data and the classification strategy.

The transformation of the signal is necessary to synthesize and to group the useful information content of the signal. The feature selection and/or extraction procedures are referred to the identification of some signal characteristics, the features, necessary for the signal classification. The classification strategy uses the identified features of a given signal to establish its class of pertinence.

Regarding the signal transformation, a time-frequency representation of the EEG signal, concerning two mental states, was obtained by a Discrete Wavelet Transform (DWT) in [27]; in particular a Daubechey 4 wavelet (DWT-Db4) was applied for multiresolution analysis considering only levels 4 and 5. The DWT till level 5 was used in [28], where it was also considered the problem of features' identification by calculating the entropy, the energy and the standard deviation of the DWT transformed signal. In [29] to analyse EEG signals, first and second order spatial and temporal features, based on a bilinear model, were proposed. Among the statistical features, the mean of the absolute values of the coefficients of the Wavelets Discrete Transform in each sub-band, the average power, the standard deviation and the ratio of the absolute mean values of adjacent sub-bands were mostly used [30].

The problem of reducing data dimension and facilitating features identification may be addressed by considering the Principal Components Analysis (PCA), the Independent Component Analysis (ICA) and the Linear Discriminant Analysis (LDA) [30]. In [31] an efficient algorithm based on PCA was proposed and it inspired part of the procedure adopted in this paper.

The classification strategy is used to assign a given signal to a specific class. An efficient method for the classification of the EEG signals is the Support Vector Machines (SVM) [32,33] that take as input the features in order to classify the given signals.

Generally the SVM are applied as a binary classifier, but they can be extended for multi-class classifiers as in [34] and in [35] where multilayer perceptron neural network, probabilistic neural network and multiclass support vector machines were considered for EEG signals.

The paradigm proposed in [22] allowed the construction of a binary BCI in the class of affective BCI, based on the measurement of one of the emotions, the disgust, produced by remembering an unpleasant odor. It was presented a specific classification method, based on the short time

Fourier Transform. The method was simple and effective in classifying binary data ("disgust" compared to "relax"), but it was based on a simplifying assumption: only the channels located in the right hemisphere of the brain were considered and averaged together in order to improve the signal quality. The aim of this paper is to propose an alternative classification strategy for the paradigm proposed in [22]. There are many reasons to explore a different strategy:

- 1) to improve the classification accuracy and to obtain a robust classification;
- 2) to generalize the strategy by considering all the recorded signals (this is particularly important to increase the cardinality of the alphabet, if new emotions would be added, and to study the relationship between signals and features corresponding to different channels);
- 3) to improve spatial resolution;
- 4) to propose a minimal subset of features and channels in order to design a minimal helmet configuration for the proposed task.

In what follows, the EEG signal classification problem is addressed by considering the peculiarity of low-amplitude signals: the particular aspects of signal processing, analysis and classification were pointed out exploring and adapting strategies like the Wavelet Transform, the PCA and the SVM. The method was applied to the signals generated by the disgust produced by remembering an unpleasant odor. The procedure was a 2-steps one. First the signals (belonging to a training set) were suitably filtered, a set of features was extracted, the more informative ones were selected, and an ad hoc classifier was tuned by applying the SVM method. The first step was concluded by performing the validation of the optimized classifier on a specific dataset. The second step, an on line one, was the application of the obtained classifier to new, real-time, data (in our tests it was applied to the test set). The classification method proposed therein was tuned on low-amplitude EEG signals, as those generated by a self-induced stimulus.

The paper is organized as follows. In Section 2 the acquisition system, the signal processing strategy and the classification method are described; the numerical results are presented in Section 3; conclusions and future work are outlined in Section 4.

II. MATERIALS AND METHODS

The first step of the experimental process was to collect the EEG signals produced by the self-induced stimulation of the disgust produced by remembering an unpleasant odor.

Being the resulting signals produced by a self-induced task, they were weaker than those produced by external stimulation and the first operation that had to be applied was a pre-filtering one in order to improve the signal quality by preserving only the useful information. To this purpose a Wavelet Transform was applied with the selection of the characterizing signal structures related to the proposed task.

After preprocessing, a set of features was computed on the resulting signals and, to avoid redundancy, a PCA features' selection was applied. On the reduced set of features the Support Vector Machines technique was applied to obtain the classification. Figure 1 shows a block diagram synthesizing the overall procedure that will be described in the next subsections.

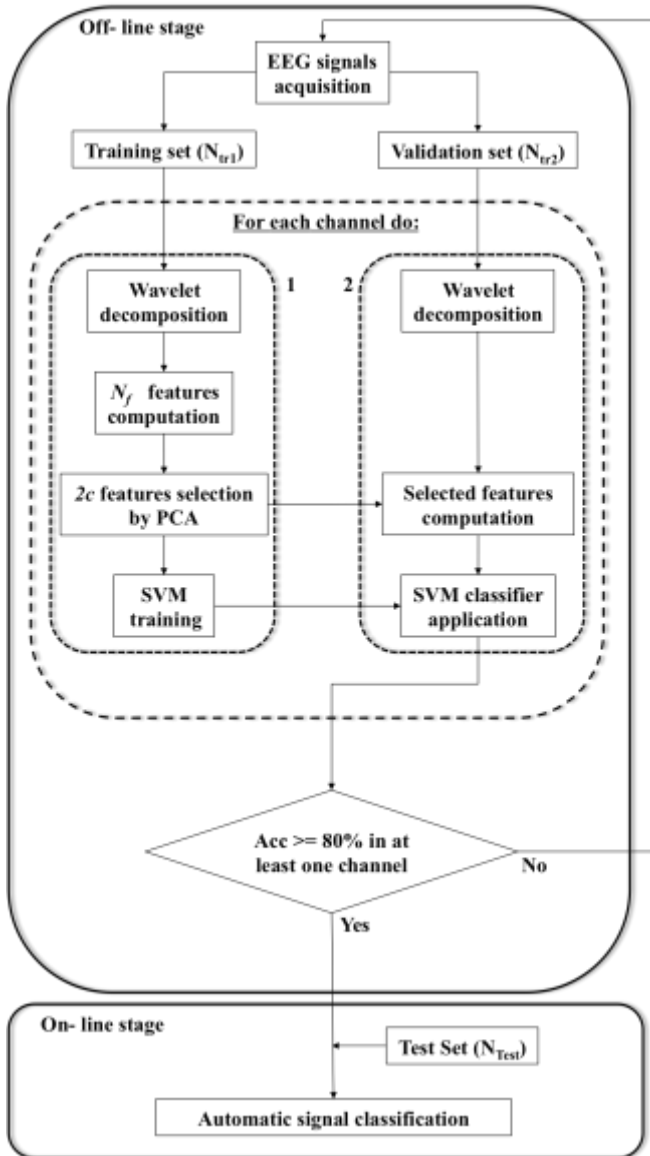


Fig.1. Block diagram of the classification procedure. Note that steps #1 and #2 were executed in sequence and not in parallel. Moreover, note that the on-line stage could use the test set (whose labels were known) or a new unknown (on-line) signal.

