# Unsupervised Analysis of Event-related Potentials (ERPs) during an emotional Go/NoGo Task

Paolo Masulli[1], Francesco Masulli[2,3], Stefano Rovetta[2], Alessandra Lintas[1], and Alessandro E. P. Villa[1]

[1] NeuroHeuristic Research Group, University of Lausanne
Internef, Quartier UNIL Dorigny, 1015 Lausanne (Switzerland)
[2] Department of Computer Science, Bioengineering, Robotics and Systems Engineering
DIBRIS, University of Genova, 16146 Genova (Italy)
[3] Sbarro Institute for Cancer Research and Molecular Medicine, Temple University, Philadelphia, PA (USA)
{paolo.masulli,alessandra.lintas,alessandro.villa}@unil.ch
{francesco.masulli,stefano.rovetta}@unige.it

**Abstract.** We propose a framework for an unsupervised analysis of electroencephalography (EEG) data based on possibilistic clustering, including a preliminary noise and artefact rejection. The proposed data flow identifies the existing similarities in a set of segments of EEG signals and their grouping according to relevant experimental conditions. The analysis is applied to a set of event-related potentials (ERPs) recorded during the performance of an emotional Go/NoGo task. We show that the clusterization rate of trials in two experimental conditions is able to characterize the participants. The extension of the method and its generalization is discussed.

**Keywords:** EEG, ERP, possibilistic clustering, decoding, brain activity

## 1 Introduction

The recording of human brain activity is generally performed by electroencephalography (EEG), which has the advantage of being a non invasive technique using external electrodes placed over many standard locations determined by skull landmarks. It is recognized that brain circuits that are activated at any occurrence of the same mental or physical stimulation generate transient electric potentials than can be averaged over repeated trials [19]. Such brain signals recorded by EEG and triggered by a specific event are referred to as event-related potentials (ERPs).

The conventional analysis of ERPs rely on a supervised approach, mainly on the experimenter's experience in order to discard outliers and detect wave components that are associated with different neural processes. In order to decrease the impact of the bias due to the human supervision, machine learning techniques have been recently presented [15]. Techniques using Echo State Networks [2], or interval features [13], or statistical techniques such as Independent Component Analysis [17, 24] have been successfully applied to tackle a specific classification problem following a phase of feature extraction. However, these techniques bear the disadvantage of depending on

a careful choice of the feature extraction technique which captures a pre-determined aspect of the variability. In a way the bias of human supervision is moved to another step of the analysis, but the pre-selection of specific features might be prone to learning biases or over-fitting of the training data.

It has been recognized for a long time the value of single-trial analysis [5] and the importance of analyzing all time points to reveal the complete time course of the effect of a triggering event, either mental or physicial [22]. In order to reduce the bias of an *a priori* criteria for a supervised classification we propose an unsupervised data processing flow based on a probabilistic clustering aimed to capture similarities between different trials in the data set, without previously extracting features. Our approach is not intended to provide a direct answer to a specific classification problem. It is rather thought as a general tool to clusterize ERPs based on their internal structure. We present the application of this technique to a set of ERPs recorded during the performance of an emotional Go/NoGo task.

## 2 Methods

### 2.1 Graded Possibilistic Clustering

The *Graded Possibilistic Clustering* approach [16] is a central clustering model able to achieve fast convergence and low outliers rejection of fuzzy clustering methods with probabilistc constraint (as in Fuzzy C-Means [1, 6], and Deterministic Annealing [21, 20]), while at the same time avoiding the difficult convergence and high outliers rejection of fuzzy clustering methods with possibilistic constraint (such as Possibilistic C-Means [11, 12]).

