2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

Single Channel Single Trial P300 Detection Using Extreme Learning Machine

Compared with BPNN and SVM

Songyun Xie, You Wu, Yunpeng Zhang, Juanli Zhang, Chang Liu
School of Electronics and Information
Northwestern Polytechnical University
Xi'an, China

***Abstract*—A Brain Computer Interface (BCI) is a communication system designed to allow the users to directly interact with external devices using their minds without using any muscle activities. P300, a component of Event Related Potentials (ERPs), is a widely used feature component of EEG signal for BCI applications. However, single trial analysis is difficult since ERPs such as P300 signals have a very low signal to noise ratio, which bring down the communication rate. And the numerous number of channels needed to record EEG prevents the popularization of BCI applications due to the complexity and high cost of the system. In this paper, a new efficient method, extreme learning machine (ELM), is presented to detect P300 components using a single channel data from a visual stimuli Oddball paradigm experiment. It reaches an average accuracy above 85% and performs better than BPNN and SVM.**

***Keywords*—single channel EEG; ERP; P300; extreme learning machine (ELM)**

I. INTRODUCTION

The electroencephalogram (EEG) is defined as recording of the brain's electrical activity which usually can be recorded by electrodes from the surface of the scalp. Event-related potentials (ERPs), which have wide usage in research purposes and clinical diagnostic, are one of important EEG, and psycho-physiological correlates of neuro-cognitive functions that reflect the responses of the brain to the external or internal environment of the organism changes (events). Because of the high temporal resolution, low cost and ease of use (compared with the other acquisition techniques, such as fMRI, EMG), EEG signals have been used widely in brain computer interface (BCI), a communication system designed to allow the users directly interact with external environment that does not depend on the brain's normal output pathways of peripheral nerves and muscles [1]. So far, Many BCI systems have been introduced with their own application, among which many are Event Related Potentials (ERPs) based. P300 speller is one of the most common BCI systems, which was introduced by Farwell and Donchin and attracts a lot of attention [2]. It uses a positive evoked potential called P300, a robust positive ERP component which appears after a visual or auditory stimulus with a latency of about 300 milliseconds and is often used as an indicator for target task in an oddball paradigm [3].

However, there are still some problems which should be addressed in these systems. Since ERPs such as P300 signals have a very low Signal to Noise Ratio (SNR), it is a common method that synchronously averaging over many trials, which effectively diminishes the random noise. But this is not a practical method in BCI applications because it is pretty slow and meanwhile reduces the communication rate greatly. Many investigators tried to overcome this problem by reducing the number of averaged trials or going toward the single trial detection [4]-[6]. Nevertheless, the efficiency of single trial P300 detection still needs to be improved. Another major problem in many BCI systems is the large number of channels needed to record EEG signals in order to have a reliable system, which prevents the popularization of BCI applications since multi-channels recording will cause complexity of the system and high cost [7]. It is of great help to reduce the record channels.

Currently, Back Propagation Neural Network (BPNN) and Support vector machine (SVM) are the most succeeded and commonly used algorithms to classify the EEG data to detect P300 component. These two method need to adjust learning parameters iteratively. Extreme Learning Machine (ELM) is a new efficient tuning-free algorithm to train single-hidden-layer feedforward neural networks (SLFNs), proposed by Huang [8]. This new method has been used in many fields, such as image processing and neural information processing [9]-[12]. Usually, ELM shows similar or better results than SVM, and needs much less time to train networks than BPNN[13], [14].

In this paper, we tried to detect P300 over single trial using just one single channel to record EEG data. ELM was used to process the data from a visual stimuli oddball paradigm experiment. Channel and parameters selection were studied in detail. And the performance of ELM is

This work was supported by the graduate starting seed fund of Northwestern Polytechnical University (Z2014138) and the National Natural Science Foundation of China (61273250).

compared with BPNN and SVM in terms of training time and testing classification accuracy.

II. EXPERIMENTAL SETUP

A. Experiment in Oddball Paradigm

To evoke P300, a visual stimuli Oddball paradigm was implemented, in which the participants saw a randomized Bernoulli series of letter “X” and “O”. In this oddball task, letter “X” was presented as the task-irrelevant standard stimulus with a probability of 80% (240 out of all 300 trials), whereas the letter “O” was the target stimulus with a probability of 20% (60 out of all 300 trials). All the stimuli were located in the center of a black background, with Courier New font, white color, bold and 40 pt size. We used “+” as the fixation, with the same configuration as the stimuli except for the size (34 pt). The letters subtended 0.8 degrees visual angle in height and width. E-Prime 2.0 was used for stimulus presentation and response collection.

