JHMI
Journal of Health Management and Informatics

Original Article

Temporal Convolutional Learning: A New Sequence-based Structure to Promote the Performance of Convolutional Neural Networks in Recognizing P300 Signals

Mahnaz Mardi[1], Mohamad Reza Keyvanpour[1*], Seyed Vahab Shojaedini[2]

[1]Department of Computer Engineering, Faculty of Engineering, Alzahra University, Tehran, Iran
[2]Department of Biomedical Engineering, Iranian Research Organization for Science and Technology, Tehran, Iran

Abstract

Distinguishing P300 signals from other components of the EEG is one of the most challenging issues in Brain Computer Interface (BCI) applications, and machine learning methods have vastly been utilized as effective tools to perform such separation. Although in recent years deep neural networks have significantly improved the quality of the above detection, the significant similarity between P300 and other components of EEG in parallel with their unrepeatable nature have led to P300 detection, which are still an open problem in BCI domain. In this study, a novel architecture is proposed in order to detect P300 signal among EEG, in which the temporal learning concept is engaged as a new substructure inside the main Convolutional Neural Network (CNN). The above Temporal Convolutional Network (TCN) may better address the problem of P300 detection, thanks to its potential in involving time sequence properties in modelling of these signals. The performance of the proposed method is evaluated on the EPFL BCI dataset, and the obtained results are compared in two inter-subject and intra-subject scenarios with the results of classical CNN in which temporal properties of input are not considered. Increased True Positive Rate of the proposed method (an average of 4 percent) and its accuracy (an average of 2.9 percent) in parallel with the decrease in its False Positive Rate (averagely 3.1 percent) shows the effectiveness of the TCN structure in promoting the detection procedure of P300 signals in BCI applications.

Keywords: EEG signals, P300, Convolutional Neural Networks, Temporal Convolutional Networks, Deep Learning, Brain-Computer Interface.

***Correspondence to:**
Mohamad Reza Keyvanpour,
Department of Computer Engineering, Faculty of Engineering, Alzahra University, Tehran, Iran
Tel: +98 21 88041469
Email: keyvanpour@alzahra.ac.ir

Introduction

Brain Computer Interface (BCI) is a form of human-machine paradigm which may assist people in neuromuscular disorders (1) such as cerebral palsy, multiple sclerosis, amyotrophic lateral sclerosis (ALS), muscular dystrophies, to communicate with other persons or control their environment (2, 3).

Electroencephalogram (EEG) is the most common electrophysiological signal which is utilized in non-invasive BCI systems (4). A BCI system often measures specific components of EEG activity and employs the results as a control signal (5). The most important type of the above components is Event-Related Potentials (ERPs) which arise as brain response (6) to the external visual, tactile, or auditory stimuli and are widely employed in BCI applications (1, 7).

P300 ERP is a positive peak in the EEG and often appears about 300 *ms* after related events happen (8, 9). Based on this fact that P300 BCI users need no special training (10, 11), this type of ERP signals is increasingly applied in BCI systems.

Farwell and Donchin proposed the use of the P300 based BCI in 1988 in the form of "oddball" response (12, 13) which mainly includes a square matrix (i.e. 6×6) of characters such as numeric, alphabetic, and underline symbols (14). This matrix is displayed on a computer screen, and, subsequently the, symbols are flashed in order to evoke a P300 response for the predefined word's characters (15).

Unfortunately, the P300 signal amplitude is much lower than the background EEG activity; therefore, low Signal to Noise Ratio (SNR) is the most challenging issue for accurately detecting the P300 signal (16). The simplest preprocessing method for detecting P300 signal is bandpass filtering of raw EEG signals (17). Make use of averaging is another

naïve technique to amplify the P300 component and gradually improve the SNR challenge. However, averaging also decreases the bit rate and distorts the ERP waveform (18).

Recently, P300 signal detection by using artificial intelligence are used to improve SNR without losing useful information. Therefore, a wide variety of linear and nonlinear machine learning techniques have been applied to classify the P300 signals in BCI. For instance, the linear methods, such as Linear Discriminant Analysis (LDA) (17), Bayesian classification (3) and Support Vector Machine (SVM) (19), have widely been employed in BCI applications. Despite their ease of use, they are weak in the face of complex real-world problems such as overfitting (20).

Artificial Neural Networks (ANNs), as nonlinear techniques, have been utilized to address BCI detection problem which leads to increased SNR and detection of P300 signal (21). The main drawback of such techniques is their high sensitivity to quantity and quality of features which are extracted from the input data (22).

In recent years, the ability of so-called Deep Neural Networks (DNNs) in many applications of recognition and classification in the field of medical signals and images has been well demonstrated. Accordingly, the use of this tool in detection of P300 signal has become an important research topic in BCI domain. DNNs use the recorded signal directly as their input data and are capable of extracting high-level features automatically (22). The most popular DNNs methods, which have been used in the BCI, are the Convolutional Neural Networks (CNN) and Restricted Boltzmann Machines (RBM) (23).

Despite the considerable improvement that CNNs has made in the detection of P300 signals, in many cases this type of networks have not been able to provide sufficient and stable accuracy in their performance.