2.1 The acquisition procedure

The experimental protocol consisted in a combination of two different tasks: one produced by self-induction of the disgust derived by remembering an unpleasant odor (identified as task #1, or concentration); the other produced by relaxing or any other brain activity except task #1 (identified as task #2, or relax).

In order to collect data, each examined subject was sat in a comfortable armchair with the arms lying relaxed, in a quiet

and lit room. The experiment consisted in showing a sequence of symbols “↓” or “+”, each presented for 3.6 seconds on a screen. All the examined subjects were previously informed that when they saw “↓” they had to concentrate on the unpleasant odor producing the disgust; whereas when they saw “+” they had to relax. The stimulus or the relax tasks had to be maintained until the symbol changed. During this time, the EEG signals were recorded. The order of presentation was random but the number of symbols “↓” was equal to the number of symbols “+”. Anonymous symbols were used only to synchronize the tasks, but not to elicit any mental state. The subjects received instructions through the projection of a short video, on the modality and duration of the experiment. In addition, at the end of the experiments, the subjects were asked to yield a feedback, by filling a form, about their difficulties in concentrating during the session. For more details please refer to [22].

Two sequences, each containing 100 trials (50 trials for each class, where a class was associate to “↓” and another to “+”) were recorded for each subject. In each sequence the trials of the two classes were mixed in a random order and executed without interruptions for 6 minutes.

The system used to record the EEG was EnobioNE® (<http://www.neuroelectrics.com/enobio>), a precise and robust 8 channels (two more channels were used one as reference and another for ground) wireless EEG equipment that uses a neoprene cap to fix the channels in the desired brain locations. Thanks to the supporting software, the channels can be dynamically associated to variable locations in the international 10-20 positioning system [36]. The positions we used were: P4, C4, T8, P8, P3, C3, T7 and P7 (Figure 2).

Dry copper electrodes (coated by a silver layer) fastened to the cap helped to ensure the contact with the subject’s scalp. The electric conduction was ensured by contact: the electrodes terminated with a circular disposition of contact tips to pass through the hair. For reference and ground channels, located just behind the right ear in an area where the skin was free, the fastening strategy was ensured by adhesive, disposable, flat gel-containing connectors to whom the electrodes were plugged in. Main characteristics of EnobioNE® were: amplitude resolution of 24 bits (0.05 uV); sampling rate of 500 Hz; low-pass filtering between 1 Hz and 46 Hz. Data were analyzed and classified by a specific algorithm (described herein) implemented in Matlab®. The signals were directly captured through the BCI2000 software that contains the utilities to acquire and to process them.

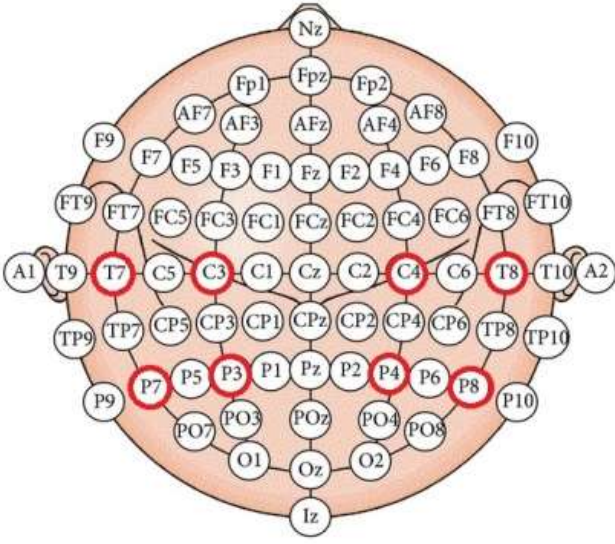


Fig.2. International 10-20 positioning system. The used channels are in bold.

2.2 Signal processing: wavelet decomposition and features' selection

To test the rule and the importance of each signal corresponding to a specific channel, the analysis was performed on the data of all the channels separately, even if the verification of the efficacy of the classification was performed in all the channels (see Fig.1 and below).

The proposed classification method was divided in two steps. First, a set of collected signals was used for calibration. This off-line stage allowed the identification of the most predominant features referring to the two distinct conditions, the one due 'activation' (task #1) and the other due to 'relax' (task #2), considering that the class of a given signal was known in advance. The second step was the on-line classification of a signal whose pertinence to the task #1 or #2 was unknown in advance.

A set of N trials was considered for all the analyzed subjects. Each trial consisted of a set of signals collected contemporarily by all the considered channels for a given time interval. At this moment the N trials were divided into two sets, N_{tr} used for the training and N_{test} used for the test. The only attention paid in this partition was using the same number of trials of the two different classes within each group; this shrewdness was motivated by the aim not to polarize the training in one sense or in another.

A suitable wavelet transformation was applied on each trial [37]. For EEG data different wavelet functions could be considered. It has been demonstrated that orthogonal wavelet transformations are particularly useful in analyzing brain signals for their capability in separating different frequency bands [38], and for their de-noising effects, as pointed out in [39]. In the present case the Meyer wavelet was chosen: it is an orthogonal symmetric wavelet, infinitely differentiable and with infinite support.

The wavelet transformation represents the similarity between the analyzed signal and the wavelet function ϕ :

$$C_d(k2^{-j}, 2^{-j}) = 2^{j/2} \sum_n s(n) \phi(2^j k - n)$$

It decomposes the signal in the approximation and in the detail coefficients.

The choice of the decomposition level depends on the band of frequencies of interest. More precisely, assuming a sampling frequency of 500 Hz (the sampling frequency of the EEG system used for the experiments), and therefore the presence of frequencies in the signal up to 250 Hz, at the ℓ -level of a dyadic decomposition the approximation CA_ℓ

contains frequencies in the band $\left[0, \frac{250}{2^\ell}\right]$ Hz, whereas the

detail CD_ℓ yields information in the band $\left(\frac{250}{2^\ell}, \frac{250}{2^{\ell-1}}\right]$ Hz.

The reconstructed signal s_ℓ^{rec} may be represented as follows:

$$s_\ell^{rec} = CA_\ell + \sum_{i=1}^{\ell} CD_i$$

The dyadic wavelet decomposition was used to select and retain only specific frequency bands of the original signal before skipping to the following step of the classification method.

After wavelet filtering, features calculation and selection were performed.

To this aim, by considering a generic subject and a generic channel among those measured, let us denote with $s_i, i = 1, \dots, N$ the i -th trial.

