

# Toward a Solution for “Genetic Algorithm”-“Support Vector Machine” Combination to Have a Reliable P300-based BCI

**Gholamreza Salimi Khorshidi**  
School Of Cognitive Sciences, Institute For Studies In Theoretical Physics And Mathematics, Tehran, Iran  
[gsalimi@bme.aut.ac.ir](mailto:gsalimi@bme.aut.ac.ir)

**Ayyoub Jaafari**  
Faculty Of Biomedical Engineering, Amirkabir University Of Technology (Tehran Polytechnic), Tehran, Iran  
[ajafari@bme.aut.ac.ir](mailto:ajafari@bme.aut.ac.ir)

**Ali Motie Nasrabadi**  
Group Of Biomedical Engineering, Faculty Of Engineering, Shahed University, Tehran, Iran  
[nasrabadi@shahed.ac.ir](mailto:nasrabadi@shahed.ac.ir)

**Mohammadreza Hashemi Golpayegani**  
Faculty of biomedical engineering, Amirkabir University of technology, Tehran, Iran  
[hashemi\\_g@morva.net](mailto:hashemi_g@morva.net)

**Caro Lucas**  
School Of Cognitive Sciences, Institute For Studies In Theoretical Physics And Mathematics, Tehran, Iran  
And  
Control And Intelligent Processing Center Of Excellence, Electrical And Computer Engineering Department, University Of Tehran, Tehran, Iran  
[lucas@ipm.ir](mailto:lucas@ipm.ir)

## Abstract

In this study, we present a new framework to combine the Genetic Algorithm (GA) and Support Vector Machine (SVM) for an accurate P300 detection. P300 is the most important cognitive component used in a Brain Computer Interface (BCI) occurring around 300ms after stimulus onset. To detect this component and discriminate it from ongoing EEG, a typical Continuous Wavelet Transform (CWT) feature extraction is adopted in combination with GA and SVM for feature selection and feature classification, respectively. SVM is a powerful classifier that yields good classification results in its optimum point. In this study we use a new technique in GA to select the best feature subset (optimum feature set) and tune the SVM (for the selected feature subset) simultaneously, which showed a significant improvement in classification results.

**Keywords:** BCI, P300, CWT, SVM, GA.

## 1 General formatting instructions

A reliable brain-computer interface (BCI) can be a big step forward for rehabilitation and different clinical and non-clinical applications. There are different approaches in designing a BCI system from its biological input from the subject; like MEG and

EEG, among which EEG-based and especially ERP-based BCI systems are the most favorite ones from applicability, cost, time resolution and some other viewpoints [1]. The most common ERP component for a BCI system is P300 which is a positive peak occurring 300ms after stimulus onset. Compared to other well-known components, P300 has a robust and obvious waveform that makes it suitable for a typical BCI system. In a P300-based BCI system, the major task is to discriminate trials with P300 from those without it or to detect the P300 cognitive component. A typical P300 detector BCI, comprises following steps; EEG acquisition, EEG Preprocessing, Feature Extraction from EEG, Feature Selection and Classification. Using all or a subset of these blocks is a case dependent issue.

In this study, we used the BCI Competition 2003 EEG dataset for P300 speller paradigm (Ib dataset) and then used a CWT process for both preprocessing and feature extraction [2]. Because of their high resolution both in time and frequency domain, wavelet coefficients from the CWT process are good features for classification, especially for a P300 detection task. As a classifier, we adopted the SVM in three ways; Linear SVM (LSVM), SVM with Gaussian kernel (GSVM) and SVM with polynomial kernel (PSVM). For all these classifiers, GA is used to both select the optimum feature set and tune the SVM for its optimum point. To describe all these

issues, this paper is organized as follows. After a brief introduction, EEG dataset will be introduced and then the wavelet analysis procedure for feature extraction from EEG time series will be described. As the classifier, SVM, its parameters and their importance will be briefly introduced. Finally after describing the GA process, the results will be evaluated and discussed.