Conventional CNN stablishes spatial relationships in order to model the nature of its input signal, but unfortunately it does not inherently have a mechanism in order to model the temporal sequences at its input. However, the nature of P300 signals is mainly time- dependent; therefore, in this paper, an upgraded spatial-temporal CNN is presented to improve P300 signal detection. In the proposed method, causal convolutional concept has been utilized beside classical CNN to take advantage of temporal relationships in P300 signals which leads to more accurate results in distinguishing P300 and non-P300 signals.

The rest of the paper is organized as follows: Section 2 includes description of the proposed protocol. In Section 3, the results of applying the proposed structure for P300 detection are demonstrated. In Section 4, the obtained results are compared to some state-of-the-art structures by using their effective indexes. The conclusion is presented in the last section of the paper.

Materials and Methods

The composition of many natural signals is hierarchical, and deep neural networks (DNNs) may lead to better modelling of more complex signals than shallower neural networks in such a way that higher-level features are achieved by composing lower-level ones (24, 25).

2.1. Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are a deep learning approach for successful and robust training, composed of several interconnected layers (26, 27). Each layer consists of neurons that are the essential parts of learning and extracting features from the input layer (26). CNN uses spatial relationships to reduce the number of parameters that must be learned, thus improving general feed-forward backpropagation training (27). The principal part of CNN is the convolutional layer, which performs complex computations. The convolutional layer has mainly a set of learning filters; each of them slides over the input to perform dot products between the filter's weights and the input, and then passes it through a forward path (7). Usually, following convolutional layers, fully-connected (FC) layers are employed to perform final classification. Some other layers, such as the non-linearity, pooling, and normalization, may also be applied in CNN, besides convolutional and FC layers (24). Typically, a non-linear activation function is utilized after each convolutional or FC layer like a sigmoid, hyperbolic tangent or a rectified linear unit (ReLU), which lead to faster training (24, 28).

By applying the convolution kernel on the input signal, several feature maps are obtained (29). Suppose X_j^l is j^{th} feature map in l^{th} convolutional layer (30) as follows:

$$X_j^l = f\left(\sum_{i \in M_j} X_j^{l-1} * W_{ij}^l + b_j^l\right) \qquad (1)$$

In which W^l demonstrates l^{th} layer corresponding weight matrix, $*$ represents the convolution symbol, b_j^l and $f(.)$ show the bias and non-linear activation function, respectively. Usually, after the convolutional procedure, the pooling layer is employed to perform down-sampling which is demonstrated as bellow (30, 31):

$$X_j^l = \beta_j^i \, pooling \, (X_j^{l-1}) \qquad (2)$$

Where *pooling*(.) indicates the rule of down-sampling function, and β_j^i denotes the pooling's weight.

In the next step, the output of the fully-connected layer may be computed as is demonstrated in equation below (32):

$$X^l = f(\, W^l X^{l-1} + b^l\,) \qquad (3)$$

The softmax function is applied after the ultimate FC layer in order to normalize its output and give a probability over all n classes (26). Let a_i be i^{th} neuron output in the final FC layer; then, the above function may generate the final classification probability as described in equation (4), in such a way that the largest P_i shows the most predicted class (26):

$$P_i = \frac{e^{a_i}}{\sum_n e^{a_n}} \qquad (4)$$

2.2. Optimizer Algorithm

Assume $\delta(\theta)$ is a differentiable stochastic scalar function with parameter θ, and $\mathbb{E}[\delta(\theta)]$ is the expected value of this function, which should be minimized to make error least (33). Furthermore, g_t illustrates the changes in the above function in the form of the gradient concept as follows (34):

$$g_t = \nabla_\theta \delta_t(\theta) \qquad (5)$$

In which $\delta_1(\theta), \cdots, \delta_T(\theta)$ denote realizations of the stochastic function at subsequent time steps $1, \cdots, T$. Adam's update rule chooses the step sizes carefully; therefore, it updates m_t and v_t, which are exponential moving averages of the gradient (first moment estimate) and the squared gradient (second raw moment estimate), respectively (33, 35).

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t \qquad (6)$$

$$v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2 \qquad (7)$$

Where $\beta_1, \beta_2 \in [0,1)$ are exponential decay rates for the moment estimates which control the exponential decay rates of m_t and v_t moving averages; additionally, g_t^2 represents $g_t \odot g_t$. Setting the initial value to 0 for moving average vectors causes moment estimates biased towards zero; therefore, to counteract this biasness, $\hat{m}_t$ and $\hat{v}_t$ may be applied as represented in equations (8-9), respectively (33, 34).

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \qquad (8)$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \qquad (9)$$

Finally, equation (10) indicates updating of θ, in which α represents step size and ε demonstrates the smoothing term which avoids division by zero.

$$\theta_t = \theta_{t-1} - \alpha \cdot \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \varepsilon} \qquad (10)$$

2.3. Temporal Convolutional Neural Networks

As mentioned in previous section, although CNN may express spatial connections in the data well, it is mainly weak in modeling information related to time sequence. However, the nature of our signals is completely time dependent. Therefore, Temporal Convolutional Networks (TCN) may be useful to learn the temporal dependencies (36). This network is based on the three basic comments on Fully Convolutional Network (FCN), causal convolution, and dilated convolution. Firstly, by using zero padding in the FCN layer, the output sequence becomes the same as the input sequence length, assuring that the data length is unchanged (37). In the next step, causal convolution, the output at time t is convolved only with the elements corresponding to time t and earlier in the previous layer. Therefore, there is no information leakage from the future to the past. Assume x and u are the input and output sequences, respectively. Furthermore, $f \in \mathbb{R}^{k \times d}$ is a convolution filter with size k. Equation (11) indicates that the output sequence u is well-defined over each time step, and prediction u_t only depends on input x≤t (38).