The aim was to classify each trial as referring to the task #1 or to the task #2. As stated above, each trial was measured during a defined time interval. In theory, the time interval could be chosen short enough (as will be clarified below, the information of interest could be contained in a short time-window of signal of the duration of about 0.5 sec), having also ensured that the subject maintained his concentration during the whole time interval. However, if the trial duration was too short, the subject could not have enough time to concentrate on the task and/or the classification algorithm could not have enough information to come to the right decision, due to the low amplitude of the signals (being weak signals, it was necessary to improve their sensitivity also by increasing the interval of measurement). The adopted strategy was using an interval of measurement of 3.6 seconds for each trial in order to break the signal into pieces and average them together (see below). At the same time, we implemented a strategy for discarding pieces of the signals where useful information

were absent, for example, because of the subject's loss of concentration.

To this aim, each trial s_i , consisting of a number ν of samples corresponding to a record of N_s seconds, was divided into q sub-trials $s_{ih}, h=1, \dots, q$ of N_q seconds, corresponding to $\bar{\nu}$ samples, with an overlapping factor of p elements (Fig.3).

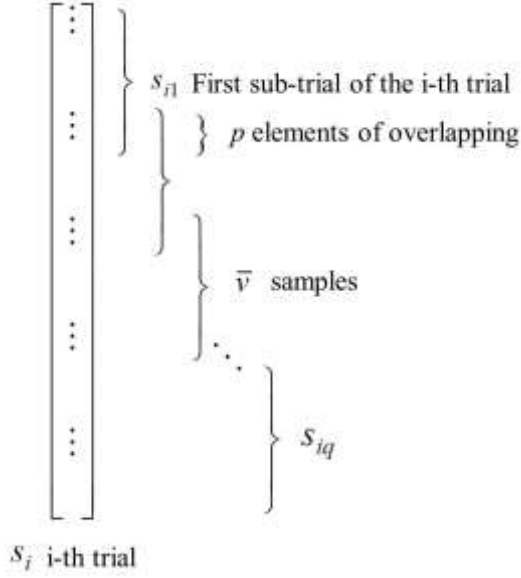


Fig.3 Trial division in q sub-trials. p represents the number of overlapping points between consecutive sub-trials.

Regarding the trials in the N_{tr} set, a set of N_f features was computed in each sub-trial.

Given a subject, a trial and a channel, the matrix of features F was defined as follows:

$$F = \{f_{uv}\}, \quad u=1, \dots, q \quad v=1, \dots, N_f$$

where the element f_{uv} , $u=1, \dots, q$, $v=1, \dots, N_f$, was the v -th feature f_v of the u -th sub-trial. Therefore the matrix of features for each trial had dimension $q \times N_f$.

To avoid noisy sub-trials of signals, among the q sub-trials it would be recommendable to recognize the values of the features that could be assumed as ‘‘singular’’ and discard the sub-trials generating these ‘‘noisy’’ feature values. To this aim in each trial the sum of the absolute values of the differences among one feature and the others, of the same type, was computed as follows:

$$d_{hv} = \sum_{j=1}^q |f_{hv} - f_{jv}|, \quad h=1, \dots, q$$

thus obtaining the matrix $D = \{d_{hv}\}_{q \times N_f}$ of differences.

For each column of this matrix D only the r smaller values were preserved and the corresponding features f_{hv} were collected in the matrix $F_{reduced}$ of dimension $r \times N_f$. This

means that in the trial, for a given type of feature, among the values of the feature evaluated on the q sub-trials, $q-r$ values with significant difference among each other were disregarded as effect of ‘‘noise’’.

Finally, for each $v=1, \dots, N_f$, the mean value of these r features was evaluated obtaining a unique value for each of the N_f features (this operation ‘‘integrates’’ the information contained in the remaining useful sub-trials). Therefore, for each trial i a vector of the N_f features $(\bar{f}_{i1} \bar{f}_{i2} \dots \bar{f}_{iN_f})$ was available. The vector of features for each trial had dimension $1 \times N_f$.

This procedure was applied to all the trials; to obtain comparable data, within each trial the elements of the features vector were normalized. More precisely, for each $j=1, \dots, N_f$ the new elements of the vector of features of the i -th trial were:

$$f_{ij}^{norm} = \frac{\bar{f}_{ij} - \min_{1 \leq j \leq N_f}(\bar{f}_{ij})}{\max_{1 \leq j \leq N_f}(\bar{f}_{ij}) - \min_{1 \leq j \leq N_f}(\bar{f}_{ij})}$$

Therefore for each kind of feature, the data that we classified were the N_f normalized features, one for each

trial, of the set N_{tr} . The matrix $\bar{F} = \{f_{ij}^{norm}\}_{N_{tr} \times N_f}$ of the

normalized features had dimension $N_{tr} \times N_f$.

The procedure of features selection, using the PCA, was inspired by [31]. First, the PCA evaluated the correlation matrix \bar{F}_c of \bar{F} ; the eigenvalues $\lambda_i, i=1, 2, \dots, N_f$ of \bar{F}_c along with the principal components coefficients W_{ji} , $j=1, 2, \dots, N_f$ (i.e. the components of the corresponding eigenvectors V_i) were computed. The larger c eigenvalues and the corresponding eigenvectors were retained. The other neglected eigenvectors were considered as noise or redundant information.

By considering only the first C principal components, the percentage of the retrieved information was given by:

$$P_c = \frac{\sum_{i=1}^c \lambda_i}{\sum_{i=1}^{N_f} \lambda_i} \cdot 100. \quad (1)$$

For each of the considered principal components, the two most significant features were considered, i.e. the features with higher weights W_{ji} , thus providing robust information.

A rationale in the choice of C must be $c < N_f / 2$. If, among the features with the higher weight, there was a repetition, it was considered the next one.

A set of $2c$ features for each of the considered channels (P4, C4, T8, P8, P3, C3, T7 and P7) was selected as output of the PCA features' selection: the selected features could be different for different channels.

Once for each channel the $2c$ features were selected and validated (see next Subsection), they were evaluated also on the set N_{test} , by repeating the procedures for sub-trials elimination and features integration used for the N_{tr} trials.

2.3 Training, validation and classification

Once the number of features was reduced, the SVM was applied to determine the optimal hyperplane in order to split the data into two groups and allocate each trial to task #1 (activation stage) or to task #2 (relaxing stage). The optimal hyperplane was obtained as a tradeoff between the requirement of maximizing the Euclidean distance between the closest points and the requirement of minimizing the error on misclassified points [40, 41]. A penalty parameter $H > 0$ on the error term was introduced.

The optimal separating hyperplane was obtained by solving the following Quadratic Programming Problem:

$$\min_{w,b,\xi} \frac{1}{2} w^T w + H \sum_i \xi_i \quad (2)$$

with the constraint:

$$y_i (w^T x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad (3)$$

where W was the vector of the points perpendicular to the separating hyperplane, x_i was the i -th $1 \times 2c$ normalized features' vector and $y_j \in \{1, -1\}$ was the label that expressed the class to which x_i belonged. Then, $y_j = -1$ if x_j corresponded to a stimulus situation (task #1) and $y_j = 1$ if x_j represented a relax one (task #2).