## 2 Materials and Methods

A typical P300 waveform can be extracted with a simple averaging technique. As shown in Fig.1, based on this technique, P300 waveform can have different waveforms (from its amplitude and shape viewpoints) in different scalp places which is because of different activities in different brain areas in response to each category of stimulation. The maximum peak and better discrimination between average waveforms can be seen in Cz and Fz channels while the worst discrimination is shown in Oz and Pz channels. Having this data, the next step is to extract the features from it to help the classifier discriminate between classes, easily.

In the present article, the EEG data is BCI Competition 2003 EEG dataset recorded from 64 scalp positions with P300 speller paradigm in which the subject was supposed to distinguish between two classes of stimulus; one class which induces P300 cognitive component and the other one which lacks this component. The aim of this competition is to detect the P300 component in scalp recorded EEG; which seems identical to our aim in P300 detection in this study. After data acquisition, there will be a 64-channel EEG data with number of trials equal to the number of stimuli. Each trial will contain 156 samples which are associated with a nearly 650ms time duration sampled with 256Hz.

## 3 Continuous Wavelet Transform

As a very useful and applicable method for EEG trials' analysis, wavelet transform is getting used in a wide range of data analyses in this field. In this study we used continuous wavelet transform (CWT) for extracting time-frequency characteristics of EEG trials as the features describing time series' dynamics. For this reason, we used four wavelets for a multi resolution analysis; two general purpose basis functions (Matlab's Bior3.9 and Db4) and two

special purpose wavelets (Quiroga's manually tuned wavelet for ERP analyses and Glassman's SNAP wavelet based on EEG's origin). In this method, each EEG trial will be decomposed into its frequency sub-bands and in each sub-band there will be some coefficients representing the time interval associated with it. This process is shown in Fig.2 for a sample trial with P300 and a sample trial without P300. These coefficients can be used as our features. Using all the scalp channels and all the extracted coefficients, it's time to select the best wavelet and the best features associated with that particular wavelet. Fig.3 shows the D-values (Eq.1) for some features extracted by using all four aforementioned wavelets (first 4000 features sorted by their D-value). As Fig.3 shows, Quiroga's manually tuned wavelet can make more statistically discriminant features than three other wavelets.

$$D = (\mu_1 - \mu_2) / (\sigma_1^2 + \sigma_2^2)^{0.5} \quad (1)$$

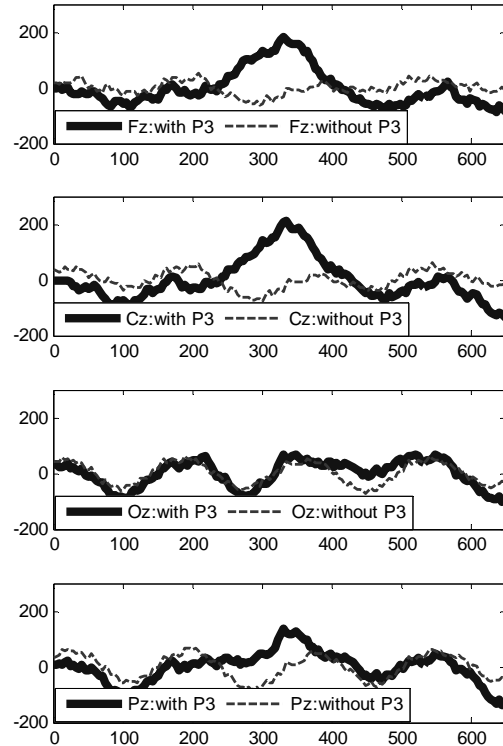


Fig. 1 Typical average waveform of trials with/without P300 over different scalp places which shows an obvious peak on central regions (Cz and Fz channels).

Having the result of Fig.3, we can now have a primary feature selection which will not yield the final optimum subset, but will result in a subset inside which we can have this optimum subset. So finally based on aforementioned D-value, we selected 150 most statistically discriminant features which will get fed into the “feature selection”-“classifier” block set.