During the experiment, the subjects were seated on a comfortable chair in front of a monitor which was used for commands and stimuli presentation. An SRBox was used to input subjects’ response. The schema of the experiment is illustrated in Fig. 1. In each trial, after a fixation period of 800-1200ms, the stimulus was presented for 1000ms followed by a 1000ms black screen, repeatedly.

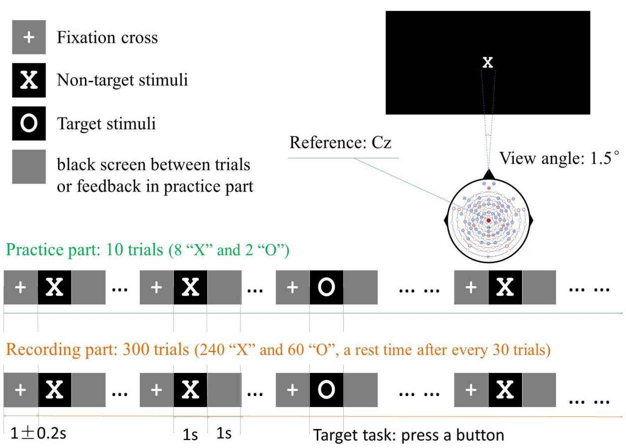

Fig. 1. Schema of the experiment. For presentation purposes the fixation cross and black screen are slightly faded in this figure. However, in the real experiments they all were black background.

To make the subjects getting familiar with the experiments, we had a practice part in the beginning of the experiment, during which eight standard stimuli “X” and two target stimuli “O” were presented randomly. Then it went to the 300 recording trials and the subjects would have a rest time as long as they want after every 30 trials. The subjects had to put one hand on the SRBox during the experiment and press a button when the target stimulus “O” appeared. They were asked to keep their eyes fixed on the fixation cross “+”.

B. Data Acquisition

The recording room was shielded with a Faraday cage. The EEG signals were acquired at 200Hz, filtered between 0.3 and 100Hz using Net Amps 300 (EGI product) and saved to a computer for off-line processing. A GSN (Geodesic sensor net) with 64 sensors was used (referenced at Cz, showed in Fig. 1) and the Impedance of all electrodes was kept below 50 .

Data were collected from nine healthy right-handed undergraduate or graduate subjects (seven male, two female; age:23±3years) with normal or corrected to normal vision. One male subject’s data were excluded because of too many artifacts. The eight subjects are named as S1, S2 to S8 in the following text. All the processing was performed in Matlab and some figures were depicted by Net Station.

III. DATA PROCESSING

A. Extreme Learning Machine

Unlike Neural networks (NN) and support vector machines (SVM), which need to adjust learning parameters iteratively, ELM is a tuning-free algorithm by randomly generating the input weights and the hidden bias before seeing the trainning data. And the hidden node parameters are not only independent of trainning data but also of each other.

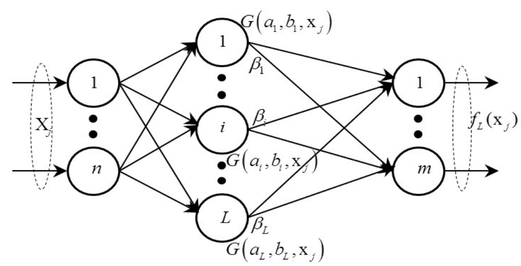

Fig. 2. Single-hidden-layer feedforward neural networks (SLFNs). ELM is a new efficient learning method to train the SLFNs by generating the input weights and the hidden bias randomly and computing a set of linear equations.