In order to make the elements x_i of the two classes linearly separable, the data were mapped into a richer space (the "feature space") by using a function ϕ , and the separating hyperplane was determined in that space. Different choices for the function ϕ were possible by considering a kernel function K . The radial basis function was assumed in this paper:

$$K(x_i, x_j) = \phi(x_i)^T \phi(x_j) = \frac{\|x_i - x_j\|_2^2}{2\alpha^2} \quad (4)$$

where $\|\cdot\|_2$ represented the L_2 -norm. The two parameters to evaluate, H and α , could be determined by using the 10-

fold cross validation, [41]. The classification was performed through the SVM algorithm LIBSVM 3.18, [42].

More precisely, the data of the training set N_{tr} were split into two sets, N_{tr1} and N_{tr2} . The first one was used to train the SVM, searching the parameters H and α on a grid and yielding the accuracy values of the training (generally, after this step, the test could be performed). In the present application, due to the specific nature of the data to be classified (features from signals with low-amplitude), two classifiers were validated on the set N_{tr2} . The used classifiers were those yielding the larger accuracy to determine the best couple (H^*, α^*) over this new set of data (N_{tr2} was not used in the training phase where just the set N_{tr1} was involved). This procedure was repeated for each channel. After the determination of the best classifiers on the set N_{tr2} , one for each channel, the list of classifiers was accepted if the percentage of success (the parameter Acc in Fig.1) was above 80% in at least one channel. Otherwise the set N_{tr} had to be integrated by collecting new data (Fig. 1). The reason behind this choice is that, having supposed that the activation task was effective, at least one area of the brain had to be active during the signals acquisition (in the same way, the non-activation of other areas of the brain justified the presence of classifiers with low accuracy in some channels). For this reason, if the accuracy was below 80% in all channels, for a binary classification, we considered the training set too poor to instruct correctly the machine (the signal-to-noise-ratio was low and more signals had to be used for training). In those channels where accuracy remained below 80%, because the activation was too weak and no significant differences occurred between the two tasks, an optimal classifier could not be found and the test was not performed. The obtained classifiers were then tested on the test set N_{test} , whose true labels were known, by evaluating the percentage of success, i.e. the ratio between correct classifications over the total number N_{test} . Note that the classification accuracy considered the percentage of data correctly classified with the optimal choice (H^*, α^*) and it represented a property of the classifier evaluated over the training set, whereas the percentage of success was evaluated over the test set N_{test} .

3 NUMERICAL RESULTS

The proposed algorithm was applied to a set of EEG signals collected from ten healthy subjects, M_i , $i=1,2,\dots,10$. Two sequences, each containing 100 trials (50 trials for each class, where a class was associate to "↓" and another to "+"), were recorded for each subject. The order of execution

of the two sequences was alternated (first sequence from the first subject; first sequence from the second subject; ... ; first sequence for the last subject; second sequence for the first subject; second sequence for the second subject; ... ; second sequence for the last subject) with 24 minutes of interruptions between subjects (for preparation of the following subject). In this way, each subject experienced a resting period of about 5 hours between the first and the second experiment.

For the same subject, data allowing to the two different tests were grouped in a single group of $N=200$ trials. Each trial consisted of $\nu=1800$ elements, corresponding to a record of $N_s=3.6$ sec; each trial was divided into $q=7$ sub-trials, corresponding to $\bar{\nu}=300$ elements including the overlapping segments. This number was chosen to point out, if present, significant features in the different parts of the signal. Consecutive sub-trials had an overlapping segment of $p=50$ samples, corresponding to 0.1sec: the overlapping segment helped to avoid the exclusion of useful information that might be present on the tails of the sub-trials.

As shown in Fig.1 and explained in Section 2, the algorithm was based on splitting the data into two main sets, the $N_{tr}=120$, used for the training phase, and $N_{test}=80$, used for the test. We wanted to be sure that the specific division of the data into these sets wouldn't influence the final result. For this reason, we considered different random choices assuming only the constraint of using the same number of trials in both classes (stimulus/not stimulus) and all the results proposed in this paper were obtained as the mean value of all the performed analyses.

In [22] the relevant information for the self-induced stimulus of disgust was in the interval 30-42 Hz and, to confirm the activation of the memory, also in the interval 8-12 Hz. The frequency band is relevant when a subject is concentrating to any task, and not specifically to the self-induced stimulus considered in this paper. Therefore, a level $\ell=3$ Meyer wavelet decomposition was assumed and the detail CD_3 was considered thus retrieving the information of interest of this specific stimulus into the interval 30-42 Hz.

From the detail signals CD_3 a set of features was chosen: the mean and the median values, f_1 and f_2 respectively, the mode f_3 (i.e. the most frequent value in a sub-trial), the largest and the smallest elements, f_4 and f_5 respectively, the range f_6 of the values and their standard deviation, the mean value f_8 and the median f_9 of the absolute value of the difference between the vector and its mean value, the sum f_{10} of all the elements, the norm f_{11} and the maximum value f_{12} . The considered features were

the most common in the relevant literature [43-45]. To determine the matrix $F_{reduced}$ of dimension $r \times N_f$ the value $r=4$ was assumed; therefore the 3 sub-trials, corresponding to about 1.5sec, with higher differences from the others were disregarded. This choice was rather conservative, but reasonable if we consider that a subject could experience difficulties to keep focused. In order to reduce the number of features useful for the classification, we applied PCA and assumed $c=2$. In fact, by applying Eq. (1) to the available data of the ten considered subjects, the choice $c=2$ represented a good compromise between the performance improvement and the reduced complexity. Thus, for each channel a set of $2c=4$ features was considered, as suggested by robustness arguments in Subsection 2.3.

The four selected features for each subject S and each channel Ch were reported in Table 1.

Table 1: Features selected through the PCA by considering the first two Principal Components. Note that the selected features were the same in the different random choices of the training sets, but changed for different channels of the same subject.