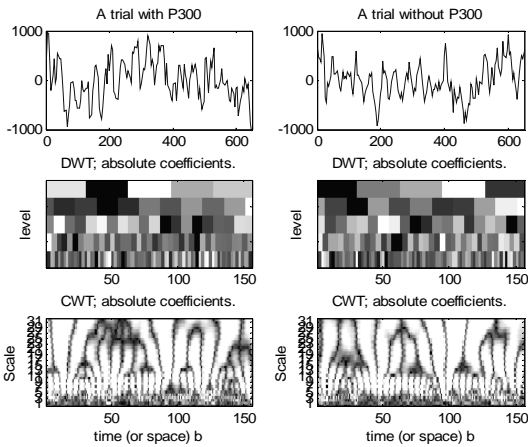


Fig. 2 A typical multi resolution analysis (MRA) on two sample EEG trials; one with P300 (a) and one without P300 (b) with Quiroga’s wavelet. Obviously, we can say that the coefficients are different in low frequencies and in times around 300 ms. The DWT is for scales 2, 4, 8, 16, 32 and CWT is for scales 1 to 32.

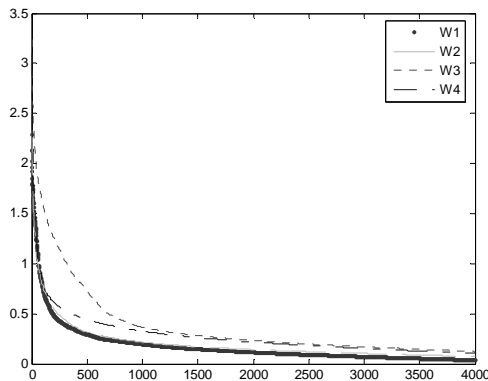


Fig. 3 Statistical discriminancy of 4000 best features (based on their D-value) for four wavelets; W1 (Db4), W2 (Bior3.9), W3 (Quiroga’s) and W4 (SNAP).

#### 4 “Genetic Algorithm”-“Support Vector Machine” Combination

The feature set resulting from primary statistical feature selection (using D-value) can not assure us to

have the best classification accuracy. It means using all of these features together will not guarantee the best accuracy and also maybe destroy the result compared to an optimum subset of it. For this reason we used genetic algorithm for a final feature selection which will yield the best feature subset in order to make the most accurate classification.

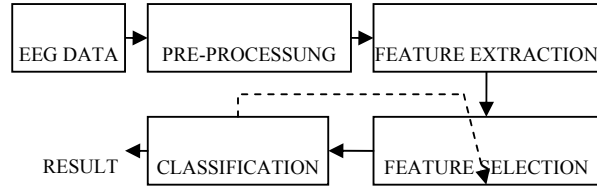


Fig.4 combination of GA and the SVM classifier in which the classification accuracy will affect the feature selection block’s output. Finally this interaction between classifier and the feature selection block (GA) will make the system to evolve and select the most discriminant feature set.

SVM classifiers have two important parameters to be tuned;  $C$  and  $\sigma$  (standard deviation of the Gaussian kernel). In the easiest approach to tune these parameters, we can use a k-fold cross validation over a specific range of these parameters and finally find the best point from the classification accuracy viewpoint. This approach is not the best way of finding the optimum values for  $C$  (trade off coefficient) and  $\sigma$  (Gaussian kernel’s variance) in an evolutionary process. Because, after some generations, the resulting feature set is not the same as the one in the first generations. It means, while using GA for feature selection, the input space is changing over generation from the dimension viewpoint which makes it require a new discriminant hyperplane for the optimum classification.