$$F(x_t) = (x * f)(t) = \sum_{j=0}^{k-1} f_j^T x_{t-j} \quad x \leq 0 := 0 \quad (11)$$

$$u = \big(F(x_1), F(x_2), \ldots, F(x_n)\big)$$

However, simple causal convolution may only look back at history with size linear in the network's depth (37). Dilated convolution is a method that allows for receptive fields exponential to the number of layers, so the rth level dilated convolutional layer may be represented as (38):

$$F(x_t) = (x * l_r f)(t) = \sum_{j=0}^{k-1} f_j^T x_{t-l_r \cdot j} \quad x \leq 0 := 0 \quad (12)$$

$$u = \big(F(x_1), F(x_2), \ldots, F(x_n)\big)$$

In which l_r is the dilation factor that can be set as $(k-1)^{r-l}$ to obtain an exponentially large receptive field. By setting a proper size of filter and number of layers, u_{t+1} may depend on the full historical interactions x≤t. Figure 1 illustrates dilated causal convolution with dilation factors D=1,2,4 (39).

Based on above descriptions and as represented in Figure 2, the proposed method is composed of two parts to take advantage of the EEG data's temporal and spatial information. For this purpose, the TCN section is utilized, which gets the input layer's temporal information and consists of four dilated causal convolutional layers and four ReLUs as
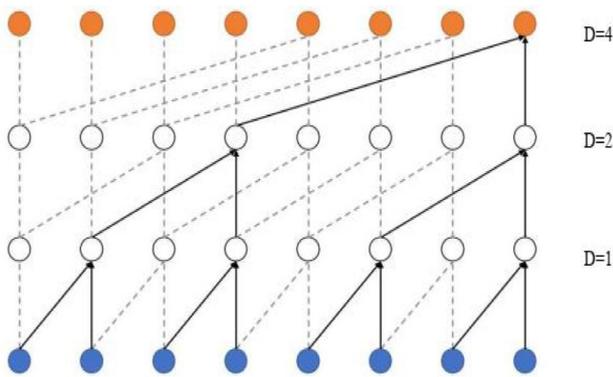
Figure 1: Dilated causal convolution with dilation factors 1, 2 and 4

activation functions between them. Then, the output of this part is fed to the CNN section. The second section is composed of two convolutional layers and two ReLUs. Finally, for classification, the CNN part's output is fed to the FC layer, which computes the score of each class and classifies the signal in the class with a high score.

Results

The proposed method was implemented by utilizing Python 3 Keras framework and then examined on a testbed which was prepared by using Google Colaboratory with NVIDIA Tesla T4 GPU allocation.

The EPFL BCI dataset was used to evaluate the performance of the proposed method. Nine volunteers were enrolled to record the above data, which are mentioned in experiments with the names of subjects 1-9. Among them, subjects 6-9 were healthy, and the others had some disabilities (Figure 3). For each subject, six images, including a television, telephone, lamp, door, window, and radio, were displayed on a laptop screen. Four recording sessions were considered for each subject, and each session consisted of one run for each of the six images.

Data for subject 5 is not in the results because the subject's consciousness level fluctuated strongly. Despite the help of a speech therapist to communicate with subject 5 during the experiments, it was not clear whether he understood the instructions given before
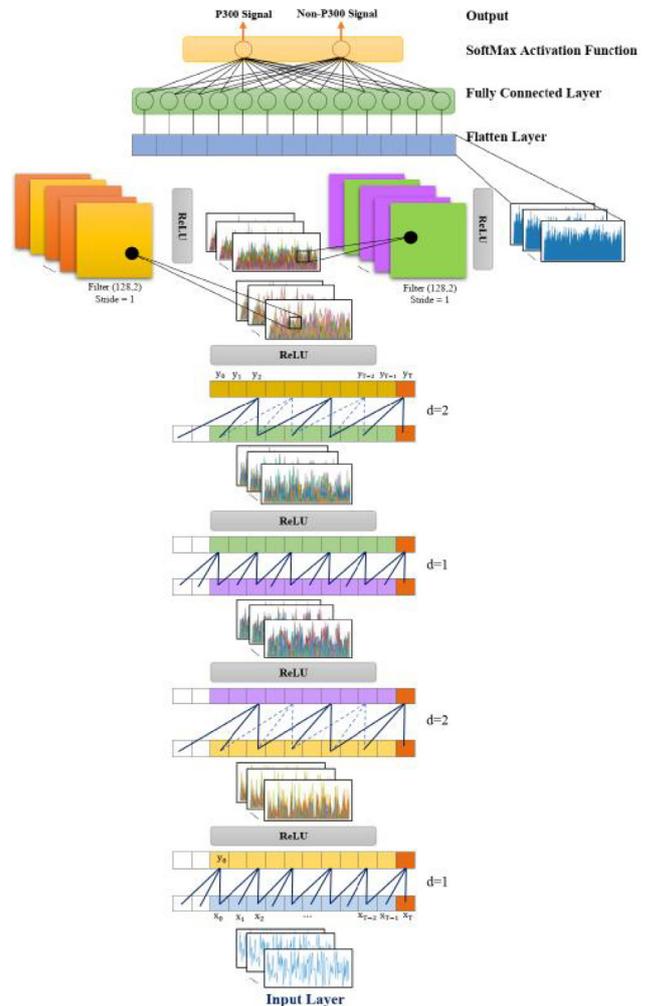
Figure 2: The structure of the proposed method