S	Ch P4	Ch C4	Ch T8	Ch P8	Ch P3	Ch C3	Ch T7	Ch P7
M ₁	$f_3; f_5$ $f_7; f_{11}$	$f_1; f_6$ $f_9; f_{12}$	$f_3; f_5$ $f_8; f_{10}$	$f_3; f_5$ $f_8; f_{10}$	$f_1; f_2$ $f_6; f_{12}$	$f_3; f_5$ $f_6; f_7$	$f_3; f_8$ $f_9; f_{12}$	$f_3; f_5$ $f_7; f_{11}$
M ₂	$f_5; f_7$ $f_{11}; f_{12}$	$f_3; f_5$ $f_8; f_{10}$	$f_1; f_2$ $f_7; f_{11}$	$f_3; f_5$ $f_8; f_{10}$	$f_3; f_5$ $f_7; f_{11}$	$f_3; f_5$ $f_8; f_{10}$	$f_4; f_8$ $f_{10}; f_{12}$	$f_3; f_5$ $f_7; f_{11}$
M ₃	$f_3; f_5$ $f_7; f_{10}$	$f_1; f_2$ $f_7; f_{11}$	$f_3; f_5$ $f_8; f_{10}$	$f_1; f_5$ $f_8; f_{10}$	$f_3; f_5$ $f_8; f_{10}$	$f_3; f_5$ $f_7; f_{11}$	$f_3; f_5$ $f_8; f_{10}$	$f_2; f_5$ $f_8; f_{10}$
M ₄	$f_3; f_5$ $f_7; f_{12}$	$f_1; f_6$ $f_9; f_{12}$	$f_1; f_2$ $f_7; f_{11}$	$f_1; f_3$ $f_5; f_{10}$	$f_3; f_5$ $f_8; f_{10}$	$f_3; f_6$ $f_7; f_8$	$f_3; f_8$ $f_9; f_{12}$	$f_3; f_5$ $f_8; f_{11}$
M ₅	$f_1; f_3$ $f_5; f_{10}$	$f_1; f_3$ $f_5; f_{10}$	$f_3; f_5$ $f_8; f_{10}$	$f_3; f_5$ $f_8; f_{10}$	$f_1; f_2$ $f_3; f_5$	$f_3; f_5$ $f_7; f_8$	$f_3; f_9$ $f_{10}; f_{12}$	$f_2; f_5$ $f_7; f_{11}$
M ₆	$f_5; f_7$ $f_{11}; f_{12}$	$f_1; f_2$ $f_7; f_{11}$	$f_3; f_5$ $f_8; f_{11}$	$f_3; f_6$ $f_9; f_{12}$	$f_1; f_6$ $f_9; f_{12}$	$f_3; f_6$ $f_7; f_{11}$	$f_3; f_8$ $f_9; f_{12}$	$f_3; f_5$ $f_7; f_{11}$
M ₇	$f_5; f_7$ $f_{11}; f_{12}$	$f_3; f_5$ $f_8; f_{10}$	$f_1; f_2$ $f_7; f_{11}$	$f_3; f_5$ $f_7; f_8$	$f_3; f_5$ $f_8; f_{10}$	$f_3; f_5$ $f_6; f_7$	$f_3; f_8$ $f_9; f_{12}$	$f_3; f_4$ $f_7; f_{11}$
M ₈	$f_3; f_7$ $f_{11}; f_{12}$	$f_1; f_6$ $f_9; f_{12}$	$f_3; f_5$ $f_7; f_{11}$	$f_5; f_6$ $f_9; f_{12}$	$f_1; f_2$ $f_6; f_{12}$	$f_3; f_5$ $f_7; f_8$	$f_3; f_8$ $f_{11}; f_{12}$	$f_2; f_5$ $f_7; f_{11}$
M ₉	$f_5; f_7$ $f_{11}; f_{12}$	$f_1; f_6$ $f_9; f_{12}$	$f_3; f_5$ $f_8; f_{10}$	$f_1; f_6$ $f_9; f_{12}$	$f_1; f_2$ $f_6; f_{12}$	$f_3; f_5$ $f_6; f_{12}$	$f_3; f_8$ $f_9; f_{12}$	$f_3; f_5$ $f_8; f_{12}$
M ₁₀	$f_3; f_7$ $f_{11}; f_{12}$	$f_3; f_5$ $f_8; f_{10}$	$f_1; f_2$ $f_7; f_{11}$	$f_3; f_5$ $f_7; f_8$	$f_3; f_5$ $f_6; f_{12}$	$f_2; f_3$ $f_5; f_{11}$	$f_5; f_7$ $f_9; f_{12}$	$f_3; f_5$ $f_7; f_{11}$

At a first glance the sets of features selected by the PCA were different both among different channels in the same subject and among the same channel in different subjects, but some features were selected more frequently than others, as we will discuss later.

For the classification, given the low number of features, the adopted kernel function was the previously recalled RBF (Eq. 4); this was generally assumed as the reasonable first choice, more general and versatile than linear kernel.

As already said, the two parameters H and α were determined by using the 10-fold cross validation and the classification was performed by LIBSVM 3.18, that provided also the optimized values for parameter b .

In Subsection 2.3 it was described the adopted procedure for the training; as far as the two sets N_{tr1} and N_{tr2} they were chosen in the percentage of 60% and 40% respectively of the training set N_{tr} .

To evaluate the goodness of the obtained classifiers the mean training accuracy T_a (in %) was considered (Table 2) as the average value of the accuracy calculated for twenty different random choices of the training sets. The optimal values for H^* and α^* were also reported.

Table 2: H^* , α^* and test accuracy of the classifiers obtained for the ten considered subjects.

S		SVM P4	SVM C4	SVM T8	SVM P8	SVM P3	SVM C3	SVM T7	SVM P7
M ₁	H^*	30	0.01	0.01	22	20	20	20	28
	α^*	-20	2	0.01	-16	-2	-2	-2	16
	T_a	92.4	87.2	100	89.1	80	77.1	92.6	67.2
M ₂	H^*	-20	30	4	14	22	8	-20	0.01
	α^*	4	-10	-6	-8	-6	-4	-6	4
	T_a	87.5	52.5	100	61.7	62.5	72.5	82.7	75
M ₃	H^*	-18	24	0.1	12	20	6	-10	12
	α^*	4	-4	-6	-2	2	-2	-4	-2
	T_a	92.5	72.4	100	68.5	72.5	63.9	84.7	50.2
M ₄	H^*	-28	4	0.2	16	10	8	-10	10
	α^*	2	-4	-6	-4	2	0.01	-6	-2
	T_a	80.5	60.4	98	74.6	57.5	52.3	68.7	90
M ₅	H^*	-20	10	0.1	12	0.01	6	-6	0.01
	α^*	4	-4	-6	-16	2	2	-4	-2
	T_a	84.2	88.2	96	80.5	54.3	54.7	92.7	73
M ₆	H^*	0.01	2	10	16	20	0.01	-10	10
	α^*	2	-2	-4	0.01	4	2	-2	0.01
	T_a	82.6	70.2	96	84.1	45.6	80.5	90.7	94
M ₇	H^*	30	0.01	10	22	10	0.01	10	26
	α^*	0.01	0.01	-4	2	-6	2	2	4
	T_a	98	100	96	98	80	96	100	98
M ₈	H^*	20	0.01	4	10	20	10	10	16
	α^*	-20	0.01	2	-16	2	0.01	-2	2
	T_a	96	96	98	100	85	98	100	100
M ₉	H^*	30	0.01	0.01	12	20	20	20	16
	α^*	-20	2	0.01	-8	-2	-2	0.01	4
	T_a	88	75.2	94.4	88.1	73.2	70.6	91.2	69.2
M ₁₀	H^*	20	8	0.01	10	2	30	4	0.01
	α^*	4	0.01	4	0.01	0.01	-2	-4	0.01
	T_a	64	85.1	98	88.4	45.3	60.1	94.8	81.5

For all the analyzed subjects, the first considered training set allowed to obtain an accuracy value above 80% in at least one channel.