In the current work we propose a new version of the Graded Possibilistic Clustering model (GPC-II). Let us consider a set $X$ of $k$ observations (or instances) $\mathbf{x}_l$, for $l \in \{1,\dots,n\}$, and a set $C$ of $c$ fuzzy clusters denoted $C_1,\dots,C_c$. The clusters are represented via their centroids $\mathbf{y}_j$, for $j \in \{1,\dots,c\}$. To an observation $\mathbf{x}_l$, each cluster associates a fuzzy cluster indicator (or membership) function

$$u_{lj} \in [0,1] \subset \mathbb{R}. \tag{1}$$

The *total membership mass* of an observation $\mathbf{x}_l$ is defined as:

$$\zeta_l = \sum_{j=1}^{c} u_{lj}. \tag{2}$$

The membership of the observation $\mathbf{x}_l$ to the cluster $u_j$ can be expressed as:

$$u_{lj} = \frac{v_{lj}}{Z_l}, \tag{3}$$

where

$$v_{lj} = e^{-d_{lj}/\beta_j} \tag{4}$$

is the *free membership*, and

$$Z_l = \zeta_l^\alpha = \left( \sum_{j=1}^{c} v_{lj} \right)^\alpha \quad , \quad \alpha \in [0,1] \subset \mathbb{R} \tag{5}$$

is the *generalized partition function*.

In the last two equations, $d_{lj}$ is the distance between the $j$-th centroid and the observation $\mathbf{x}_l$, coefficients $\beta_j$ are model parameters playing a role in the representation of data as cluster widths, and the parameter $\alpha$ controls the "possibility level" of the GPC-II, from a totally probabilistic ($\alpha = 1$) to a totally possibilistic ($\alpha = 0$) model, with all intermediate cases for $0 < \alpha < 1$.

In GPC-II the clusters' centroids are related to the membership vectors via the equation:

$$\mathbf{y}_j = \frac{\sum_{l=1}^{n} u_{lj} \mathbf{x}_l}{\sum_{l=1}^{n} u_{lj}} \ . \tag{6}$$

The implementation of the GPC-II is based on a a Picard iteration of Eq.s 3 and 6 after a random initialization of centroids.

Note that for $\alpha = 1$, the representation properties of the method coincide with those of Deterministic Annealing [20, 21]). When $\alpha = 0$, they are equivalent to those of Possibilistic C-Means [12]). In the intermediate cases, as soon as $\alpha > 0$, there is a degree of competition between clusters, as in probabilistic models, but memberships eventually vanish for points sufficiently far away from the centroids, as in the possibilistic case.

A deterministic annealing version of the Graded Possibilistic Clustering model (DAGPC-II) can be implemented by decomposing the model parameters $\beta_j$ as $\beta_j = \beta b_j$, where $\beta$ is the optimization parameter for the deterministic annealing procedure (starting small and enlarging each time the Picard iteration converges) and $b_j$ are the *relative* cluster scales obtained from some heuristic like those proposed in [12]. In this way the Graded Possibilistic Clustering can benefit of the powerful optimization technique proposed in [20, 21] that, after starting from fully overlapped clusters, with the increasing of $\beta$ performs a hierarchical clustering (with overlapped clusters splitting), able to find the "natural" aggregations of observations.

### 2.2 Overlapping of clusters

The Jaccard index expresses how much two subsets are overlapping. It is defined by the following expression:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}. \tag{7}$$

For fuzzy clusters as the ones we consider, the belonging of an element to a cluster is expressed via its membership indicator, which is a real value in the interval $[0,1]$. In this case we can use following definitions of fuzzy cardinality for the fuzzy clusters to make sense of Eq. (7).

$$|C_l \cap C_m| = \sum_{j=1}^{m} \min(u_{lj}, u_{mj}), \qquad |C_l \cup C_m| = \sum_{j=1}^{m} \max(u_{lj}, u_{mj}),$$

The resulting Jaccard index is a real value in the interval $[0,1]$.

### 2.3 Experimental data

Nineteen volunteers (5 females, mean age (SD) 28 (6.69)) were fitted with EEG equipment and recorded using 64 scalp active electrodes (ActiveTwo MARK II Biosemi EEG System) at a sampling frequency of 2048 Hz. We present here an analysis limited to brain signals recorded on the electrode Fz.