Requirement: Human raw EEG signals
1- Data preparation:
2- Pre-processing the raw EEG signals
3- Make train, validation, and test datasets using the data of step 2
4- While (training data is available) %fit the model on training data using step 5 to 12
   5- Apply the first dilated causal convolutional layer
   6- Use the output of step 5 to the input of the second dilated causal convolutional layer
   7- Using the output of step 6 as the input of the 3rd dilated causal convolutional layer
   8- Apply the previous step output as input to the 4th dilated causal convolutional layer
   9- Apply first convolutional layer to the output of the last dilated causal convolutional layer
   10- Apply the previous step output as input to the 2nd convolutional layer
   11- Use the flat layer to reshape the 2-D output of the step 10 to a vector
   12- Use the dense layer to map the output of the last layer to P300 and non-P300 signals
13- End while
14- While (validation data is available)
   15- Get a batch of validation data
   16- Use the fitted model to tune batch of validation data
   17- Apply the next batch
18- End while
19- Use the test data to evaluate the trained network

Figure 3: Description of the pseudocode of the proposed method

the experiments. Therefore, in studies that have used the EPFL dataset such as (3, 7, 12), no results have been reported for subject 5. Some other descriptions of the data used are shown in Table 1 and in more detail in (3).

The EEG signals were recorded at the 2048 Hz sampling rate from 32 electrodes placed on the eight subjects' scalp according to standard positions of 10-20 international system. A Biosemi Active Two amplifier was employed to convert analog to digital EEG signals and amplify them.

For each subject, EEG signals were recorded in four separate sessions and six runs. First, each subject's data is pre-processed, including referencing, filtering, down- sampling, signal trial extraction, windsorizing, and normalizing, respectively.

In the next and main step, the CNN and TCN methods were applied on pre-processed data belonging to each subject. Four-fold cross-validation was used. Therefore, the data of each session were used once as a train, validation, and test data. In each fold, the data of two sessions were used for training, and the data of the other two sessions were used in order to validate and test, respectively.

Through trial and error, the best configuration and hyperparameters were obtained for the proposed structure, as described in Table 2. The weights were initialized randomly for the training procedure employing Adaptive moment estimation optimizer with a learning rate of 1e-3 and decay rate of 1e-6.

Good default values for the tested machine learning problems proposed in (33) are: α=0.001, β_1=0.9 , β_2 =0.999 and ε=10^{-8}. The ReLU activation function was employed after each layer, and L2-regularization with a rate of 1e-6 on the weights was applied for all the layers in order to control the overfitting problem.

As shown in (40), using a convolutional layer with stride makes network structures simple without accuracy loss. Therefore, convolutional layers with stride are used instead of applying a pooling layer.

It is certain that the proposed method of this paper, like its competitors, still faces the errors in detecting non-P300 signal instead of P300 and vice versa. The reason for this is the striking similarity of these two types of signals in some cases; for example, some sample signals from subjects 1, 4 and 8 are shown in Figure 4. These figures obviously show no considerable difference between two recorded EEG signals belonging to P300 and non-P300 components.

Eventually, for each subject, standard parameters were applied to compare the efficiency of the examined methods, including True Positive Ratio (TPR), False Positive Ratio (FPR), and accuracy of the classification. The obtained results are reported in Table 3. Interpretation and comparison of the results are made in two scenarios, inter-subject and intra-subject, as described below.

Inter-subject Scenario

In this scenario, the effectiveness of the proposed

Table 1: Details of EPFL BCI

Status	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6-9
	A	A	A	A	NA	A
Diagnosis	Cerebral palsy	Multiple sclerosis	Late-stage amyotrophic lateral sclerosis	Traumatic brain and spinal-cord injury, C4 level	Post-anoxic encephalopathy	Able-body
Age	56	51	47	33	43	30±2.3
Sex	M	M	M	F	M	M

A: subject's data is available in dataset. NA: subject's data are not available in dataset

Table 2: Description of main parameters of proposed method and its deep based alternatives

	TCN+CNN	CNN
Number of layers	8	5
Input layer	1024×1	32×32×1
Dilated Causal Convolution 1	1-D conv (32×2), d=1, Stride=1	
Dilated Causal Convolution 2	1-D conv (20×5), d=2, Stride=1	
Dilated Causal Convolution 3	1-D conv (32×2), d=1, Stride=1	
Dilated Causal Convolution 4	1-D conv (20×5), d=2, Stride=1	
Convolution 1	1-D conv (128×2), Stride=1	2-D conv (5×5), 20, MaxPool (2×2)
Convolution 2	1-D conv (128×2), Stride=1	2-D conv (5×5), 50
Convolution 3		2-D conv (2×2), 32
Output	Softmax	Softmax