For this reason, we used another approach for determining the best point of using GA-SVM combination from feature set and classifier viewpoint. In a binary GA for feature selection (the classic viewpoint), the chromosome length is equal to the number of features (150 bits for this problem). Finally, bits with 1 value represent the selected features while those with 0 values represent the rejected features. To make the SVM tuning process an evolutionary process, we added 20 new bits to the end of chromosomes which were tuning-specific bits; 10 bits for  $C$  and 10 bits for  $\sigma$ . The decimal equivalent of these 10 bits divided by 10 will

represent the final value for each parameter. So the chromosomes will be of length 170 from which 150 bits are representing the selected features and 20 bits are representing the optimally tuned values for C and  $\sigma$  parameters (Fig.5).

150 bits (feature selection)	10 bits (C)	10 bits ( $\sigma$ )
------------------------------	-------------	----------------------

Fig.5 a typical chromosome used for simultaneous feature selection and SVM tuning. As this figure shows, this chromosome comprises three parts; Feature Selection, C tuning and  $\sigma$  tuning bits.

There are some other important parameters in the GA that are supposed to get determined to have a better and more effective evolution; like mutation rate, crossover rate, initial population size and the number of generations to reach the end of evolution. For this reason, we used some tests over our dataset and selected the optimum parameters (as follows).

*Crossover and Mutation Rate (c, m):* to tune these two parameters, we selected four values for them; (0.3, 0.3), (0.5, 0.5), (0.7, 0.3) and (0.9, 0.1) for population size of 100, elite number of 10 and 2000 generations. The result is shown in Fig.6.

As the result in Fig.6 shows, increasing the crossover rate can make the mating more effective and faster for reaching its optimum point. The problem with having only the crossover is the lack of new evolutionary actions in the final generations as shown in Fig.6-A. In this figure, the fitness has a flat area after reaching a fairly good value and will remain unchanged. To evaluate the effect of mutation rate, we selected the values as in Fig.6-B in which the mutation rate is equal to 50% which shows a jump for better fitness at the middle of the evolution (after the flat area). This means that the mutation rate can show its effect in the final generation when the crossover can not make any new evolution. The other important point about Fig.6-B is that increasing the crossover rate can make the final fitness better. To confirm this idea, we increased the crossover rate to 0.9 with a small mutation rate which results the Fig.6-D in which there's no evolutionary change in the last generations (the small mutation rate) but reaches its optimum point very fast. So we can say having a big crossover can help us have a good evolution.

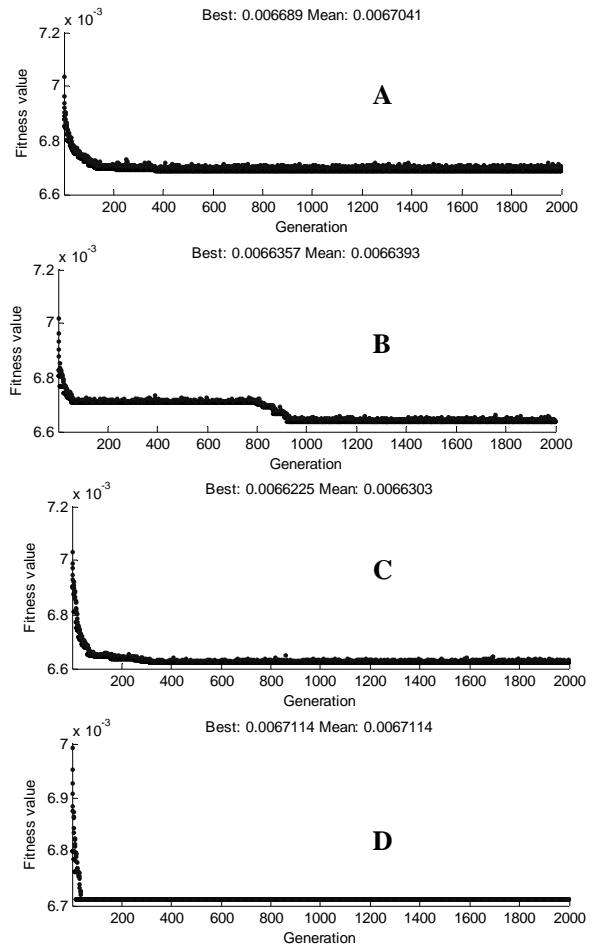


Fig.6 the best fitness value vs. the number of generations in the GA for different (c, m) values; A.(0.3, 0.3), B.(0.5, 0.5), C.(0.7, 0.3) and D.(0.9, 0.1).

