Use of genetic algorithm for the selection of EEG features

P Asvestas¹, A Korda², S Kostopoulos¹, I Karanasiou³, A Ouzounoglou², K Sidiropoulos¹, E Ventouras¹ and G Matsopoulos²
¹ Department of Biomedical Engineering, Technological Educational Institute of Athens, Greece
² School of Electrical and Computer Engineering, National Technical University of Athens, Greece
³ Institute of Communications and Computer Systems, Athens, Greece

E-mail: pasv@teiath.gr

Abstract. Genetic Algorithm (GA) is a popular optimization technique that can detect the global optimum of a multivariable function containing several local optima. GA has been widely used in the field of biomedical informatics, especially in the context of designing decision support systems that classify biomedical signals or images into classes of interest. The aim of this paper is to present a methodology, based on GA, for the selection of the optimal subset of features that can be used for the efficient classification of Event Related Potentials (ERPs), which are recorded during the observation of correct or incorrect actions. In our experiment, ERP recordings were acquired from sixteen (16) healthy volunteers who observed correct or incorrect actions of other subjects. The brain electrical activity was recorded at 47 locations on the scalp. The GA was formulated as a combinatorial optimizer for the selection of the combination of electrodes that maximizes the performance of the Fuzzy C Means (FCM) classification algorithm. In particular, during the evolution of the GA, for each candidate combination of electrodes, the well-known ($\Sigma, \Phi, \Omega$) features were calculated and were evaluated by means of the FCM method. The proposed methodology provided a combination of 8 electrodes, with classification accuracy 93.8%. Thus, GA can be the basis for the selection of features that discriminate ERP recordings of observations of correct or incorrect actions.

1. Introduction
A major task of medical signal analysis is to develop algorithms (also known as classifiers) that will be capable to discriminate between signals from various classes of interest (for example, normal vs. pathological). A significant step towards this direction is to select quantitative indices (features) that: a/ provide a compact representation of the signals under consideration and b/ present different values for signals that do not belong to the same class. Thus, instead of “feeding” the classifier with the whole signal (which can contain a large number of samples), a much smaller number of feature values is used. The process of feature selection helps in understanding data, reducing computation requirement, and improving the performance of the classifier [1].

Usually, an initial set of candidate features is formed and an optimization method is applied in order to select the optimum subset of features. The optimization process involves the definition of an objective function that evaluates a subset of features, a generator of candidate subsets of features and a termination condition. The subset of features that provides the best value of the objective function is selected. One of the most popular global optimization methods is Genetic Algorithm (GA), which is a heuristic method that searches for the best solutions in a large search space by simulating the process of natural selection and genetic evolution.
algorithm that mimics the process of natural selection [2]. GA have been widely used for feature selection [3], [4], wherein the chromosome bits represent whether the feature is selected or not.

The aim of this paper is to present a methodology, based on GA, for the selection of the optimal subset of features that can be used for the efficient classification of Event Related Potentials (ERPs), which are recorded during the observation of correct or incorrect actions. ERPs are a special category of electroencephalographic (EEG) signals, which are recorded from various locations on a subject’s scalp when the subject is presented with external stimuli or events. ERPs provide non-invasive measurements of the electrical activity of the brain and describe the specific cognitive processes that are responsible for processing the stimuli or the events [5]. A significant area of application of ERPs is for generating models that represent the brain activity when a subject is committing errors or observing other people’s errors. The reliable detection of correct/incorrect actions is the basis for the implementation of brain computer interface (BCI) systems that decode brain electrical activity into actions controlling devices that will assist the users of the system.