First step of ELM is making a SLFN which has L hidden neurons and randomly generate the input weights and the hidden bias $(a_i, b_i), i = 1, \cdots, L$. Assume we have N train data $(\mathrm{x}_j, \mathrm{t}_j), j = 1, \cdots, N$, where $\mathrm{x}_j = [x_1^j, x_2^j, \cdots, x_n^j]^T \in R^n$ is input vector and its target output value is $\mathrm{t}_j = [t_1^j, t_2^j, \cdots, t_m^j]^T \in R^m$. Trainning the SLFN (showed in Fig. 2) to learn the map between the input and output patterns is mathmaticaly to solve the following set of equations:

$$f_L(\mathrm{x}_j) = \sum_{i=1}^{L} \beta_i G\left(a_i, b_i, \mathrm{x}_j\right) = \mathrm{t}_j ; j = 1, ..., N . \quad (1)$$

where $\beta_i = [\beta_1^i, \beta_2^i, \cdots, \beta_m^i]^T \in R^{\mathrm{m}}$ are the output weights from the hidden unit to the output unit. And $G\left(a_i, b_i, \mathrm{x}_j\right) = g(a_i \mathrm{x}_j + b_i)$ where g denotes the nonlinear activitation function of the hidden node. It can be the sigmoid, sine, hardlim, tribas or radial function and choosed before training. The problem becomes a set of linear equations once the output of the hidden units is fixed and can be solved by minimum square error estimation. Equation (1) can be written in matrix form as: $\mathrm{H}\beta = \mathrm{T}$.

$$\mathrm{H}=\begin{vmatrix} G(a_1,b_1,\mathrm{x}_1) & \cdots & G(a_L,b_L,\mathrm{x}_1) \\ \vdots & \ddots & \vdots \\ G(a_1,b_1,\mathrm{x}_N) & \cdots & G(a_L,b_L,\mathrm{x}_N) \end{vmatrix}_{N\times L}, \quad \beta=\begin{vmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{vmatrix}_{L\times m}, \quad \mathrm{T}=\begin{vmatrix} \mathrm{t}_1^T \\ \vdots \\ \mathrm{t}_N^T \end{vmatrix}_{N\times m} \tag{2}$$

H is called the hidden layer output matrix of the network and T is the target matrix of output layer. After calculate the output matrix of the hidden layer H, We can solving this linear system by computing the least-squares estimation $\hat{\beta}=\mathrm{H}^{+}\mathrm{T}$, where H^{+} is the moore-penrose inverse of H. The ELM training algorithm is summarized in Table I.

TABLE I. ELM TRAINNING ALGORITHM

➢	inputs parameters : Number of Hidden Neurons L, activation function, train data $(\mathrm{x}_j,\mathrm{t}_j), j=1,\cdots,N$.
1)	Generate the hidden node parameters $(a_i,b_i), i=1,\cdots,L$ randomly;
2)	Calculate the hidden layer output matrix H;
3)	Calculate the output weights $\hat{\beta}=\mathrm{H}^{+}\mathrm{T}$;

B. Data preprocessing

After band pass filtering with a range of 1-15Hz, the EEG was partly denoised. Then it was segmented by a time window of $[-100, 700)$ ms, 160 samples, post-stimulus to extract all 300 trials waveforms, which were classified in two categories (240 Non-target, 60 Target). Baseline of each segment was corrected by subtracting mean of 100-0 ms prior to stimulus onset.

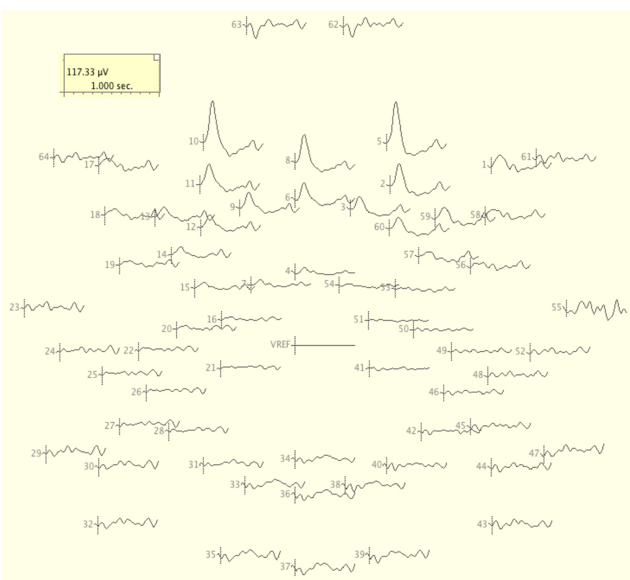

Fig. 3. One trial marked as "ARTIFACTS" at 64 channels. This figure shows eye blink artifacts from one subject.

