A New Method to Improve the Performance of Deep Neural Networks in Detecting P300 Signals: Optimizing Curvature of Error Surface Using Genetic Algorithm

Seyed Vahab Shojaedini[1]*[ORCID], Sajedeh Morabbi[2], Mohamad Reza Keyvanpour[3]

[1]PhD, Associate professor in Biomedical Engineering, Department of Biomedical Engineering, Iranian Research Organization for Science and Technology, Tehran, Iran
[2]MSc, Department of Computer Engineering, Alzahra University, Tehran, Iran
[3]PhD, Associate professor in Computer Engineering, Department of Computer Engineering, Alzahra University, Tehran, Iran

ABSTRACT

Background: Deep neural networks have been widely used in detection of P300 signal in Brain Machine Interface (BMI) systems which are rely on Event-Related Potentials (ERPs) (i.e. P300 signals). Such networks have high curvature variation in their error surface hampering their favorable performance. Therefore, the variations in curvature of the error surface must be minimized to improve the performance of these networks in detecting P300 signals.

Objective: The aim of this paper is to introduce a method for minimizing the curvature of the error surface during training Convolutional Neural Network (CNN). The curvature variation of the error surface is highly dependent on model parameters of deep neural network; therefore, we try to minimize this curvature by optimizing the model parameters.

Material and Methods: In this experimental study an attempt is made to tune the CNN parameters affecting the curvature of its error surface in order to obtain the best possible learning. For achieving this goal, Genetic Algorithm is utilized to optimize the above parameters in order to minimize the curvature variations.

Results: The performance of the proposed algorithm was evaluated on EPFL dataset. The obtained results demonstrated that the proposed method detected the P300 signals with maximally 98.91% classification accuracy and 98.54% True Positive Ratio (TPR).

Conclusion: The obtained results showed that using genetic algorithm for minimizing curvature of the error surface in CNN increased its accuracy in parallel with decreasing the variance of the results. Consequently, it may be concluded that the proposed method has considerable potential to be used as P300 detection module in BMI applications.

Introduction

BMI (Brain-Machine Interface) system allows people who are unable to communicate with control devices using their EEG (Electroencephalogram) signals. A vital part of this procedure is detecting P300 signal from evoked human brain potentials, which is

*Corresponding author:
Seyed Vahab Shojaedini
Department of Biomedical Engineering, Iranian Research Organization for Science and Technology, Tehran, Iran
E-mail: shojadini@irost.ir

a significant factor for establishing the BMIs. P300 is a type of ERP (Event-Related Potential) signals, which has a positive deflection occurring in the recorded EEG and typically elicited approximately 300 ms after the presentation of an infrequent stimulus [1]. Furthermore, P300 signals are widely used in other applications such as lie detection and diagnosis of neurological disease [2-4].

Farwell and Donchin were the first researchers who employed P300 as a control signal in BMI. They introduced an "oddball" paradigm to evoke P300 signals. The oddball is a square matrix containing letters of the alphabet and other symbols; they are displayed on a computer screen. Rows and columns of the matrix are flashed in random order and the person under test (i.e. subject) is asked to concentrate mentally on some target characters through counting them. When a character is flashed in a certain row and column, a P300 signal is evoked automatically and appeared on EEG signal. Then it may be detected by an appropriate method for further operations [5].

The main challenges in P300 detection are variability and low Signal to Noise Ratio (SNR); therefore, several methods have been introduced to improve distinguishing P300 from other parts of EEG signal [3, 6]. Averaging is a simple method, which obtains the higher detection rate by increasing the SNR, but it reduces the bit rate and deforms the ERP waveform [7].

Moreover, linear and nonlinear methods for BMI usage have been developed. For instance, Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), K-Nearest Neighbors (KNN) and Bayesian have already been used for P300 signal detection [2-4, 8-13].