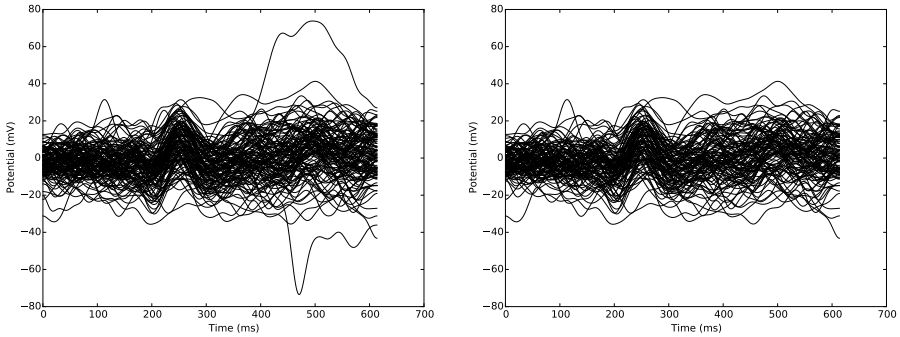
The behavioral task consisted in an emotional Go/NoGo task [7]. The stimulus presentation and response collection software was programmed using the software EPrime (Psychology Software Tools, Inc., Sharpsburg, PA 15215-2821, USA). The Go-cues of the task required participants to look at a picture with a face presented in the center of a computer screen and respond as fast as possible by pressing a button when a face expressing neutral emotion was displayed. In the NoGo-cues trials the participants withheld responses to non-target stimuli. For each trial the stimulus presentation lasted for a duration of 500 ms, followed by a fixation mark (+) for 1000 ms. The task consisted of four blocks of 30 pseudo-randomized "Go/NoGo" trials. Each block contained 20 "Go" and 10 "NoGo" trials. A neutral emotional expression of a face was paired with emotional expressions (happiness, fear, anger, or sadness) of the same face. The neutral expression served as the "Go" cue, while anyone of the emotional expressions was the "NoGo" cue.

### 2.4 Within-participant data processing flow

For each participant the analysis of all Go/NoGo trials is aimed to exclude the trials that contain artefacts, and then identify clusters of the remaining trials, without any information whether the stimulus was a "GO" or a "NoGo" cue. The data processing flow consists of the following phases.

1. **EEG Pre-processing.** The data of each participant are imported in EEGLAB [4] and re-referenced with respect to the two mastoidals M1 and M2. Subsequently, the filter IIR Butterworth (Basic Filter for continuous EEG in the software) is applied, with Half-Power parameters 0.1 Hz for High-Pass and 27.6 Hz for Low-Pass.
2. **Segments extraction.** The resulting signals are read with a Python script using the EDF reader present in the package `eegtools`. The relevant segments of the EEG recording are extracted according to the triggers which identify the experimental trials. We have considered only the correct trials (i.e., the participants either pressed the button after a "Go" cue, or they did not press the button for a "NoGo" cue), on average 108 (out of 120) trials per participant. We considered a time course of 600 ms for the ERPs, starting at the beginning of the trial (the instant when the picture with the face is presented to the participant), in order to avoid the presence of muscular artifacts produced at the time the participant pressed the button. For each trial, we apply a baseline correction by subtracting the average value of amplitude of the signal in the 200ms immediately before the trial onset. With a sampling rate of 2048 Hz, the segments with a duration of 600 ms correspond to vectors in $\mathbb{R}^{1228}$ and they form a subset of$\mathbb{R}^{1228}$.
3. **Smoothing.** We apply the 1-dimensional Anisotropic Diffusion algorithm by Perona & Malik [18] to smooth each signal segment (parameters: 1000 iterations, `delta_t` = 1/3, `kappa` = 3).

4. **Artefact rejection.** The peak-to-peak amplitude was computed for each segment. We rejected those segments characterized by a difference larger than twice the median absolute deviation (MAD) between the peak-to-peak value of a trial and the median peak-to-peak value computed over all correct trials. On average this procedure rejected approximately 14 segments per participant. The remaining segments ($n = 94$ on average, out of 120 initial trials) for each participant were considered valid for further analysis. The set of valid data is denoted $X \subset \mathbb{R}^{1228}$. An example is displayed in Figure 1.