d: Represents the dilation rate
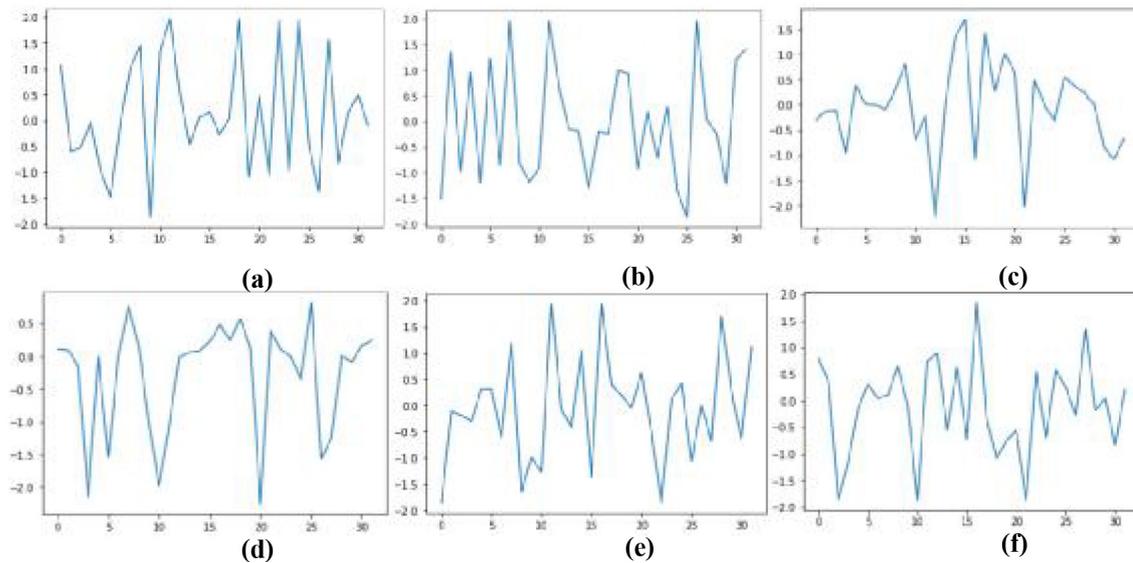
Figure 4: EEG recorded signals belong to subject 1,4 and 8, respectively; (a), (c) and (e) are P300 signals. (b), (d) and (f) are non-P300 signals

Table 3: The parameters over each subject are in percentage (data for subject 5 is not considered in the results)

Subject	Method	TPR	FPR	Accuracy	Subject	Method	TPR	FPR	Accuracy
1	CNN	62.32	16.66	82.12	6	CNN	61.31	14.45	81.50
	TCN	63.04	13.62	82.48		TCN	69.34	12.99	84.06
2	CNN	60.43	21.58	75.41	7	CNN	71.32	14.54	83.10
	TCN	64.49	16.52	80.31		TCN	77.53	9.85	88.04
3	CNN	71.32	17.62	80.53	8	CNN	82.60	12.17	86.95
	TCN	77.94	8.97	88.84		TCN	87.05	11.94	87.88
4	CNN	66.9	13.95	82.85	9	CNN	70.14	15.82	81.84
	TCN	68.34	9.92	86.45		TCN	71.64	15.52	82.33

and alternative methods in the same subjects was compared. Based on TPRs, the proposed method showed a better performance in all subjects than its alternative method. The most notable superiority occurred in subject 6, in which

the proposed scheme increased TPR by almost 8 percent. Moreover, the minimum amount of TPR improvement occurred in subject 1 by 0.72 percent (i.e. there is no meaningful difference between the obtained TPRs). It is worth noting that the average improvement was almost 4 percent in terms of TPR.

Likewise, the results showed that the proposed method performed better than its alternative in terms of FPR. The minimum and maximum improvement in the proposed method's FPR values against the CNN method was 0.23 percent and 8.65 percent, respectively, over subjects 8 and 3. Investigation of the proposed method's superiority over the CNN among other subjects, yielded an average improvement of 3.1 percent in detecting false signals.

Finally, the values obtained from accuracy confirmed the better performance of the proposed

method than its alternative. Therefore, this method's minimum and maximum improved accuracy values arose from subjects 1 and 3 by 0.36 percent and 8.31 percent, respectively. The moderate superiority of our method versus CNN was obtained as 2.9 percent.

Intra-subject Scenario

In this scenario, the best results obtained from the examined methods (i.e. CNN and TCN) were compared in the subject under examination. As Table 3 show that the value of the TPRs of the proposed method was superior compared to the CNN method in all subjects. The best TPR was obtained with the proposed method was 87.05% over subject 8, which is 4.45% better than the best TPR value obtained by applying classical CNN scheme.

Similarly, in terms of the false positive rate, the proposed method showed a significant advantage over its alternative in such way that its best FPR was equal to 8.9 percent (over subject 3), which was 3.2% less than the best result of the CNN (over subject 8).

Ultimately, the classification accuracy in the proposed method for all of the subjects was

outperformed. For instance, the best accuracy was achieved over subject 3, by using the proposed method, equal to 88.8 percent. However, the best value achieved by using the classical CNN was almost 87 percent, over subject 8.

Discussion

The recorded EEG signals were captured from nine volunteers including four healthy persons in parallel with five patients (for the reasons described in Section 2, data for subject 5 was not considered in the results); then, they were classified. Because of the nature of the EEG signals, training and testing procedures were separately performed for each volunteer. The four-fold-cross-validation method was applied to evaluate the performances of data sessions in the test procedure. Based on these tests, the results showed that the proposed method was generally performed superior to its alternative in all subjects and for all evaluation parameters (i.e. TPR, FPR and accuracy).