So we finally selected the crossover rate and mutation rate the way in which we have both effects very well (a good evolution and effective mating with some new evolution in flat area of the fitness) which results the Fig.6-C. In this situation we have the best evolution in which we have all the things at the best point.

*Population Size:* the variety is another important issue to make better generations which is affected by the number of individuals in the initial population size. In this study, we tested this factor for two values of 50 and 100 which shows the better evolution for 100 individuals which confirms our initial idea (Fig.7).

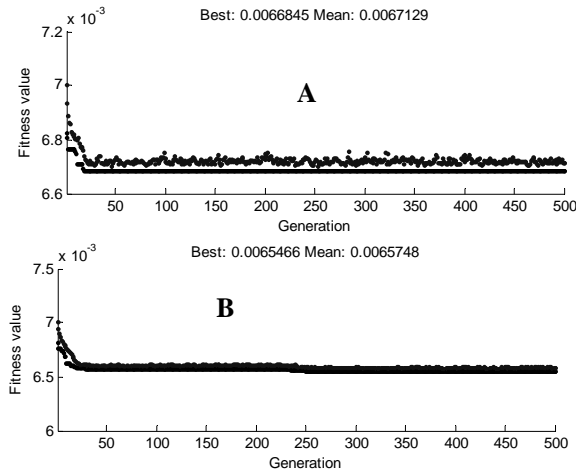


Fig.7 shows the effect of initial population size for two values of A. 50 individuals and B. 100 individuals. As the results show the bigger the initial population size, the better and the more effective the evolution is held here.

Having all these results together, we used the GA for both feature selection and SVM tuning with crossover rate of 0.7, mutation rate of 0.3, initial population size of 100 individuals, 1000 generations, binary chromosomes (individuals) with the length of 170 bits. The results of this process is described and discussed in the next section.

## 5 Results and Conclusion

To evaluate our method over real EEG data for a BCI application, we selected 2500 trials of aforementioned dataset. Having the EEG trials, we extracted the continuous time-frequency features from the data. Then selected 150 most statistically discriminant features for a simpler computation. Then GA was used with aforementioned parameters in combination with SVM classifier. Two different methods were used for SVM tuning; 1- evolutionary SVM tuning and 2- using a 10-fold cross validation for SVM tuning. For the first method this length was 170 and for the second method, the length of chromosome was 150.

Data was divided into two groups; training dataset (2000 trials: 1000 trials with P300 and 1000 trials without P300) and test dataset (500 trials: 250 trials with P300 and 250 trials without P300). Finally the classification task was done and evaluated. The result

of this classification is shown in Table1. Since the same process can be done for different classes of SVM, we tested the method for linear SVM (LSVM), SVM with Gaussian kernel (GSVM) and SVM with polynomial kernel (PSVM). For PSVM, the chromosome length is different, because the range of  $\sigma$  is different from polynomial's degree. So we added 13 bits to the end of chromosomes (163 bits to tune the PSVM).

Table1. The result of classification for different methods of combining the SVM and GA

Method	Accuracy on train	Accuracy on test
LSVM+ First Method	76.8%	74.9%
PSVM+ First Method	81.8%	82.7%
GSVM+ First Method	84.6%	80.5%
PSVM+ Second Method	87.1%	81.9%
GSVM+ Second Method	91.2%	88.7%

As the result show, using the SVM tuning in a simultaneous manner with feature selection can yield a good result compared to the first method in which the SVM is tuned separated from the feature selection process. As mentioned before, this is as a result of tuning the SVM for the specific feature subset selected at the final level.