The obtained classifiers were tested on the features of Table 1 selected from the signals of the set N_{test} and the percentage of success was determined and reported in Table 3. The results relative to the examined subjects were presented separately for the considered channels. The results were computed as the average among the ones obtained considering the twenty different random choices of the training sets; also the standard deviations were shown.

The percentage of accuracy was evaluated on the N_{tr2} set, whereas the performance of the classifier was calculated on the test set N_{test} .

Table 3: Results of the test for each subject and for all the significant (active) channels; the percentage of success (obtained as mean values in the different random choices of the sets N_{tr} and N_{test}) along with the standard deviations were reported. Note that cells were empty for all channels whose accuracy was below 80% (insignificant, inactive, channels).

S	Ch P4	Ch C4	Ch T8	Ch P8	Ch P3	Ch C3	Ch T7	Ch P7
M ₁	91.1 ±0.3	84.4 ±1.5	98.8 ±0.8	90 ±3.2	76.1 ±5.4	-	92 ±1.5	-
M ₂	84.2 ±3.9	-	89.1 ±6.5	-	-	-	83.2 ±5.3	-
M ₃	94.7 ±3.4	-	95.3 ±2.4	-	-	-	82.6 ±3.4	-
M ₄	81.1 ±4.3	-	96.8 ±3.4	-	-	-	-	87.2 ±5.3
M ₅	82.5 ±3.3	85.8 ±4.8	95.8 ±2.4	82.2 ±2.2	-	-	89.6 ±3.3	-
M ₆	81.7 ±2.1	-	95.6 ±2.6	82.2 ±1.1	-	80.2 ±3.2	88.6 ±4.1	-
M ₇	-	94 ±2.2	93.3 ±3.4	-	80.3 ±5.5	-	92.6 ±3.5	-
M ₈	80.8 ±6.4	-	95.5 ±3.2	85.4 ±2.4	-	82.8 ±4.8	87.6 ±3.8	-
M ₉	-	-	92.0 ±4.4	-	-	-	86.6 ±4.8	-
M ₁₀	-	82.3 ±3.2	96.1 ±3.2	84.4 ±2.8	-	-	92.6 ±3.8	82.6 ±3.8

By comparing the results obtained by the subjects and shown in Table 3, it could be noticed that the highest values of average percentage of success were obtained in the right hemisphere of the brain and in particular for the channel T8 (this channel was active in all the considered subjects), as could be observed by averaging inter-subjects data and maintaining distinct information for the channels. Moreover, the useful results in the right hemisphere were generally very close to the chosen threshold. On the other hand, T7 (left hemisphere) had also significant accuracy in almost all subjects but it was more isolated compared to its closer channels.

Regarding the computational time the algorithm, implemented in Matlab® on a personal computer (Intel(R) Core(TM) i7-4790 CPU @3.60 GHz 3.60 GHz RAM: 16,0 GB) took 81 minutes for processing all the trials used for the calibration step, including both training and validation (120

trials), 2.2 minutes for the test (80 trials), and $5.4 \cdot 10^{-4}$ sec. for a single trial processing (on-line).

Though the time necessary to process a single trial during the test was very low, almost real-time, the reduction of the computation time could be very important for the calibration step. At a first glance, by using the previously discussed classification results, we could reduce the computation time by a factor of 8 by using only information of the channel T8 (this could allow the use of a EEG helmet with only one electrode).

In the same direction, the computational efficiency could be further improved by considering only the features that best characterized the signals. From Table 1 it could be appreciated that the couple of features f_3 and f_5 was selected more often than the others.

With the aim of investigating the importance of these features for the classification success, we trained and tested a classifier using only the two features (and only the active channels of Table 3). The obtained results were shown in Table 4.

Table 4: Average and standard deviation percentage of classification success using only the features f_3 and f_5 in the active channels of Table 3. Note that the values of accuracy in the active channels were indicated also if the accuracy was below 80%.

S	Ch P4	Ch C4	Ch T8	Ch P8	Ch P3	Ch C3	Ch T7	Ch P7
M ₁	90.2 ±1.1	82.7 ±2.7	98.2 ±0.6	88 ±2.7	72.1 ±4.7	–	89.4 ±1.1	–
M ₂	74.7 ±7.9	–	86.1 ±3.6	–	–	–	73.2 ±4.8	–
M ₃	91.2 ±4.6	–	93.6 ±2.7	–	–	–	81.6 ±3.3	–
M ₄	65.4 ±8.1	–	94.2 ±3.1	–	–	–	–	82.1 ±6.9
M ₅	70.5 ±8.3	82.8 ±7.1	94.6 ±2.8	72.7 ±5.1	–	–	86. ±4.9	–
M ₆	71.3 ±8.7	–	95.2 ±3.4	81.3 ±2.1	–	71.3 ±5.1	85.7 ±4.3	–
M ₇	–	91.0 ±3.2	92.5 ±3.8	–	71.3 ±6.7	–	90.2 ±3.8	–
M ₈	72.1 ±8.4	–	92.0 ±3.8	83.7 ±3.9	–	70.9 ±6.8	88.3 ±5.1	–
M ₉	–	–	90.0 ±4.7	–	–	–	82.4 ±6.2	–
M ₁₀	–	80.6 ±5.0	93.4 ±4.3	81.7 ±4.6	–	–	88.2 ±5	70.9 ±6.3

By considering all the subjects, the results shown in Table 4 agreed with the ones obtained using all the features selected by the PCA; in all the subjects, the channel that provided the highest percentage values of classification success continued to be the channel T8. Since we aimed at testing the significant channels with only two features, we reported the accuracy results also where they were lower than 80%. However, the channel T8 remained above 80% in all the tested subjects. These results confirmed that the features selected by PCA were the most significant for signal classifications and allowed to consider an even smaller

number of features with a further reduction of computation time, thus making negligible also the time required for calibration (about 90sec, by considering just T8 and the couple features f_3 and f_5).