All the trials whose amplitude (max-min) exceeding 140 μV were suspicious for eye blink artifacts, eye movement artifacts or EMG related artifacts and marked as "ARTIFACTS" for further analysis. The others were marked as "CLEAN", which remains approximately 88–98% of trials across participants. Fig. 3 shows one trial segment marked as "Artifacts".

C. Channel selection

One important part of this research is selecting a best recording channel, which performs good detection accuracy and suits to different individuals. First, we calculated the time-domain grand average ERPs over all 8 subjects. Fig. 4 depicts grand average ERPs at all 64 channels from data "CLEAN" after the preprocessing. The blue waveforms are ERPs of task-irrelevant standard stimuli "X" whereas the red ones are ERPs of target stimuli "O". The P300 components of ERPs from target tasks were apparently enhanced than those from the non-target ones. And the topology maps showed on the top left corner of the Fig.4 present the voltage at the time of 349ms after stimulus when the ERPs went around the highest P3. The right one is for the target stimuli and the left one is for standard stimuli). The voltages at the occipital and parietal region for target tasks were much higher than those for the non-target ones. Here we chose 15 sensors on the occipital and parietal region to be studied as recording channels, because signals from those on the frontal field of scalp were apparently interfered by eye artifacts (showed in Fig. 3). These channels were CH26, CH28, CH31, CH32, CH33, CH34, CH35, CH36, CH37, CH38, CH39, CH40, CH42, CH43 and CH46.

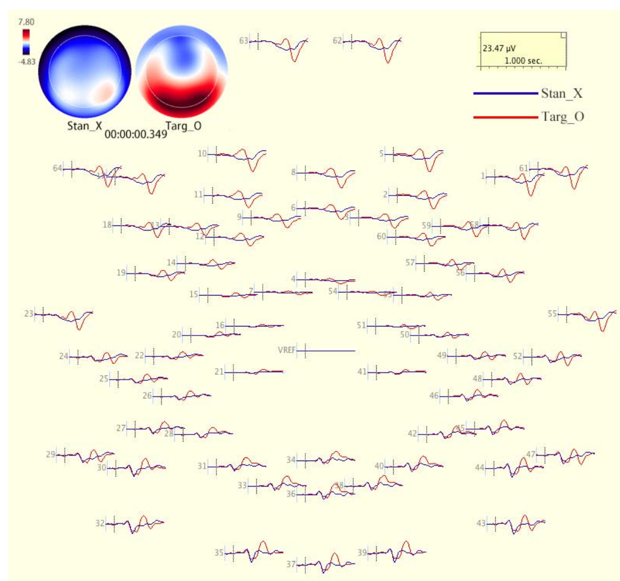

Fig. 4. Grand average ERPs at all 64 channels from data "CLEAN". Average ERPs were calculated over all 8 subjects. Red lines and blue lines show targets and non-targets respectively. And the topology maps show the ERPs voltage at time of 349ms post-stimulus (right one for the target stimuli, left one for standard stimuli).

Then the final channel was selected according to the accuracy of classification. The single trial EEG data from each single channel were classified using extreme learning machine to detect whether P300 component were contained. The influence of artifacts on single channel P300 detection was also studied in this part by comparing the results of classifying the "CLEAN" dataset and the "ALL" dataset (which contained "CLEAN" and "ARTIFACTS" trials). In this binary classification problem, target and standard stimuli were labeled to 1 and 0 respectively. The ratio of training set to testing set was 1:1. In the ELM-SLFNs, according to prior test, we used 3000 hidden neurons and chose sigmoid function as the activation function (which will be studied in detail in the next section).

After training the SLFN using the training set, the EEG data from testing set were classified and testing accuracy was calculated. All the testing accuracies in this paper were over 30 runs averagely. Fig. 5 shows the testing accuracies of the every channel from each subject. "CLEAN" accuracies are marked with small crosses "x" whereas dashed lines depict the "ALL" accuracies in plot style. The red big crosses "x" and bold solid line represent the averages of the "CLEAN" and "ALL" accuracies over all subjects. From the results between "CLEAN" and "ALL" data, which showed similar accuracies, we found little difference was caused by the artifacts through all subjects. This showed the robust of the method, so the next processing was on the "ALL" data. CH39 (O2) performed best average accuracy of "CLEAN" dataset at 86.30%. Top five average accuracies of the "ALL" dataset were showed by CH33, CH35 (O1), CH36, CH37 (Oz) and CH39 (O2) at 86.08%, 86.12%, 85.90%, 85.73%and 85.89%. CH35 was selected as the recording sensor because not only the highest accuracy in average but also more distinction between two tasks it showed in the topology map in Fig.4 and more robust efficiencies through different individuals.