(a) All segments for participant #12721 (b) Valid segments after removal of 11 trials

Fig. 1: Rejection of the ERP segments tagged as artefacts according to the peak-to-peak amplitude criterion.

5. **Possibilistic clustering.** We apply the Graded Possibilistic Clustering algorithm with deterministic annealing (Section 2.1) with parameter $\alpha = 0.85$ and initial number of clusters equal to $c_0 = 7$, which is chosen heuristically to be between 2 and 3 times the actual number of clusters we expect to find in the data, obtaining, for each cluster $C_j$, its centroid $\mathbf{y}_j \in X$ and its membership vector $u_{\cdot j} \in \mathbb{R}^{|X|}$, whose components express how much each of the segments belongs to the cluster.
6. **Singleton clusters rejection.** Any cluster with its centroid modeling only a single trial was removed. Such clusters, the singletons, are identified among those with a total membership differing from the average membership of all clusters by more than 1.5 standard deviations. In case there is only one segment, among all the segments, with maximum membership in the cluster, then the segment is removed and the clustering algorithm of point 5 is applied iteratively. Hence the number of required clusters is decreasing by one and the procedure is repeated until no more segments are removed.
7. **Merging of overlapping clusters.** The Jaccard index, defined in Section 2.2, is computed to determine the closeness for each pair of clusters. For each participant we compute independently the distribution of these values, which tends to be a multimodal distribution that can be decomposed as a mixture of distributions. The leftmost mode includes the neighbor clusters, very close to each other given

a Jaccard index towards 1. Different methods used to fit this density show that a cut-point separating the neighbor from the remaining clusters tended to be close to a value $J_0 = 0.7$. For this reason we have decided to fix this value for all participants as the significant threshold value characterizing overlapping clusters. Hence, two clusters are merged if their index surpasses the threshold value $J_0$. After merging, the new centroid is computed as the average of the centroids of the previous clusters weighed by the membership values of their elements and we determine the new membership vector as done in the clustering algorithm of point 5. Then, new Jaccard indexes are recomputed and the procedure is applied iteratively until no neighbor clusters can be merged any more.

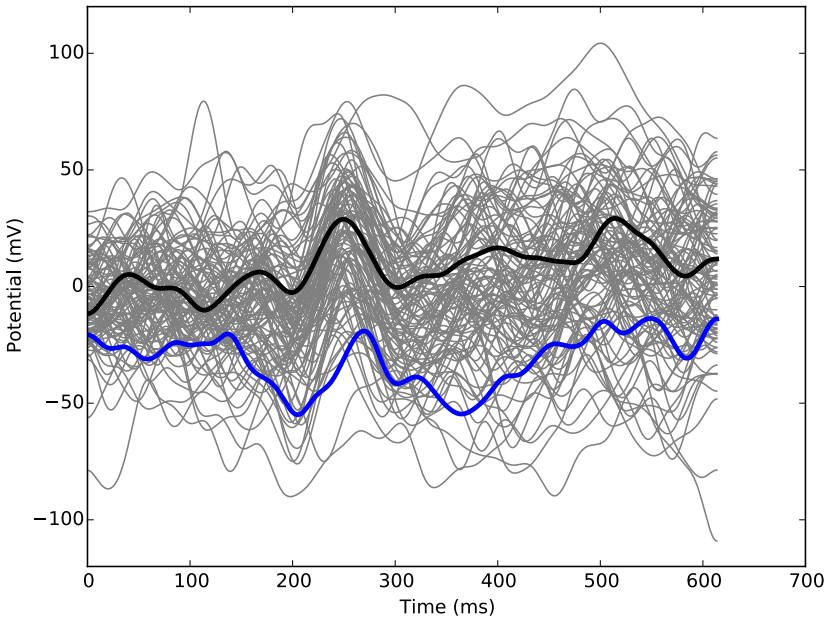



Fig. 2: ERP segments for participant participant #12721 (thin grey lines) and the two clusters (represented by their centroids with thick blue and black lines).