2. Material and Methods
The ERP data used in the present study were recorded in the study described in [6]. The data were acquired from sixteen (16) healthy volunteers (observers), who observed correct or incorrect actions of subjects (actors) performing. The actors were seated in front of a table facing an experimenter, having in front of them, on the table, two joystick devices positioned to the left and right of a LED stimulus device. The actors were asked to respond to the direction of a centre arrowhead surrounded by distracting flankers pointing either in the same direction as the centre arrow, or in opposite direction (figure 1). The brain electrical activity of the observers was recorded from 47 Ag/AgCl electrodes as well as vertical and horizontal electro-oculograms and was sampled with sampling rate 250 Hz. Electrodes were mounted in an elastic cap (Easy cap, Montage 10) configured for equal arrangement of the electrodes over the scalp (figure 2).

![Figure 1. Experimental setup.](image1)

![Figure 2. Electrode placement according to Easy Cap, Montage 10.](image2)

The electrode common was placed on the sternum. Ocular artifacts were corrected using the method described in [7]. The experimental session involved 8 runs of 100 trials of the task and the observations of correct and incorrect responses were averaged over a 800ms epoch (baseline [-100 , 0] ms before response) (figure 3). This procedure is necessary in order to discriminate the ERP signal from noise (brain activity that is not relevant to the task).
A time window, starting at -6ms and ending at 700ms (corresponding to 176 samples) after the response, was selected for analysis. A total of $32 \times 47 = 1504$ ERP recordings were available for analysis. From the available recordings, $16 \times 47 = 752$ recordings corresponded to observations of correct actions and the rest $16 \times 47 = 752$ recordings corresponded to observations of incorrect actions.

The quantitative characterisation of multichannel EEG recordings can be performed by means of the $(\Sigma$, $\Phi$, $\Omega)$ features [8]. The feature $\Sigma$ provides a measure of global signal strength. The feature $\Phi$ provides a measure of global frequency of signal changes. Finally, the feature $\Omega$ is a measure of spatial complexity. Let $K = 47$ be the number of electrodes, $N = 176$ be the number of samples per recording and $\Delta t = \frac{1}{250}$ s be the sampling period. If $u_k(n)$, $k = 1, 2, \ldots, K$ and $n = 0, 1, \ldots, N - 1$ is the recording from the electrode site $k$, then $K$-dimensional vectors can be defined as follows:

$$u(n) = [u_1(n), u_2(n), \ldots, u_K(n)]^T$$

Assuming that the vectors are detrended such that $\sum_{n=0}^{N-1} u(n) = 0$ and $\sum_{i=1}^{K} u_i(n) = 0$ for each $n$, then the $\Sigma$ feature is defined by the following equation [8]:

$$\Sigma = \sqrt{\frac{m_0}{K}}$$

where $m_0 = \frac{1}{N} \sum_{n=0}^{N-1} \|u(n)\|^2$. If $m_1 = \frac{1}{N} \sum_{n=0}^{N-1} \left\| \frac{\Delta u(n)}{\Delta t} \right\|^2$, where $\Delta u(n) = u(n) - u(n - 1)$, then the feature $\Phi$ is defined as follows [8]:

$$\Phi = \frac{1}{2\pi} \sqrt{\frac{m_1}{m_0}}$$

Furthermore, let $C = -\frac{1}{N} \sum_{n=0}^{N-1} u(n) u^T(n)$ be the $K \times K$ covariance matrix and $\lambda_1', \ldots, \lambda_K'$ be its eigenvalues. Then, the feature $\Omega$ is given by the following equation:

$$\Omega = -\sum_{k=1}^{K} \lambda_k' \log \lambda_k'$$
The next step is to find the best combination of electrodes which will be used for the calculation of the values of the features. The best combination of the electrodes can be obtained by optimizing the clustering performance (which will be defined below) of the fuzzy c-means (FCM) algorithm [9]. The FCM algorithm is an unsupervised clustering algorithm, which allows one feature vector to belong to two or more clusters according to the value of a membership function, which represents the fuzzy behaviour of this algorithm. Specifically, since 32 sets of recordings are available, for each set of recordings a vector of values can be computed resulting in 32 vectors. The FCM algorithm will group the 32 vectors in two clusters. Ideally, each cluster should contain vectors from one class only (observation of correct actions or observation of incorrect actions). In order to evaluate the clustering performance, CP, the following metric can be defined:

\[ CP = \max \{ cp_{11} + cp_{22} , cp_{12} + cp_{21} \} \]  

where \( cp_{ij} = \sum_{j=1}^{32} w_{ij} \) and \( w_{ij} \in [0,1] \) is the degree of membership of feature vector \( j (j = 1,2,\ldots,32) \) in cluster \( i (i=1,2) \). Thus, the objective is to find the combination of electrodes that will provide the best value of the clustering performance. This can be accomplished using GA optimization. Each individual of the GA population is a vector with 47 components. The components are either 0 (electrode not selected) or 1 (electrode selected). For each individual, the values of the 32 sets of recordings for the selected electrodes are calculated and the clustering performance is evaluated using (5). The GA will provide the individual with the best clustering performance.

The parameters of the GA algorithm are given in table1.

| Parameter                | Value       |
|--------------------------|-------------|
| Population size          | 200         |
| Number of generations    | 1000        |
| Selection process        | Tournament  |
| Mutation process         | Uniform     |
| Crossover process        | Scattered   |

The algorithm terminates if the predefined number of generations is completed or the weighted average relative change in the best value of the objective function over 150 generations is less than \( 10^{-6} \).

3. Results

The proposed methodology was developed in Matlab and applied on the available data resulting in the following electrodes (see figure 2): 8, 10, 11, 13, 14, 20, 23 and 48. The best value of the objective function was -27.08 (the problem was formulated as a minimization one). Sixteen (16) feature vectors corresponding to class 1 (observation of correct actions) were grouped into cluster 1 and 14 out of 16 feature vectors from class 2 (observation of incorrect actions) were grouped into cluster 2. Thus, the accuracy was 100% for cluster 1, 87.5% for cluster 2 and 93.8% in total.

Figure 4 presents the evolution of the best value and the average value of the objective function with respect to generation number. As can be seen, the best value is achieved quite early (generation 34).
4. Conclusion
In this paper, a GA based feature selection method for the characterization of ERP data that are collected during the observation of correct or incorrect actions was presented. The method finds the combination of electrodes that are used to extract the well-known $(\Sigma, \Phi, \Omega)$ features. Each combination of electrodes is evaluated using the clustering performance of the FCM algorithm. A total of eight electrodes were finally selected resulting in an at total accuracy of correct classification 93.8%.

Acknowledgements
The authors would like to thank Hein van Schie and Ellen de Bruijn from the Nijmegen Institute for Cognition and Information (NICI), The Netherlands, for kindly providing the data of their experiments and for their contribution to initial stages of the research. This research has been co-funded by the European Union (European Social Fund) and Greek national resources under the framework of the “Archimedes III: Funding of Research Groups in TEI of Athens” project of the “Education & Lifelong Learning” Operational Programme.

References
[1] Chandrashekar G and Sahin F 2014 Computers and Electrical Engineering 40 16-28
[2] Goldberg D 1989 Genetic Algorithms in Search, Optimization and Machine Learning (Boston, MA: Addison-Wesley)
[3] Sun Y, Babbs C and Delp E 2005 Conf. Proc. IEEE Eng. Med. Biol. Soc. 6 6532-5
[4] Yang J and Honavar V 1998 *IEEE Intell. Syst. Appl.* **13** 44–9
[5] Fabiani M, Gratton G and Coles M 2000 *Event-Related Potentials: Methods, Theory, and Applications* (New York: Cambridge University Press)
[6] van Schie H, Mars R B, Coles M G H and Bekkering H 2004 *Nat. Neurosci.* **7** 549-54
[7] Gratton G, Coles M G H and Donchin E 1983 *Electroencephalogr. Clin. Neurophysiol.* **55** 468-84
[8] Wackermann J 1999 *Int. J. of Psychophysiol.* **34** 65-80
[9] Bezdek J-C 1981 *Pattern Recognition with Fuzzy Objective Function Algorithms* (Norwell, MA: Plenum Press)