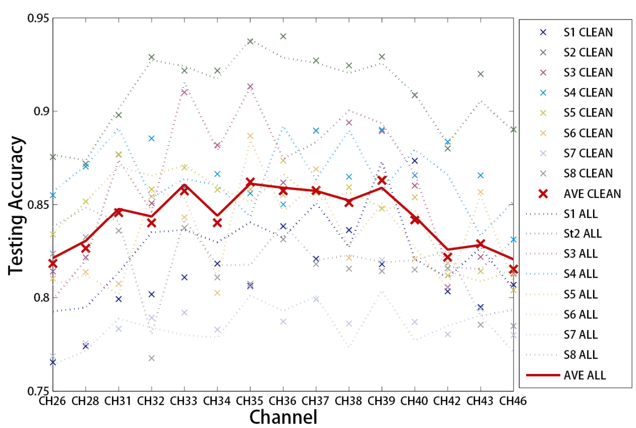

Fig. 5. Testing accuracy of each channel using ELM. The crosses "x" and the dashed lines show the accuracy of "CLEAN" data and "ALL" data (CLEAN and ARTIFACTS) at each channel from each subject. And the red big "x" and solid bold line shows the average of the "CLEAN" and "ALL" accuracy over all subjects.

D. *ELM Parameter Selection*

For classification single trial (160 samples) P300s, the sigmoid and hardlim function were best which showed similar high accuracy around 85% when the number of hidden neurons was 3000, whereas triangular basis and radial basis function performed around 72%, and sine function did worst around 50%. The activation function was set as sigmoid after comparing.

Another important parameter, number of hidden neurons, was selected according to both testing accuracy and time required by raining and testing. Empirically, number of hidden neurons should be larger than number of input neurons. Here we chose numbers from 200 to 4000 with the step of 10 to study. The Fig. 6a depicts the average testing accuracy over 30 runs against different number of hidden neurons. The increase of validation accuracy was significant until the number of hidden neurons in the SLFN was over 1500 and keeps slowly ascendant after it. Meanwhile, the training time using 150 segments and testing time of 150 segments (Fig. 6b and Fig. 6c) were increased linearly with the growth of the number. 2500 was chosen as the number of hidden neurons when accuracy and time were taken into account.

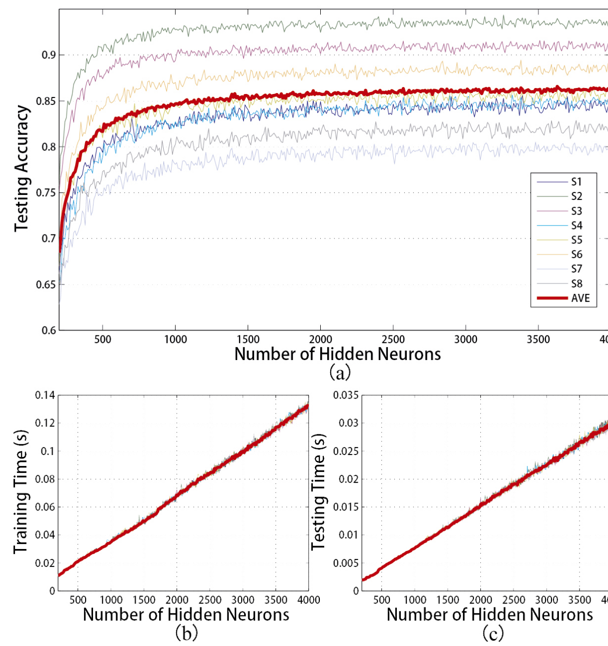

Fig. 6. Testing accuracy, training time and testing time against number of hidden neurons, from 200 to 4000 with the step of 10. All the data were averagely calculated over 30runs. Red bold lines show the average values over all subjects.

E. *Single channel-Single trial P300 Detection*

After setting all the parameters, the processing of ELM based single channel-single trail detection was confirmed. The schema is summarized in Fig. 7. The results of the proposed detection method are shown in the next section.