## References

- [1] B. Blankertz, K.-R. Müller, G. Curio, T. M. Vaughan, G. Schalk, J. R. Wolpaw, A. Schlögl, C. Neuper, G. Pfurtscheller, T. Hinterberger, M. Schröder, and N. Birbaumer, "The BCI Competition 2003: Progress and perspectives in detection and discrimination of EEG single trials", in *IEEE Trans. Biomed. Eng.*, vol. 51, pp. 1044–1051, June 2004.
- [2] *Handbook of Electroencephalographic Clinical Neurophysiology*, vol. 3, "Endogenous Event-Related Potentials", T. W. Picton and S. A. Hillyard, Eds., Elsevier, Amsterdam, the Netherlands, 1988, pp. 361–426.
- [3] N. Birbaumer, N. Ghanayim, T. Hinterberger, I. Iversen, B. Kotchoubey, A. Kübler, J. Perelmouter, E. Taub, and H. Flor, "A spelling device for the paralyzed," *Nature*, vol. 398, pp. 297–298, 1999.
- [4] Emanuel Donchin, Kevin M. Spencer, and Ranjith Wijesinghe, "The Mental Prosthesis: Assessing the Speed of a P300-Based Brain-Computer Interface," *IEEE Transactions on*

- Rehabilitation Engineering, VOL. 8, NO. 2,  
Page(s):174 – 179, JUNE 2000.
- [5] M. Kaper, P. Meinicke, U. Grossekhoefer, T. Lingner, and H. Ritter, "BCI competition 2003 - data set Iib: Support vector machines for the p300 speller paradigm," IEEE Transactions on Biomedical Engineering, VOL. 51, NO. 6, Page(s): 1073-1076, June, 2004.
- [6] Vladimir Bostanov, "BCI Competition 2003— Data Sets Ib and Iib: Feature Extraction From Event-Related Brain Potentials with the Continuous Wavelet Transform and the t-Value Scalogram," IEEE Transactions on Biomedical engineering, VOL. 51, NO. 6, Page(s): 1057-1061, June 2004.
- [7] Elena L. Glassman, "A Wavelet-Like Filter Based on Neuron Action Potentials for Analysis of Human Scalp Electroencephalographs", IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. 52, pp.1851-1862, 2005.
- [8] R. Quian Quiroga , O.W. Sakowitz , E. Basar , M. Schurmann "Wavelet Transform in the analysis of the frequency composition of Evoked potentials" Brain Research Protocols vol.8, pp.16–24, 2001.
- [9] Deon Garrett, David A. Peterson, Charles W. Anderson, Michael H. Thaut, "Comparison of Linear, Nonlinear and Feature Selection Methods for EEG signal Classification", IEEE transaction on neural systems and rehabilitation engineering, Vol.11, No.2, Page(s):141-152, June 2003.
- [10] A. Schwaighofer. SVM Toolbox for Matlab. [Online]. Available: <http://www.cis.tugraz.at/igi/aschwaig/software.html>
- [11] N. Cristianini and B. Schölkopf, "Support vector machines and Kernel methods: The new generation of learning machines," *AI Mag.*, vol. 23, no. 3, pp. 31–41, 483, 484, 2002.
- [12] C.-W. Hsu, C.-C. Chang, and C.-J. Lin. (2003) A Practical Guide to Support Vector Classification. National Taiwan Univ., Dept. Comput. Sci Inf. Eng.. [Online]. Available: <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>
- [13] Randy L. Haupt, Sue Ellen Haupt, *Practical genetic algorithm*, Wiley-Interscience; 2 edition May 31, 2004
- [14] Schroder, M. Bogdan, M. Hinterberger, T. Birbaumer, N. "Automated EEG feature selection for brain computer interfaces", pp. 626-629, 20-22 March 2003.