The weakness of linear methods obstacles solving complicated problems and nonlinear methods deals with over fitting in real world problems. Recently several machine-learning methods have been applied to raise SNR without loss of any significant information of P300

signal [14]. Most of the machine learning techniques applied for P300 detection have been based on Artificial Neural Networks (ANN) [15, 16]. Classic ANNs are not strong enough to escape from local minimums impressing their performances in distinguishing P300 from other parts of EEG signal. Recently, in the case of P300 detection Deep Neural Networks (DNNs) have been presented in which the deep structure and multiple level data representation are utilized as their basic potentials [4, 8].

Nevertheless, such networks have large-scale dimensions and high curvature variations, which lead to higher process volume in parallel with undesirable convergence [17, 18]. Minimizing curvature variation results in faster and more accurate convergence; thus, it may be considered as an important issue in P300 detection by DNNs.

Several techniques have been presented to address the high curvature variation problem, including first and second order optimization algorithms. First, order algorithms are to improve detecting P300 signal by the use of gradient information of objective function that are simple to be used and converged fast. Second, order algorithms compute Hessian matrix, which is highly dependent on dimensions of objective function. As the dimensions grow, the required memory also increases. In the field of learning DNNs, local minimums have values in same order of the global optimum. Thus, finding a local minimum is good enough to address high curvature variation problem [19].

Convolutional Neural Networks (CNNs) are the member of DNN family, widely applied in P300 detection [8, 20]. The CNN is the semantic of the weights once the network is trained and robust to variability of P300. The receptive convolution kernel may be easily interpreted, therefore can provide a diagnosis about the type of high-level features to be detected [8].

In this paper, a new method is introduced to

improve training of DNNs, which has been based on minimizing curvature variation in their error function. In proposed method, the evolutionary paradigm is used to decrease the level of curvature variation in high dimensional space. The proposed method is applied on CNN to obtain an improved network to distinguish P300 and non-P300 components.

Material and Methods

In this experimental study, at first, the dataset and pre-processing steps are briefly demonstrated. These items are effectively utilized in our experimental section. Then, the proposed scheme is described completely to minimize curvature variation in training CNN.

Dataset and Pre-processing

In this paper, EPFL BCI dataset is used. It has been captured using visual stimulation by the Biosemi system with 32 electrodes located according to standard 10-20 international system position at 2048 HZ. The EPFL BCI dataset consist of eight subjects. The first four subjects are disabled and out of them are able-bodies. The data of each subject is composed of four sessions. Each of the sessions consists of six stimulating patterns (i.e. runs), which is displayed in format of a six-cell matrix. The stimulating patterns were flashed at random order, which each of them lasts for 100 ms and then during 300 ms none of the patterns is flashed. The inter stimulus interval is 400 ms and more details about the EPFL dataset may be found in [5].

After recording EEG signals, firstly referencing was performed to remove the reflection of the reference electrode in captured signals. Afterward, to clean the signals from additional noises, a 6th order forward-backward Butterworth band pass filter at 1.0 and 12 HZ cut off frequency was utilized. The resultant filtered signals were down-sampled from 2048 to 32 HZ and single trials of duration 1000 ms were extracted. Finally, the EEG signals were mapped to the range of (0, 1) as normalizing process to reduce the computational complexity [5].

Proposed Scheme

The first important part of CNN, which is called convolutional layer is responsible for extracting fundamental features of input EEG signals. This part is a set of filter banks, applied to raw EEG signal to extracts features. Furthermore, the nonlinear activation function is applied on neurons, then the pooling, which is also known as a down-sampling layer applied. Another fundamental part of CNN (i.e. fully connected layer) is accountable to classify the data based on the extracted features. In this article Adadelta algorithm is used in this layer to perform classification.