BPNN and SVM were used to evaluate the performance of ELM. Before performance comparison, the parameters of BPNN and SVM were pre-estimated to achieve their best generalization performance. The testing accuracy of BPNN increased slightly when more hidden neurons were used, but this also requires two or three orders of magnitude more training time. Five was chosen as the number of hidden neurons in BPNN. Linear kernel, Gaussian kernel, Multilayer

Perceptron kernel and polynomial kernel were test for SVM. The order 3 polynomial kernel showed the best results.

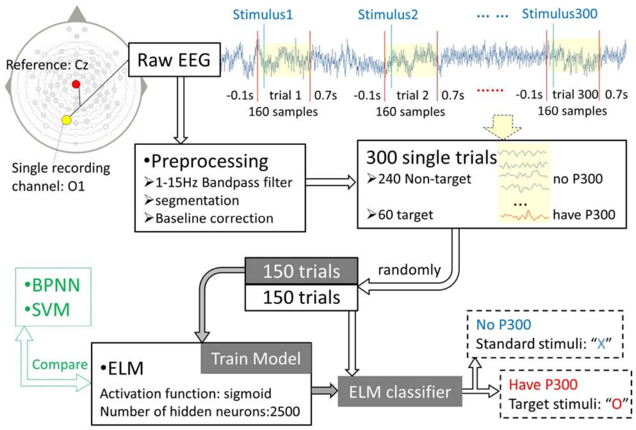

Fig. 7. The schematic diagram of the whole method

IV. Results

The performance of the method proposed in this paper to detect single trial P300 using just one recording channel was evaluated by the testing accuracy. The testing accuracy and model training time for each subject using different classification method are listed in Table II. "NH" stands for number of hidden nodes and polynomial is the kernel function used in SVM. The numbers attached after the accuracy are standard deviations between different runs whereas the average's standard deviations were computed between different subjects. Testing time is not listed because the limited space and all testing parts were done very quickly. Fig. 8 depicts the comparison of accuracies.

TABLE II. Testng Accuracy (%) and Training Time (S)

Subjects	ELM *HN=2500*	BP *HN=5*	SVM *'polynomial'*
1	84.11±2.00 *(0.0876s)*	79.67±3.36 *(1.2269s)*	84.07±1.70 *(0.0286s)*
2	93.29±1.59 *(0.0822s)*	86.49±4.35 *(1.2598s)*	90.22±1.73 *(0.0194s)*
3	91.18±1.88 *(0.0900s)*	84.71±4.22 *(1.3096s)*	86.24±2.08 *(0.0222s)*
4	83.91±2.62 *(0.0845s)*	79.76±5.89 *(1.1682s)*	79.82±3.08 *(0.0416s)*
5	84.87±1.65 *(0.0884s)*	79.00±3.81 *(1.1615s)*	82.67±1.89 *(0.0283s)*
6	88.69±1.71 *(0.0870s)*	82.71±3.04 *(1.2643s)*	84.60±1.74 *(0.0272s)*
7	78.53±2.40 *(0.0870s)*	72.27±8.30 *(1.1745s)*	80.16±2.02 *(0.0326s)*
8	81.16±2.37 *(0.0880s)*	75.49±5.36 *(1.1808s)*	80.16±2.23 *(0.0260s)*
Average	85.72±4.68 *(0.0869s)*	80.01±4.37 *(1.2182s)*	83.52±3.34 *(0.0282s)*

The experiment results show that extreme learning machine successfully detects P300 component using single channel (O1) EEG, and achieve an average accuracy above 85%. Two subjects' testing accuracy using ELM exceeded 90%.

ELM with 2500 hidden neurons performed 5% better than BPNN and using only 7% training time of BPNN to train its classifier model. When the number of hidden neurons in ELM was set as 700, ELM reached similar accuracy to SVM and cost same time in training processing. It cost more time to train ELM with 2500 hidden neurons than SVM, but ELM performed 2% more accuracy in average.

V. Conclusions

In this paper, we study using only one channel to record signals effectively and use extreme learning machines to detect the single trail P300 using just single recording channel. The results showed that this method can efficiently detect the P300 component and performs better than BPNN and SVM.