### 2.5 Clusterization rate

The interest and the validity of the clusters obtained through the procedure described above is dependent on the specific data set. In this study, we evaluate the outcome of the analysis by considering whether the “Go” and “NoGo” trials were clusterized according to the experimental conditions. For each participant we computed a *clusterization rate* of the “Go” trials (respectively, the “NoGo”) trials defined as the number of “Go”

segments (respectively, “NoGo”) characterized by a high membership to one of the identified clusters (i.e., signal segments with a membership above the 95-th percentile of the cluster) divided by the total number of “Go” trials (respectively, “NoGo”).

## 3 Results

For each participant we observed a number of clusters $2 \leq C \leq 4$. An example is displayed in Figure 2.

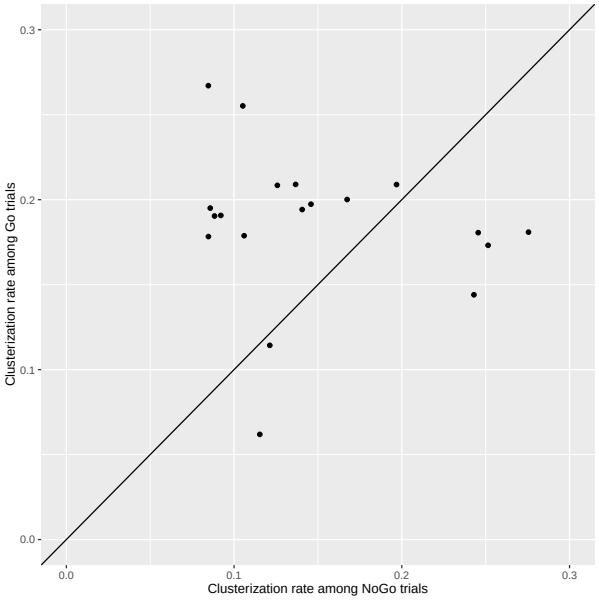
The confusion matrix of the different experimental conditions (Go/NoGo condition, type of emotion) *vs.* the membership to the clusters did not provide relevant differences between the clusters. This suggests that the ERPs features related to high cognitive functions such as affective discrimination are not associated with simple wave components of the brain signal itself and might require the choice of specific features or a different metric of the signals to be detected.

Nevertheless, we observed a significant difference of the clusterization rate of the “Go” and “NoGo” trials for most participants (Figure 3). The probabilistic clustering is implicitely depending on the random seed used at the begin of the procedure, for we repeated the same analysis with 20 different random seeds in order to evaluate the robustness of the result for each participant. We observed that for 13 participants the average clusterization rates for the Go trials were larger than the corresponding clusterization rates for the NoGo trials (Figure 3a).

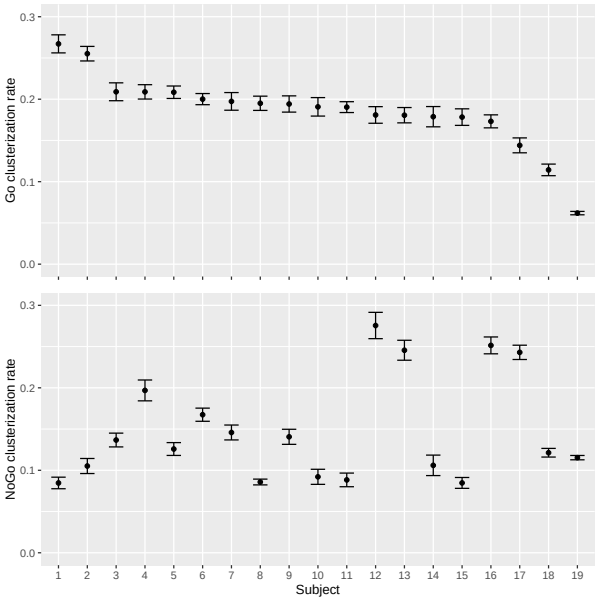
For each participant the variability of the clusterization rates depended on the variability between segments, on the number of correct available trials and on the nonlinear interaction of these factors with the random initialization of the probabilistic clustering. The variability for the Go trials was smaller than for the NoGo trials (Figure 3b) The average values of clusterization of the Go trials for different participants were more similar to each other, while the average values for the clusterization of the NoGo trials varied more between participants. Indeed, it was expected that the ERP segments of the “Go” trials, with a neutral expression of the cue, were more similar and hence would be associated with a high membership to one of the identified clusters. Conversely, the NoGo ERPs were associated with different face expressions and tended to be more scattered and differ more between each other.