Suppose Adadelta weights introduced as θ and $H(\theta)$ demonstrates the error (i.e. cost function), which may be minimized to obtain the best classification performance. Gradient descent procedure, the use of the gradient information of the error function, may be used to update weights as bellow:

$$E[\nabla\theta^2]_t = \tau.E[\nabla\theta^2]_{t-1} + (1-\tau)\nabla\theta_t^2 \qquad (1)$$

In which $E[\nabla\theta^2]_t$ shows running average over the gradient of the squared weight, which only depends on previous average and the current gradient [21] and the parameter τ refers to momentum. The derivative of $H(\theta):\mathbb{R}\to\mathbb{R}$, may be written as:

$$H'(\theta) = V(\theta)(\theta-\theta^*) \qquad (2)$$

In which $V(\theta)\in\mathbb{R}$ is generalized curvature [22, 23] and θ^* is a global minimum of $H(\theta)$. Let denote $1-\vartheta.V(\theta_t)$ as contraction of a gradient descent step, N_t as model parameter operator at time t and ϑ as learning rate; thus, the update rule may be defined as bellow:

$$\begin{pmatrix} \theta_{t+1} - \theta^* \\ \theta_t - \theta^* \end{pmatrix} = \begin{bmatrix} 1 - \vartheta.V(\theta_t) + \tau & -\tau \\ 1 & 0 \end{bmatrix} \cdot \begin{pmatrix} \theta_t - \theta^* \\ \theta_{t-1} - \theta^* \end{pmatrix} (3)$$
$$\triangleq N_t \begin{pmatrix} \theta_t - \theta^* \\ \theta_{t-1} - \theta^* \end{pmatrix}$$

If the model parameters ϑ and τ and the generalized curvature V are in the robust region, we have:

$$\left(1 - \sqrt{\tau}\right)^2 \leq \vartheta.V(\theta_t) \leq \left(1 + \sqrt{\tau}\right)^2 \qquad (4)$$

Therefore, the spectral radius of the τ only depends on $f(N_t) = \sqrt{\tau}$. Optimization is done using a quadratic model which may be considered as:

$$H(\theta) = \frac{1}{2}k\theta^2 + R = \frac{1}{m}\sum_i \frac{1}{2}k(\theta - r_i)^2 \triangleq \frac{1}{m}\sum_i H_i(\theta), \sum_i r_i = 0 \ (5)$$

The average of m component functions H_i is a common gradient based objective function $\nabla H_t(\theta)$ at time t. In equation 5, the gradient variance is defined as $R = \frac{1}{2m}\sum_i r_i^2$.

As $H(\theta)$ is introduced at equation 5, and $\theta_1 = \theta_0$ and θ_t follow the model parameter update rule with stochastic gradients $\nabla H_t(\theta_{t-1})$ for $t \geq 2$. The squared distance to the optimum θ^* is:

$$\mathbb{E}\left(\theta_{t+1} - \theta^*\right)^2 = \left(s_1^T N^t \left[\theta_1 - \theta^*, \theta_0 - \theta^*\right]^T\right)^2 + \vartheta^2 R s_1^T (I - P^t)(I - P)^{-1} s_1 \ (6)$$

In which the first and second terms refer to squared bias and variance, and their corresponding model parameter dynamics are defined as bellow:

$$N = \begin{bmatrix} 1 - \vartheta V + \tau & -\tau \\ 1 & 0 \end{bmatrix}, P = \begin{bmatrix} (1 - \vartheta V + \tau)^2 & \tau^2 & -2\tau(1 - \vartheta V + \tau) \\ 1 & 0 & 0 \\ 1 - \vartheta V + \tau & 0 & -\tau \end{bmatrix} (7)$$

A scalar and asymptotic surrogate based on the spectral radius of operators is used to simplify the problem [24].

$$\mathbb{E}\left(\theta_t - \theta^*\right)^2 \approx f(N)^{2t}\left(\theta_0 - \theta^*\right)^2 + (1 - f(P)^t)\frac{\vartheta^2 R}{1 - f(P)} \ (8)$$

The spectral radius of the variance operator, P is τ. As described in equation 4, under the

exact same condition, the variance operator, P has spectral radius τ, if $\left(1 - \sqrt{\tau}\right)^2 \leq \vartheta.V \leq \left(1 + \sqrt{\tau}\right)^2$. The new form of equation 8, in robust region then may be written as [24]:

$$\mathbb{E}\left(\theta_t - \theta^*\right)^2 \approx \tau^t \left(\theta_0 - \theta^*\right)^2 + (1 - \tau^t)\frac{\vartheta^2 R}{1 - \tau} \qquad (9)$$