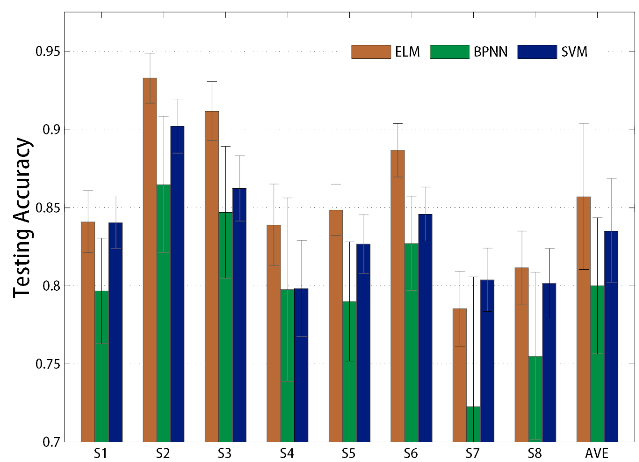

Fig. 8. Comparison of different classification method. The standard deviations between different runs were showed on the top of the bars of 8 subjects. However, the average accuracies were calculated over all 8 subjects and three standard deviations were computed between different subjects.

References

[1] H. Cecotti, "Spelling with non-invasive Brain–Computer Interfaces – Current and future trends," *Journal of Physiology-Paris,* vol. 105, pp. 106-114, 2011-01-01 2011.

[2] E. Yin, Z. Zhou, J. Jiang, F. Chen, Y. Liu, and D. Hu, "A Speedy Hybrid BCI Spelling Approach Combining P300 and SSVEP," in *Biomedical Engineering, IEEE Transactions on*. vol. PP, 2013, p. 1-11.

[3] Y. T. Qassim, T. R. H. Cutmore, D. A. James, and D. D. Rowlands, "Wavelet coherence of EEG signals for a visual oddball task," *Computers in Biology and Medicine,* vol. 43, pp. 23-31, 2013-01-01 2013.

[4] Y. Shahriari and A. Erfanian, "Improving the performance of P300-based brain–computer interface through subspace-based filtering," *Neurocomputing,* vol. 121, pp. 434-441, 2013-12-09 2013.

[5] F. Ghaderi, S. K. Kim and E. A. Kirchner, "Effects of eye artifact removal methods on single trial P300 detection, a comparative study," *Journal of Neuroscience Methods,* vol. 221, pp. 41-47, 2014-01-15 2014.

[6] N. Haghighatpanah, R. Amirfattahi, V. Abootalebi, and B. Nazari, "A single channel-single trial P300 detection algorithm," in *Electrical Engineering (ICEE), 2013 21st Iranian Conference on*, Mashhad, 2013, pp. 1-5.

[7] M. Arvaneh, G. Cuntai, K. A. Kai, and Q. Chai, "Robust EEG channel selection across sessions in brain-computer interface involving stroke patients," in *Neural Networks (IJCNN), The 2012 International Joint Conference on*, Brisbane, QLD, 2012, pp. 1-6.

[8] H. Guang-Bin, Z. Qin-Yu and S. Chee-Kheong, "Extreme learning machine: a new learning scheme of feedforward neural networks," in *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*, 2004, pp. 985-990 vol.2.

[9] L. Shi and B. Lu, "EEG-based vigilance estimation using extreme learning machines," *Neurocomputing,* vol. 102, pp. 135-143, 2013-02-15 2013.

[10] C. Chi, T. Wee-Peng and H. Guang-Bin, "Extreme learning machines for intrusion detection," in *Neural Networks (IJCNN), The 2012 International Joint Conference on*, Brisbane, QLD, 2012, pp. 1-8.

[11] Y. Kai, E. Y. Du, E. J. Delp, J. Pingge, J. Feng, C. Yaobin, R. Sherony, and H. Takahashi, "An Extreme Learning Machine-based pedestrian detection method," in *Intelligent Vehicles Symposium (IV), 2013 IEEE*, Gold Coast, QLD, 2013, pp. 1404-1409.

[12] W. Li, D. Wang and T. Chai, "Burning state recognition of rotary kiln using ELMs with heterogeneous features," *Neurocomputing,* vol. 102, pp. 144-153, 2013-02-15 2013.

[13] J. Chorowski, J. Wang and J. M. Zurada, "Review and performance comparison of SVM- and ELM- based classifiers," *Neurocomputing,* 2013.

[14] H. Zhong, C. Miao, Z. Shen, and Y. Feng, "Comparing the Learning Effectiveness of BP, ELM, I-ELM, and SVM for Corporate Credit Ratings," *Neurocomputing,* 2013.