However, in addition to the overall improvement of the results, there is another parameter that is crucial in evaluating the performance of each method. This is the degree of stability of the results of the method, which means how much the obtained results are concentrated among different folds. Accordingly, in this section, we used the variation range criterion of fold results to examine the aforementioned stability.

Figure 5 demonstrates that the concentration of accuracies obtained from different folds in the proposed method was significantly better than those obtained from the alternative scheme.

It is worth noting that this advantage was more significant for some subjects than others. For instance, exploring the results of the first subject shows that variation ranges of the accuracy for the proposed and basic structures were equal to 1.94% and 3.9%, respectively. These values showed the accuracies obtained from the proposed structure were more compact (81.64% to 84.54%) than the results of the alternative structure (79.83% to 84.17%) during the same folds by 1.96%.

The variation ranges obtained for the second subject were 5.21% and 11.6%, respectively, corresponding to the proposed and alternative methods. The results of accuracies during the same folds determined that the proposed method was more focused (80.31% to 85.03%) than its alternative (75.41% to 83.57%).

The results of the proposed method were more stable for the third subject according to variation ranges, so that its advantage over its alternative reached 17.4%. In this case, the variation ranges of the proposed and alternative method were 0.78 percent
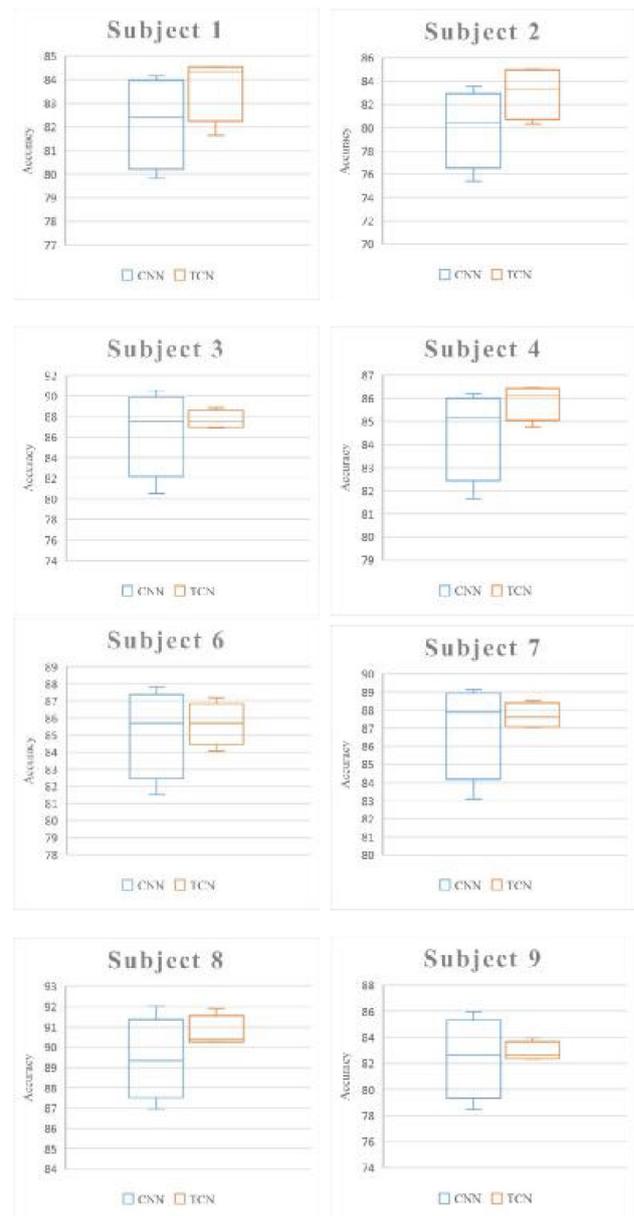
Figure 5: The concentration of accuracies obtained from different folds in the proposed method and its alternative (data for subject 5 is not considered in the results)

and almost 18.18 percent, respectively. The four-fold accuracy ranges were 86.94% to 88.84% and 80.53% to 90.56%, respectively, belonging to the proposed and alternative methods.

Investigating the results of subject four also displayed more concentration of the proposed method against classical CNN, in such a way that variation ranges of the proposed and alternative accuracy were equal to 0.6 percent and approximately 3.5 percent, respectively. In this case, the accuracy values were 84.76% to 86.45% and 81.65% to 86.19%, belonging to the proposed and its alternative method, respectively.

In a similar manner, the accuracy values of subject 6 during the same folds were 84.06% to 87.19% and 81.5% to 87.8%, corresponding to our proposed

method and its alternative, respectively. The values of variation range represented the advantage of the proposed method over its alternative by almost 5.5 percent.

For subject 7, the variation ranges equal to 0.49% and 7.24% were obtained from the proposed method and its alternative, respectively. Therefore, the proposed method was more concentrated (87.06% to 88.52%) than the CNN method (83.1% to 89.13%) by 6.75 percent.

The variation ranges achieved for subject 8 were 0.62% and 4.27%, corresponding to the proposed and alternative algorithms. This subject's accuracy values from the proposed and alternative methods were 90.24% to 91.9% and 86.95% to 92.01, respectively.