Equation 9 is used to design rule of the proposed adaptive optimization scheme, which leads to equation 10, in which B, refers to the estimated distance between the current model and a local quadratic approximation's minimum, W denotes the estimate for gradient variance. Furthermore, V_{max} and V_{min} refer to the maximum and minimum generalized curvature. They are used to estimate the variation of curvature as a fitness function in the proposed method as bellow:

$$\begin{cases} H(\tau) = \arg\min_\tau \tau B^2 + \vartheta^2 W \\ \\ \tau \geq \left(\dfrac{\sqrt{V_{max}/V_{min}} - 1}{\sqrt{V_{max}/V_{min}} + 1}\right)^2 \\ \\ \vartheta = \dfrac{(1 - \sqrt{\tau})^2}{V_{min}} \end{cases} , \qquad (10)$$

Now, the genetic algorithm is used to minimize the surrogate for the expected squared distance from the optimum of above local quadratic approximation [25]. In this minimization problem, we have a non-linear objective function H subject to equality constraint and inequality constraint. The pseudo code of genetic algorithm is presented in Figure 1.

Results

A system with an Intel Core i7 and 16 GB RAM was used to test the proposed method. The proposed method has been implemented using MATLAB 2017 package and applied on EPFL dataset as described in section 2. For

```
Initial values: Pop_Size=10, Max_Gen=20,
Convergence=false. Counter=1
    a.    Generate Pop_Size individuals randomly
    b.    while ((Counter<=Max_Gen) && !(Convergence))
    c.    Evaluate the fitness of all individuals.
    d.    Select the best individual based on fitness value
    e.    Apply crossover on the selected individuals
    f.    Apply mutation over the results of step e
    g.    if (the optimum value found)
    h.    Convergence=true
    i.    End of if
    j.    Counter=Counter+1
    k.    End of while
```

Figure 1: Pseudocode of genetic minimizer

preparing this data set, the international 10-20 protocol has been used to position 32 electrodes on scalp.

The EEG signals were recorded using 32 electrodes from 8 subjects, which each of them consists of four sessions and each session contains six runs (i.e. exciting symbols). Thus, in each subject, the EEG data from two sessions was used to train and one another applied for validation. Eventually, the last session was applied for test. This process was repeated four times. Finally, the average of four steps was evaluated and reported, separately for each of subjects.

The EEG data belonging to each subject was first pre-processed to make it ready to feed up to the CNN. Thereafter, P300 signal was distinguished using CNNs, which had been trained by two versions of Adadelta; the first was Adadelta with naive model parameter called as NMP in the rest of paper and another was Adadelta with optimum genetic model parameter called OGMP for brevity. Finally, their performances were estimated to determine how good the examined methods detect P300 signal. The important parameters used in implementation of the genetic module of the proposed method may be depicted in Table 1.

The effectiveness of the examined methods was compared using some standard param-

eters, including: True Positive Ratio (TPR), False Positive Ratio (FPR) and classification accuracy, used commonly in P300 detection paradigm [8]. Table 2 shows how the mentioned parameters were used as evaluation criterions to compare proposed method and its

Table 1: The value of parameters in Optimum Genetic Model Parameter (OGMP) algorithm.

Parameter	Value/Description
Maximum Generation	20
Population Size	10
Crossover Function	Scattered
Mutation Function	Gaussian
Selection Function	Stochastic Uniform

Table 2: The average of parameters over subjects.

Method	Subject	TPR	FPR	Accuracy
NMP	1	70.74	8.86	80.93
	2	67.45	5.35	81.05
	3	77.18	2.14	87.51
	4	68.39	3.24	82.57
	5	76	5.14	83.41
	6	72.61	3.22	84.69
	7	78.85	2.23	88.31
	8	61.79	2.75	79.52
OGMP	1	95.48	0.68	97.40
	2	94.27	0.92	96.67
	3	98.44	0.62	98.91
	4	98.10	2.43	97.83
	5	93.35	0.41	96.46
	6	97.53	0.54	98.49
	7	97.72	0.25	98.73
	8	98.54	0.86	98.83

TPR: True Positive Ratio, FPR: False Positive Ratio, NMP: Naive Model Parameter, OGMP: Optimum Genetic Model Parameter

alternative. The comparisons are performed in two distinct scenarios as described below.