## 4 Discussion and conclusions

The analysis presented here demonstrates a possible approach towards an unsupervised EEG signal analysis. The choices that we have made in the design of the data processing flow make it very general and easy to extend to other datasets, since the user is not required to choose a specific set of features to extract and analyze. The only value used in our processing flow that was derived from a post-hoc analysis is the threshold value $J_0$ of the Jaccard index used to merge neighbor clusters. We set this value $J_0 = 0.7$ and it we can foresee that for other datasets this value might not be optimal. What is the effect of choosing a value of 0.6, 0.65, or 0.8 is worth to be considered in future work and with other data sets.




(a)




(b)

Fig. 3: Clusterization rate comparison of Go and NoGo trials for all participants. (a) Scatter plot of the average values over 20 different random seed initializations (each point corresponds to one participant). For the majority of the participants (n = 13), the clusterization rate of the Go trials is higher than the one of the NoGo trials. This expresses the fact that the Go trials tend to be more similar to each other, while the NoGo trials are characterized by higher intrinsic variability which is not necessarily captured in the clusters. (b) Average value and standard error of the clusterization rate of Go and NoGo trials for each participant over 20 different random seed initializations. Notice that the clusterization rates of the Go trials are characterized by a similar variation across all the participants, while the NoGo trials have a much broader distribution of the standard error.

The trade-off of our approach is that the clustering algorithm compares the signals directly by computing their distances as vectors, which might lead to the loss of higher order features and to comparison problems between different trials given by time shifts and delayed onset of EEG peaks. The complexity of the information processed by the brain will certainly require a set of techniques able to analyze simultaneously the brain activity at various time scales [9, 3, 23, 25]. Decoding brain states cannot be expected to be achieved by totally unsupervised or partially supervised machine learning techniques [15]. For the analysis of specific features of the ERPs a necessary approach will consider the use of metrics able to detect and compare peaks and wave components and accounting for time warping [8, 14, 10, 26]. Another future development which we shall address is the comparison of the EEG signals of different participants with similar unsupervised techniques, with the aim of grouping the participnts according to their mutual similarities. This will require the application of normalization to account for inter-participant variability of the amplitude of the signals.

Beyond brain research and clinical applications, these results will be applicable to end-user applications of EEG recordings and might be integrated in brain-machine interfaces featuring EEG sensors, thereby bringing important societal benefits through applications to the domains of health, wellness and well-being. Additional goals will be the generalization of the analyses to different types of neurophysiological signals, such as spike trains recorded by intra-cranial electrodes, and to time-series data coming from different sources.

**Acknowledgments** This work was partially supported by the Swiss National Science Foundation grant CR13I1-138032.

## References

1. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Springer US, Boston, MA (1981)
2. Bozhkov, L., Koprinkova-Hristova, P., Georgieva, P.: Learning to decode human emotions with Echo State Networks. Neural Networks 78, 112–119 (Jun 2016)
3. Del Prete, V., Martignon, L., Villa, A.E.: Detection of syntonies between multiple spike trains using a coarse-grain binarization of spike count distributions. Network 15(1), 13–28 (Feb 2004)
4. Delorme, A., Makeig, S.: EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of Neuroscience Methods 134(1), 9–21 (Mar 2004)
5. Donchin, E.: Discriminant analysis in average evoked response studies: the study of single trial data. Electroencephalogr Clin Neurophysiol 27(3), 311–4 (Sep 1969)
6. Dunn, J.C.: Some Recent Investigations of a New Fuzzy Partitioning Algorithm and its Application to Pattern Classification Problems. Journal of Cybernetics 4(2), 1–15 (Jan 1974)
7. Hare, T.A., Tottenham, N., Davidson, M.C., Glover, G.H., Casey, B.J.: Contributions of amygdala and striatal activity in emotion regulation. Biological Psychiatry 57(6), 624–632 (Mar 2005)
8. Ihrke, M., Schrobsdorff, H., Herrmann, J.M.: Recurrence-based estimation of time-distortion functions for ERP waveform reconstruction. Int J Neural Syst 21(1), 65–78 (Feb 2011)