3.1 Comparison with previous works

As stated in the Introduction section, the self-induced stimulus of disgust has been proposed for the first time in a recent paper [22] where also a classification strategy was implemented. Despite the methodological differences, the two classification approaches provided very similar results, in both classification accuracy and computational efficiency, by considering as distinct all the channels in the method proposed therein. However, the machine learning method proposed herein has the following advantages compared to [22]:

- 1) *it was able to highlight the activation differences between channels, thus obtaining a better spatial resolution than [22], as the latter considered the average among all the channels in the right hemisphere;*
- 2) *it demonstrated the possibility of using only one channel (T8) in the classification to obtain an average accuracy that was above 90% (a minimal helmet could be constructed, with only one channel, to distinguish “disgust” from “relax”);*
- 3) *as a consequence of point 2) it could improve computational efficiency (the processing time could be reduced by a factor of 8);*
- 4) *as a consequence of point 1), it could allow the discrimination of a greater set of emotional states compared to [22].*

The present method could also be improved along the following directions:

- a) *it could eventually improve the classification accuracy by combining the information coming from more than just a single channel;*
- b) *it could be generalized by searching different activation patterns for different activation stimuli, if more activation tasks would be implemented.*

These aspects, being rather demanding (in particular, point a would require that all possible combinations of groups of channels are explored and analyzed), will be the objects of future investigations.

4 CONCLUSIONS

In this paper it was presented a fully automatic real time algorithm for the classification of low amplitude EEG signals (the disgust self-induced by remembering an unpleasant odor).

The signals were classified by a two stages algorithm: the first one was an off-line stage, aiming at the training of a suitable classifier whose input was the set of selected features; the second stage was the application of the classifier to new data. The specific considered signals required the use and the adaptation of mathematical tools like Wavelet Signal Decomposition theory, Principal Components Analysis and Support Vector Machine. The specific decomposition level to be considered depended on

the frequency of interest; in theory the proposed method allowed also to detect the frequency bands interested by a particular stimulus. The procedure, applied to real data coming from 10 different subjects, yielded satisfactory results in both accuracy and efficiency. Furthermore, the predominance of the right hemisphere in revealing the stimulus, and specifically of the channel T8, was confirmed. Ongoing work will be done in the following directions:

- to merge the results obtained from single channels in order to improve the classification accuracy;
- to study a minimal headset device (minimally cumbersome) to implement a robust binary BCI;
- to enlarge the number of different emotional states;
- to apply the proposed procedure to data coming from subjects affected by severe disabilities by using an adapted single channel device.

ACKNOWLEDGMENT

The Authors are very grateful to the “Fondazione Fabio Sciacca Onlus” for having supported this research project.

The Authors acknowledge GEA Soluzioni (Mr Giorgio Tonon) for having furnished the latest hardware/software upgrades for the used EEG measurement system.

Finally, the Authors would like to thank the anonymous Reviewers for their helpful comments and suggestions.

REFERENCES

- [1] J. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, and T.M. Vaughan, “Brain-computer interfaces for communication and control”, *Electroencephalography and Clinical Neurophysiology*, vol.113, no. 6, pp. 767–791, Jun, 2002.
- [2] N. Birbaumer, “Breaking the silence: Brain–computer interfaces (BCI) for communication and motor control”, *Psychophysiology*, Vol. 43, no. 6, pp. 517–532, Nov, 2006.
- [3] R. Liu, Y.Xuan Wang, L. Zhang, “An FDES-Based Shared Control Method for Asynchronous Brain-Actuated Robot”, *IEEE Trans. on Cybernetics*, in press.
- [4] A. Frisoli, C. Loconsole, D. Leonardis, F. Bann`o, M. Barsotti, C. Chisari, M. Bergamasco, “A New Gaze-BCI-Driven Control of an Upper Limb Exoskeleton for Rehabilitation in Real-World Tasks”, *IEEE Transactions on Systems, Man, And Cybernetics—Part C: Applications And Reviews*, vol. 42, pp. 1169–1179, Nov, 2012.
- [5] A.Vallabhaneni, T.Wang, B.He, “Brain–computer interface”, *Neural Eng.*, Springer, pp. 85–121, 2005.
- [6] R. Hari, and R. Salmelin, “Human cortical oscillations: a neuromagnetic view through the skull”, *Trends in Neurosciences*, vol. 20, no.1, pp. 44-49, 1997.
- [7] J. Donoghue, “Connecting cortex to machines: Recent advances in brain interfaces”, *Nature Neuroscience*, vol. 5, pp. 1085–1088, Nov. 2002.
- [8] R. Srinivasan, “Methods to improve the spatial resolution of EEG”, *International Journal of Bioelectromagnetism*, vol.1, no.1, pp. 102–111, 1999.
- [9] N. Birbaumer, L.G. Cohen, “Brain-computer interfaces: Communication and restoration of movement in paralysis”, *Journal of Physiology*, vol. 579, no. 3, pp. 621–636, Mar, 2007.
- [10] A. Kübler, F. Nijboer, J. Mellinger, T.M. Vaughan, H. Pawelzik, G. Schalk, D.J. McFarland, N. Birbaumer, and J.R. Wolpaw, “Patients with ALS can use sensorimotor rhythms to operate a brain computer interface”, *Neurology*, vol. 64, no. 10, pp. 1775–1777, May, 2005.
- [11] L.R. Hochberg, M.D. Serruya, G.M. Friehs, J.A. Mukand, M. Saleh, A.H. Caplan, A. Branner, D. Chen, R.D. Penn, and J.P. Donoghue, “Neuronal ensemble control of prosthetic devices by a human with tetraplegia”, *Nature*, vol. 442, no. 7099, pp. 164–171, Jul, 2006.
- [12] J.R.Wolpaw,D.J.McFarland, “Control of a two-dimensional movement signal by a non invasive brain–computer interface in humans”, *Proc.Natl.Acad.Sci.U.S.A.* 101(51)(2004)17849–17854.
- [13] C. Neuper, G.R. Müller-Putz, R.Scherer, G. Pfurtscheller, “Motor imagery and EEG-based control of spelling devices and neuro prostheses”, *Prog. Brain Res.* 159, pp. 393–409, 2006.
- [14] Y.Han, H.Bin, “Brain–computer interfaces using sensori motor rhythms: current state and future perspectives”, *IEEE Trans. Biomed. Eng.*, vol. 61, no. 5, pp 1425–1435, May, 2014.
- [15] L.A.Farwell, E. Donchin, “Talking off the top of your head: toward mental prosthesis utilizing event-related brain potentials”, *Electroencephalogr. Clin. Neurophysiol.* Vol.70, no. 65, pp.10–523, Dec, 1988.
- [16] A.Furdea, S.Halder, D.J. Krusienski, D.Bross, F.Nijboer, N.Birbaumer, A.Kübler, “An auditory odd ball (P300) spelling system for brain–computer interfaces”, *Psychophysiology*, vol. 46, no. 3, pp. 617–625, 2009.
- [17] A.Kübler, A.Furdea, S.Halder, E.M.Hammer, F.Nijboer, B.Kotthoube, “A brain–computer interface controlled auditory event-related potential (p300) spelling system for locked-in patients”, *Ann. N.Y. Acad. Sci.*, vol.1157, pp.90–100, Mar, 2009.
- [18] G. Shangkai, W.Yijun, G.Xiaorong, H.Bo, “Visual and auditory brain–computer interfaces”, *IEEE Trans. Biomed. Eng.* vol.61, no.5, pp.1436–1447, May, 2014.
- [19] A.M.Brouwer, J.B.F.vanErp, “A tactile P300 brain –computer interface”, *Front. Neurosci.*, vol. 4, no.19, May, 2010.
- [20] G.R. Müller-Putz, R.Scherer, C.Neuper, G.Pfurtscheller, “Steady-state somato- sensory evoked potentials: suitable brain signals for brain–computer inter- faces?”, *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol.14, no.1, pp.30–37, March, 2006.
- [21] Y.P. Lin, C.H. Wang, T.L. Wu, S.K. Jeng, J.K. Chen, “Support Vector Machine for EEG Signal Classification during Listening to Emotional Music”, on *Proceedings of IEEE 10th Workshop Multimedial Signal Processing IEEE*, Cairns, 2008, pp. 127-130.
- [22] G. Placidi, D. Avola, A. Petracca, F. Scallari, M. Spezialetti, “Basis for the Implementation of an EEG-based Single-Trial Binary Brain Computer Interface through the Disgust Produced by Remembering Unpleasant Odors”, *Neurocomputing*, vol.160, no. 21, pp. 308-318, Jul, 2015.
- [23] J.L.M. Levy, R.I.Henkin, C.S.Lin, A.Hutter, D.Schellinger, “Odor memory induces brain activation as measured by functional MRI”, *J. Comput. Assist. Tomogr.*, vol.23, 487–498, Jul, 1999.
- [24] B.W. Dahanayake, A.R.M. Upton, “On line methodology for detection of suspected epileptic seizures from electroencephalogram”, *Control Engineering Practice*, vol. 2, no. 5, pp. 187-190, Jul, 1994.
- [25] A. Subasi, and E. Erceleb, “Classification of EEG signals using neural network and logistic regression”, *Computer Methods and Programs in Biomedicine*, vol. 78, pp. 87-99, May, 2005.
- [26] J. Štastný, P. Sovka, A. Stancák, “EEG Signal Classification: Introduction to the Problem”, *Radioengineering*, vol. 12, no. 3, pp. 51-55, Sept, 2003.
- [27] F. Benimeli, and K. Sharman, “Electroencephalogram signal classification for brain computer interfaces using wavelets and support vector machines”, in *ESAN’2007 Proceedings European Symposium on Artificial Neural Networks*, Bruges (Belgium), 2007, pp 361-366.
- [28] R. Panda, P.S. Khobragade, P.D. Jambhule, S.N. Jengthe, P.R. Pal, T.K. Gandhi, “Classification of EEG signal using wavelet transform and support vector machine for epileptic seizure diction”, in *Proceedings International Conference on Systems in Medicine and Biology*, Kharagpur (India), 2010, pp.405-408.
- [29] C. Christoforou, R. Haralick, P. Sajda, L.C. Parra, “Second-Order Bilinear Discriminant Analysis”, *Journal of Machine Learning Research*, vol. 11, pp. 665-685, 2010.
- [30] A. Subasi, and M.I. Gürsoy, “Comparison of PCA, ICA and LDA in EEG signal classification using DWT and SVM”, *Expert Systems with Applications*, vol. 37, Jan, 2010.
- [31] F.Song, Z.Guo, D.Mei, “Feature selection using principal component analysis”, 2010 International Conference on System Science, Engineering Design and Manufacturing Informatization, 2010, pp. 27-30.