Intra-Subject Scenario

In this type of analysis, the best results obtained from the examined methods were compared out of the subject under test. According to TPRs, the proposed method (i.e. OGMP) was outperformed in all of the subjects in such way that the best TPR gained by this method has been 98.54%, over subject 8. This value has been 19.69% better than the best value for alternative's TPR, occurred in subject 7.

In similar manner, the FPR value also demonstrated the superiority of the proposed algorithm against NMP method. By investigating the obtained FPRs, it is observed that the proposed scheme has achieved FPR equal to 0.25 percent over subject 7, which has been the best among all obtained FPRs and also 1.89 percent lower than the best of alternative obtained over subject 3.

Finally, the classification accuracy confirmed better performance of our proposed method in such way that the best accuracy of this method has been obtained over subject 3 to extent of 98.91%. This accuracy has been 10.6% better compared to the best accuracy, obtained using NMP method (i.e. 88.31% over subject 7).

Inter-Subject Scenario

In another type of analysis, the performances of the proposed and NMP methods were compared in the same subjects. By investigating TPRs, it was observed that, the highest superiority of the proposed method against NMP method has occurred in subject 8 by extent of 36.75 percent. The lowest superiority of the proposed method against NMP method was observed in subject 5 by extent of 17.35 percent. Exploiting such superiority in the rest of subjects led to the moderate value of 24.83 percent for inter-subject superiority of our method (i.e. OGMP) against alternative based on true detections.
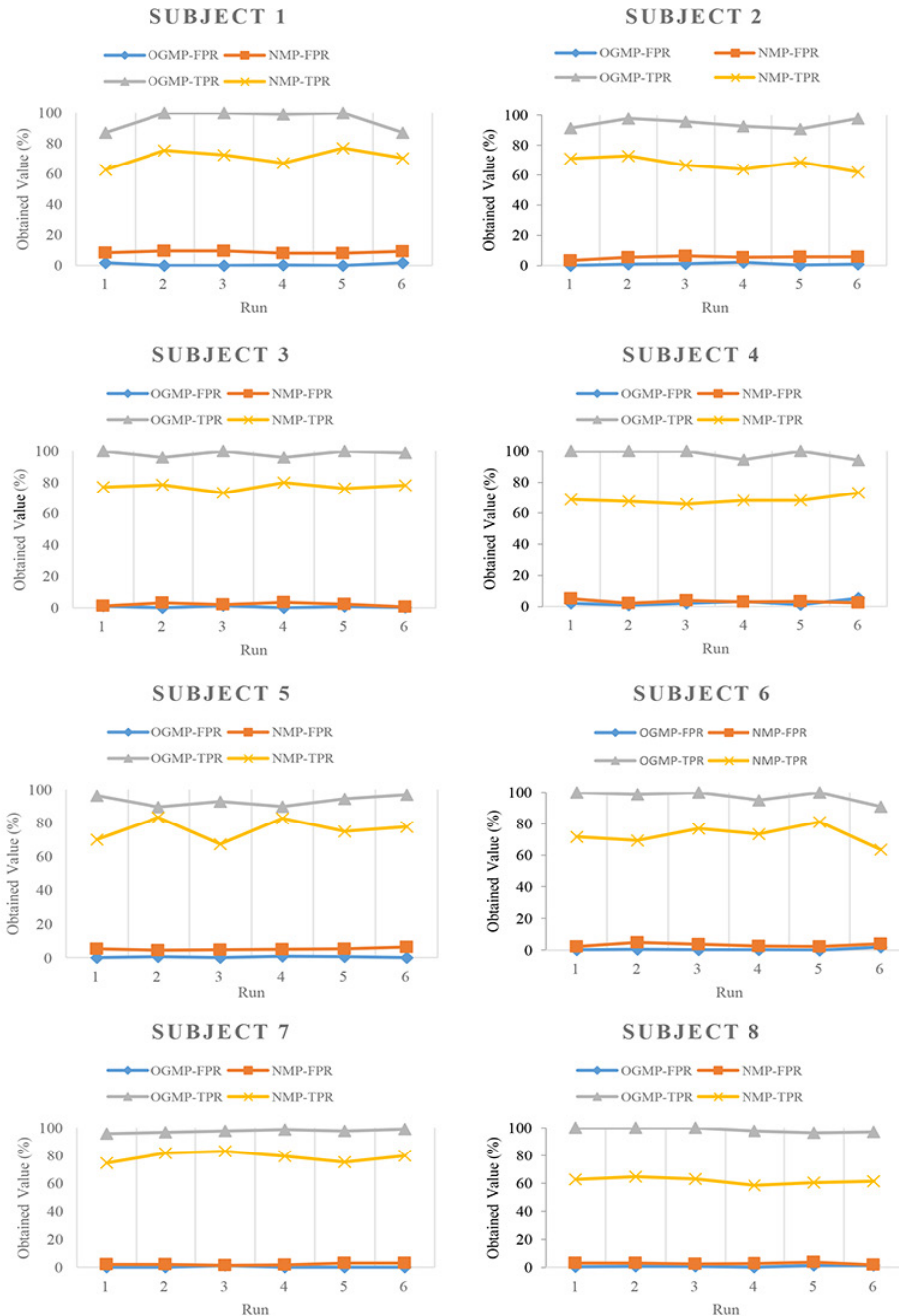
In the same way, based on FPRs, the superiorities of the proposed method against alternative were obtained in the range of (0.81- 8.18) percent which arose from subjects 4 and 1, respectively. By investigation of such superiority for other subjects, the value of 2.33 percent was obtained as moderate superiority of proposed against NMP methods in detecting false signals.