As proved by other subjects, by exploring the obtained values, the advantage of the proposed method for subject 9 was also observed. The variation ranges achieved from the proposed and alternative methods were almost 0.6 percent and 9.7 percent. The obtained accuracy variation values indicated that the advantage of our scheme (82.33% to 83.95%) over its alternative (78.5% to 85.94%) reached 9.1%.

Finally, the degree of focus of the accuracy was investigated in terms of two categories of data related to healthy and disabled subjects. Based on the numbers reported in the previous lines, it may be seen that the variances of the accuracies obtained from the proposed method with respect to the same parameter in the alternative method led to relatively corresponding numbers for all the tested subjects. Consequently, the focus of the proposed method compared to the classical CNN indicated that there was no significant correlation between the improving effect of the proposed method on the accuracy concentration and the nature of the data, (i.e., whether healthy or disabled).

Conclusion
In this study, a novel structure was proposed to promote deep neural networks' ability in detecting P300 components among recorded EEG in BCI systems. The proposed method causes the time sequence concept to be considered in constructing the model created by convolutional neural networks. This is done by adding an infrastructure called Temporal Convolutional Networks (i.e., TCNs) to the conventional CNN architecture, thus increasing its potential in detecting P300 signals that are temporal in their nature. To evaluate the effectiveness of the proposed structure, it was applied in a set of real dataset along with its classic alternative (CNN) in forms of inter-subject and intra-subject scenarios;

then, the performance of the examined methods was compared in terms of TPR, FPR and accuracy parameters.

The above comparisons confirmed the superiority of the proposed method over it classic alternative in distinguishing P300 signal from other EEG components. Exploring the achieved TPR values confirmed that the proposed method performed almost 4% and 4.45% better than CNN based on inter-subject and intra-subject scenarios, respectively. Investigating obtained FPR values demonstrated that these superiorities were approximately 3% and 3.2%. In similar manner, the values obtained from accuracy confirmed the better performance of the proposed method than its alternative by 2.9% and almost 1% in inter-subject and intra-subject scenarios, respectively.

In terms of uniformity in the results obtained from different tests, the proposed method showed more stability than its alternative.

The variation ranges of the results obtained in several folds were computed in order to better demonstrate the stability, which showed the proposed method was more concentrated than CNN method among different examinations. Based on the above investigations, it may be concluded that the proposed method may be considered as a high potential candidate for P300 detection in BCI applications.

Conflict of Interest: None declared.

References
1. Ramele R, Villar AJ, Santos JM. EEG Waveform Analysis of P300 ERP with Applications to Brain Computer Interfaces. *Brain Sci*. 2018;8(11). doi: 10.3390/brainsci8110199.
2. Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM. Brain-computer interfaces for communication and control. *Clin Neurophysiol*. 2002;113(6):767-91. doi: 10.1016/s1388-2457(02)00057-3.
3. Hoffmann U, Vesin JM, Ebrahimi T, Diserens K. An efficient P300-based brain-computer interface for disabled subjects. *J Neurosci Methods*. 2008;167(1):115-25. doi: 10.1016/j.

jneumeth.2007.03.005.

4. Panoulas KJ, Hadjileontiadis LJ, Panas SM. Brain-computer interface (BCI): types, processing perspectives and applications. Multimedia services in intelligent environments: Springer; 2010. p. 299-321. doi: 10.1007/978-3-642-13396-1_14.

5. McFarland DJ, Lefkowicz AT, Wolpaw JR. Design and operation of an EEG-based brain-computer interface with digital signal processing technology. *Behavior Research Methods, Instruments, & Computers*. 1997;29(3):337-45. doi: 10.3758/BF03200585.

6. Walter WG, Cooper R, Aldridge VJ, McCallum WC, Winter AL. Contingent Negative Variation: An Electric Sign of Sensorimotor Association and Expectancy in the Human Brain. *Nature*. 1964;203:380-4. doi: 10.1038/203380a0.

7. Shojaedini SV, Morabbi S. Sparse Representation-Based Classification (SRC): A Novel Method to Improve the Performance of Convolutional Neural Networks in Detecting P300 Signals. *Journal of Health Management & Informatics*. 2019;6(2):37-46.

8. Shojaedini SV, Morabbi S, Keyvanpour M. A New Method for Detecting P300 Signals by Using Deep Learning: Hyperparameter Tuning in High-Dimensional Space by Minimizing Nonconvex Error Function. *J Med Signals Sens*. 2018;8(4):205-14. doi: 10.4103/jmss.JMSS_7_18.

9. Nicolas-Alonso LF, Gomez-Gil J. Brain computer interfaces, a review. *Sensors (Basel)*. 2012;12(2):1211-79. doi: 10.3390/s120201211.

10. McCane LM, Heckman SM, McFarland DJ, Townsend G, Mak JN, Sellers EW, et al. P300-based brain-computer interface (BCI) event-related potentials (ERPs): People with amyotrophic lateral sclerosis (ALS) vs. age-matched controls. *Clin Neurophysiol*. 2015;126(11):2124-31. doi: 10.1016/j.clinph.2015.01.013.

11. Lafuente V, Gorriz JM, Ramirez J, Gonzalez E. P300 brainwave extraction from EEG signals: An unsupervised approach. *Expert systems with applications*. 2017;74:1-10. doi: 10.1016/j.eswa.2016.12.038.

12. Shojaedini S, Morabbi S, Keyvanpour M. A New Method to Improve the Performance of Deep Neural Networks in Detecting P300 Signals: Optimizing Curvature of Error Surface Using Genetic Algorithm. *Journal of Biomedical Physics and Engineering*. 2020.