- [32] P. Bhuvaneshwari, and J.S. Kumar, "Support vector machine technique for EEG signals", *International Journal of Computer Applications*, vol. 63, no. 13, pp. 1-5, Feb, 2013.
- [33] Y. Yin, D. Xu, X. Wang, M. Bai, "Online State-Based Structured SVM Combined With Incremental PCA for Robust Visual Tracking", *IEEE Trans. on Cybernetics*, 45, pp.1988-2000, Sep, 2015.
- [34] G.B. Huang, H. Zhou, X. Ding, R. Zhang, "Extreme Learning Machine for Regression and Multiclass Classification", *IEEE Transactions on Systems, Man, And Cybernetics—Part B: Cybernetics*, 42, pp.513-529, Apr, 2012
- [35] I. Guler, and E.D. Ubeyli, "Multiclass support vector machines for EEG-signals classification", *IEEE Transactions on Information Technology in Biomedicine*, vol. 11, no. 2, pp. 117-126, Mar, 2007.
- [36] H.H. Jasper, "The ten-twenty electrode system of the International Federation", *Electroencephalogr Clinical Neurophysiology*, vol. 10, no. 2, pp. 371-375, 1958.
- [37] I. Daubechies, *Ten lectures on wavelet*, Philadelphia, 1992.
- [38] S. Blanco, A. Figliola, R. Quian Quiroga, O.A. Rosso, E. Serrano, "Time-frequency analysis of electroencephalogram series. III. Wavelet packets and information cost function", *Physical Review E*, vol.57, no. 1, pp. 932-940, Jan, 1998.
- [39] Md. Mamun, Mahmoud Al-Kadi, M. Marufuzzaman, "Effectiveness of Wavelet Denoising on Electroencephalogram Signals", *Journal of Applied Research and Technology*, vol. 11, no.1, pp. 156-160, Feb, 2013.
- [40] C.W. Hsu, C.C.Chan, C.J. Lin. A Practical Guide to Support Vector Classification, *Bioinformatics*, vol.1, no.1, 2003, available at <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>.
- [41] B. Efron, "Estimating the error rate of a prediction rule: improvement on cross-validation", *Journals- American Statistical Association*, vol. 78, no. 382, pp. 316-331, Jun, 1983.
- [42] C-C.Chang, and C-J. Lin, "LIBSVM: A library for support vector machines", *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no.27, pp.1-27, 2011. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [43] D. Cvetkovic, E.D. Übeyli, I. Cosic, "Wavelet transform feature extraction from human PPG, ECG, and EEG signal responses to ELF PEMF exposures: A pilot study", *Digital Signal Processing*, vol. 18, pp. 861-874, 2008.
- [44] A. Subasi, "EEG signal classification using wavelet feature extraction and a mixture of expert model", *Expert Systems with Applications*, vol. 32, no.4, pp. 1084-1093, May, 2007.
- [45] I. Omerhodzic, S. Avdakovic, A. Nuhanovic, K. Dizdarevic, "Energy Distribution of EEG signals: EEG signal wavelet-neural network classifier", *World Academy of Science, Engineering and Technology*, vol. 61, pp. 1190-1195, 2013.