By considering the obtained accuracies, the minimum and maximum of the mentioned superiorities were obtained as 10.42% and 19.31%, obtained in subjects 7 and 8, respectively. Therefore, the moderate superiority of our method against NMP was obtained as 14.53 percent.

Discussion

The EEG signals were recorded from subjects who watched certain and predefined pictures as exciting symbols on a laptop screen during four sessions. Each session included six runs, which each of runs was representative of a predefined picture. The averaged EEG signal of each run was evaluated based on TPR and FPR as depicted in Figure 2. Exploiting the trend of TPRs obtained from our proposed method in graphs shows that in subject 1, some significant growths have occurred in runs 2, 3 and 5 (i.e. TPR= 100%).

Furthermore, the lowest value of FPRs was 0%, obtained in runs 2, 3 and 5. In similar manner, the highest and lowest growths in TPR and FPR belonging to subject 2, have been 97.82% (run 2) and 0% (run 1), respectively. A similar trend may be observed in subjects 3, 4, 6 and 8, in all of them, the best upward of TPRs has been 100 percent. In subjects 5 and 7, the TPRs were 97 and 98.95 percent, respectively. Similarly, the best downwards of FPRs in subjects 3, 5, 6, 7 and 8 have also been 0 percent and for subject 4 was 0.83 percent. The obtained results also depict that the proposed method is more stable than its alternative by exploiting changes of TPRs against different exciting

Figure 2: The comparison of the sensitivity of true positive ratio (TPR) and false positive ratio (FPR) of examined methods to exciting symbols.

symbols (i.e. runs). It may be observed that in subjects 2, 3, 5, 6, 7 and 8 over different runs, the NMP method gained some TPRs, which has varied in the range of (62.02%- 72.91%), (73.3%- 79.83%), (67.29%- 83.3%), (63.44%-

81.29%), (74.36%- 83.04%) and (58.34%-64.58%), respectively.