9. Indic, P.: Time scale dependence of human brain dynamics. Int J Neurosci 99(1-4), 195–199 (Aug 1999)
10. Karamzadeh, N., Medvedev, A., Azari, A., Gandjbakhche, A., Najafizadeh, L.: Capturing dynamic patterns of task-based functional connectivity with EEG. Neuroimage 66, 311–317 (Feb 2013)
11. Krishnapuram, R., Keller, J.M.: A possibilistic approach to clustering. IEEE transactions on fuzzy systems 1(2), 98–110 (1993)
12. Krishnapuram, R., Keller, J.M.: The possibilistic c-means algorithm: insights and recommendations. IEEE transactions on Fuzzy Systems 4(3), 385–393 (1996)
13. Kuncheva, L.I., Rodríguez, J.J.: Interval feature extraction for classification of event-related potentials (ERP) in EEG data analysis. Progress in Artificial Intelligence 2(1), 65–72 (2013)
14. Lederman, D., Tabrikian, J.: Classification of multichannel EEG patterns using parallel hidden Markov models. Med Biol Eng Comput 50(4), 319–328 (Apr 2012)
15. Lemm, S., Blankertz, B., Dickhaus, T., Müller, K.R.: Introduction to machine learning for brain imaging. Neuroimage 56(2), 387–99 (May 2011)
16. Masulli, F., Rovetta, S.: Soft transition from probabilistic to possibilistic fuzzy clustering. IEEE Transactions on Fuzzy Systems 14(4), 516–527 (Aug 2006)
17. Mueller, A., Candrian, G., Kropotov, J.D., Ponomarev, V.A., Baschera, G.M.: Classification of ADHD patients on the basis of independent ERP components using a machine learning system. Nonlinear biomedical physics 4(Suppl 1), S1 (2010)
18. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. IEEE Transactions on pattern analysis and machine intelligence 12(7), 629–639 (1990)
19. Picton, T.W., Bentin, S., Berg, P., Donchin, E., Hillyard, S.A., Johnson, R., Miller, G.A., Ritter, W., Ruchkin, D.S., Rugg, M.D., Taylor, M.J.: Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. Psychophysiology 37(2), 127–152 (Mar 2000)
20. Rose, K., Gurewitz, E., Fox, G.: A deterministic annealing approach to clustering. Pattern Recognition Letters 11(9), 589–594 (1990)
21. Rose, K., Gurewitz, E., Fox, G.C.: Statistical mechanics and phase transitions in clustering. Phys. Rev. Lett. 65(8), 945–948 (Aug 1990)
22. Rousselet, G.A., Pernet, C.R.: Quantifying the Time Course of Visual Object Processing Using ERPs: It's Time to Up the Game. Front Psychol 2, 107 (2011)
23. Smith, R.X., Yan, L., Wang, D.J.J.: Multiple time scale complexity analysis of resting state FMRI. Brain Imaging Behav 8(2), 284–91 (Jun 2014)
24. Stewart, A.X., Nuthmann, A., Sanguinetti, G.: Single-trial classification of EEG in a visual object task using ICA and machine learning. Journal of Neuroscience Methods 228, 1–14 (May 2014)
25. Wohrer, A., Machens, C.K.: On the number of neurons and time scale of integration underlying the formation of percepts in the brain. PLoS Comput Biol 11(3), e1004082 (Mar 2015)
26. Zoumpoulaki, A., Alsufyani, A., Filetti, M., Brammer, M., Bowman, H.: Latency as a region contrast: Measuring ERP latency differences with Dynamic Time Warping. Psychophysiology 52(12), 1559–1576 (Dec 2015)