By contrast, according to the proposed method, the trends of TPRs in the same subjects have been in the smaller ranges as (90.68%-

97.82%), (95.83%- 100%), (89.61%- 97%), (91%- 100%), (95.95%- 98.95%) and (96.42%- 100%). The above results may be summarized in such manner that the maximum variances of TPRs in the proposed and NMP methods have been evaluated as 13.05% and 17.85% over subjects 1 and 6, respectively.

Moreover, the minimum variances of TPRs, obtained from proposed method and its alternative were 3% in subject 7 and 6.24% in subject 8, respectively. These results show the more robust behavior of detection rate against changes of exciting symbols (i.e. runs) in the proposed method compared to NMP method. Furthermore, the investigation of variances in FPRs obtained by examined methods showed that the considerable superiority in TPR robustness in the proposed method not only has not increased false detections, but also has a little decreased in false positive rates.

Conclusion

This article presented a new method to improve detection of P300 signals using deep neural networks. The proposed scheme tries to address the curvature variation in large-scale deep networks by evolutionary optimization of their model parameters. The genetic algorithm was used to find the best model parameters of convolutional neural networks, which led to minimize the curvature variations of the error function. Such a minimization may improve the performance of detecting P300 component from other parts of EEG signals.

To evaluate the performance of the proposed algorithm (i.e. OGMP), it was examined on EEG data in parallel with existing NMP method and the obtained results were compared using their effective parameters including TPR, FPR and accuracy. The comparisons showed superiority of the proposed method against its alternative in such way it has distinguished P300 component from other parts of EEG signal, 19.69% and 24.83% better than NMP method using intra-subject and inter-subject analysis, respectively. These superiorities were obtained as 1.89% and 2.33% by investigating the obtained FPR parameters. Furthermore, the accuracy of the proposed method has been better than its alternative by extents of 10.6% and 14.53% by using intra-subject and inter-subject analysis, respectively.

In another type of analysis, it was investigated that the performance of the proposed method is more robust against different stimulating patterns (i.e. runs) than NMP method. The detection rates obtained from the proposed method, due to several runs, showed the variances approximately half of those values obtained for NMP. Based on the above analyzes, it may be concluded that the proposed method has considerable potential to be used as P300 detection module in BMI applications.

Acknowledgment

Conflict of Interest

None

References

1. Alvarado-Gonzalez M, Garduno E, Bribiesca E, Yanez-Suarez O, Medina-Banuelos V. P300 Detection Based on EEG Shape Features. *Comput Math Methods Med*. 2016;**2016**:2029791. doi: 10.1155/2016/2029791. PubMed PMID: 26881010. PubMed PMCID: PMC4736976.

2. Chowdhury A, Raza H, Meena YK, Dutta A, Prasad G. Online covariate shift detection based adaptive brain-computer interface to trigger hand exoskeleton feedback for neuro-rehabilitation. *IEEE Transactions on Cognitive and Developmental Systems*. 2017. doi: 10.1109/TCDS.2017.2787040.

3. Mubeen MA, Knuth KH. Evidence-Based Filters for Signal Detection: Application to Evoked Brain Responses. *ArXiv preprint arXiv:11071257*. 2011.

4. Zhang J, Yan C, Gong X, editors. Deep convolutional neural network for decoding mo-

tor imagery based brain computer interface. 2017 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC); Xiamen, China: IEEE; 2017. p. 1-5. doi: 10.1109/icspcc.2017.8242581.

5. Hoffmann U, Vesin JM, Ebrahimi T, Diserens K. An efficient P300-based brain-computer interface for disabled subjects. *J Neurosci Methods*. 2008;**167**:115-25. doi: 10.1016/j.jneumeth.2007.03.005. PubMed PMID: 17445904.

6. Changoluisa V, Varona P, Rodriguez FB, editors. How to Reduce Classification Error in ERP-Based BCI: Maximum Relative Areas as a Feature for P300 Detection. International Work-Conference on Artificial Neural Networks; Springer; 2017. p. 486-97. doi: 10.1007/978-3-319-59147-6_42.

7. Vareka L, Mautner P, editors. Using the Windowed means paradigm for single trial P300 detection. 38th International Conference on Telecommunications and Signal Processing (TSP); Prague, Czech Republic: IEEE; 2015. p. 1-4. doi: 10.1109/tsp.2015.7296414.

8. Cecotti H, Graser A. Convolutional neural networks for P300 detection with application to brain-computer interfaces. *IEEE Trans Pattern Anal Mach Intell*. 2011;**33**:433-45. doi: 10.1109/TPAMI.2010.125. PubMed PMID: 20567055.

9. Lazăr AM, Ursulean R, editors. The P300 event-related potential detection-A morphological approach. E-Health and Bioengineering Conference (EHB); Iasi, Romania: IEEE; 2013. doi: 10.1109/ehb.2013.6707413.

10. Li Y, Guan C, Li H, Chin Z. A self-training semi-supervised SVM algorithm and its application in an EEG-based brain computer interface speller system. *Pattern Recognition Letters*. 2008;**29**:1285-94. doi: 10.1016/j.patrec.2008.01.030.

11. Sharma N. Single-trial P300 Classification using PCA with LDA, QDA and Neural Networks. *ArXiv preprint arXiv:171201977*. 2017.

12. Cecotti H. Toward shift invariant detection of event-related potentials in non-invasive brain-computer interface. *Pattern Recognition Letters*. 2015;**66**:127-34. doi: 10.1016/j.patrec.2015.01.015.

13. Kindermans PJ, Verstraeten D, Schrauwen B. A bayesian model for exploiting application constraints to enable unsupervised training of a P300-based BCI. *PLoS One*. 2012;**7**:e33758. doi: 10.1371/journal.pone.0033758. PubMed PMID: 22496763. PubMed PMCID: PMC3319551.

14. Chen S-W, Lai Y-C. A signal-processing-based technique for P300 evoked potential detection with the applications into automated character recognition. *EURASIP J Adv Signal Process*. 2014;**2014**:152. doi: 10.1186/1687-6180-2014-152.

15. Kottaimalai R, Rajasekaran MP, Selvam V, Kannapiran B, editors. EEG signal classification using principal component analysis with neural network in brain computer interface applications. IEEE International Conference ON Emerging Trends in Computing, Communication and Nanotechnology (ICECCN); Tirunelveli, India: IEEE; 2013. p. 227-31.

16. Cecotti H, Gräser A. Neural network pruning for feature selection-Application to a P300 Brain-Computer Interface. *ESANN*. 2009:473-8.

17. Iyer RK, Jegelka S, Bilmes JA. Curvature and optimal algorithms for learning and minimizing submodular functions. *Adv Neural Inf Process Syst*. 2013:2742-50.

18. Dauphin YN, Pascanu R, Gulcehre C, Cho K, Ganguli S, Bengio Y. Identifying and attacking the saddle point problem in high-dimensional non-convex optimization. *Adv Neural Inf Process Syst*. 2014:2933-41.

19. Goodfellow I, Bengio Y, Courville A. Deep learning. Cambridge: MIT press; 2016.

20. Lawhern VJ, Solon AJ, Waytowich NR, Gordon SM, Hung CP, Lance BJ. EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces. *J Neural Eng*. 2018;**15**:056013. doi: 10.1088/1741-2552/aace8c. PubMed PMID: 29932424.

21. Zeiler MD. ADADELTA: an adaptive learning rate method. *ArXiv preprint arXiv:12125701*. 2012.

22. Zhang H, Sra S. First-order methods for geo-

desically convex optimization. Conference on Learning Theory; New York, USA: Columbia University; 2016. p. 1617-38.

23. Nesterov Y. Introductory lectures on convex optimization: A basic course. New York: Springer Science & Business Media; 2013.

24. Schaul T, Zhang S, LeCun Y. No more pesky learning rates. *International Conference on Machine Learning*. 2013:343-51.

25. Wah BW, Chen Y-X. Constrained genetic algorithms and their applications in nonlinear constrained optimization. Evolutionary Optimization. New York: Springer; 2003. p. 253-75.