13. Sellers EW, Donchin E. A P300-based brain-computer interface: initial tests by ALS patients.

Clin Neurophysiol. 2006;117(3):538-48. doi: 10.1016/j.clinph.2005.06.027.

14. Shojaedini SV, Adeli M. A New method for Improvement of the Accuracy of Character Recognition in P300 Speller System: Optimization of Channel Selection by Using Recursive Channel Elimination Algorithm Based on Deep Learning. *Journal of Health Management & Informatics*. 2020;7(1):40-9.

15. Guy V, Soriani MH, Bruno M, Papadopoulo T, Desnuelle C, Clerc M. Brain computer interface with the P300 speller: Usability for disabled people with amyotrophic lateral sclerosis. *Ann Phys Rehabil Med*. 2018;61(1):5-11. doi: 10.1016/j.rehab.2017.09.004.

16. Morabbi S, Keyvanpour M, Shojaedini SV. A new method for P300 detection in deep belief networks: Nesterov momentum and drop based learning rate. *Health and Technology*. 2019;9(4):615-30. doi: 10.1007/s12553-018-0276-9.

17. Vareka L, Mautner P, editors. Using the Windowed means paradigm for single trial P300 detection. 2015 38th International Conference on Telecommunications and Signal Processing (TSP); 2015. doi: 10.1109/TSP.2015.7296414.

18. Luck SJ. An introduction to the event-related potential technique. Cambridge: MIT press; 2005.

19. 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(9):1285-94. doi: 10.1016/j.patrec.2008.01.030.

20. Sobhani A, editor P300 classification using deep belief nets. European Symposium on Artificial Neural Networks (ESANN); 2014.

21. Norani NM, Mansor W, Khuan L, editors. A review of signal processing in brain computer interface system. 2010 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES); 2010. doi: 10.1109/IECBES.2010.5742278.

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

23. Lotte F, Bougrain L, Cichocki A, Clerc M, Congedo M, Rakotomamonjy A, et al. A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update. *J Neural Eng*. 2018;15(3):031005. doi: 10.1088/1741-2552/aab2f2.

24. Sze V, Chen Y-H, Yang T-J, Emer JS. Efficient processing of deep neural networks: A tutorial and survey. *Proceedings of the IEEE*. 2017;105(12):2295-329. doi: 10.1109/JPROC.2017.2761740.

25. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436-44. doi: 10.1038/nature14539.

26. Qin Z, Yu F, Liu C, Chen X. How convolutional neural network see the world-A survey of convolutional neural network visualization methods. *arXiv preprint arXiv:180411191*. 2018. doi: 10.3934/mfc.2018008.

27. Arel I, Rose DC, Karnowski TP. Deep machine learning-a new frontier in artificial intelligence research [research frontier]. *IEEE computational intelligence magazine*. 2010;5(4):13-8. doi: 10.1109/MCI.2010.938364.

28. Nair V, Hinton GE, editors. Rectified linear units improve restricted boltzmann machines. *ICML*. 2010..

29. Wang W, Yang Y, Wang X, Wang W, Li J. Development of convolutional neural network and its application in image classification: a survey. *Optical Engineering*. 2019;58(4):040901. doi: 10.1117/1.OE.58.4.040901.

30. Zhiqiang W, Jun L, editors. A review of object detection based on convolutional neural network. 2017 36th Chinese Control Conference (CCC); 2017. doi: 10.23919/ChiCC.2017.8029130.

31. Wu H, Gu X, editors. Max-pooling dropout for regularization of convolutional neural networks. International Conference on Neural Information Processing; 2015. doi: 10.1007/978-3-319-26532-2_6.

32. Bouvrie J. Notes on convolutional neural networks. *Neural Nets*. 2006.

33. Kingma DP, Ba J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:14126980*. 2014.

34. Khan AH, Cao X, Li S, Katsikis VN, Liao L. BAS-ADAM: an ADAM based approach to improve the performance of beetle antennae search optimizer. *IEEE/CAA Journal of Automatica Sinica*. 2020;7(2):461-71. doi: 10.1109/JAS.2020.1003048.

35. Yazan E, Talu MF, editors. Comparison of the stochastic gradient descent based optimization techniques. 2017 International Artificial Intelligence and Data Processing Symposium (IDAP); 2017 doi: 10.1109/IDAP.2017.8090299.

36. Zhang P, Wang X, Chen J, You W, Zhang W. Spectral and Temporal Feature Learning With Two-Stream Neural Networks for Mental Workload Assessment. *IEEE Trans Neural Syst Rehabil Eng*. 2019;27(6):1149-59. doi: 10.1109/TNSRE.2019.2913400.

37. Bai S, Kolter JZ, Koltun V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:180301271*. 2018.

38. You J, Wang Y, Pal A, Eksombatchai P, Rosenburg C, Leskovec J, editors. Hierarchical temporal convolutional networks for dynamic recommender systems. The world wide web conference; 2019. doi: 10.1145/3308558.3313747.

39. Lu N, Yin T, Jing X, editors. A Temporal Convolution Network Solution for EEG Motor Imagery Classification. 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE); 2019. doi: 10.1109/BIBE.2019.00148.

40. Springenberg JT, Dosovitskiy A, Brox T, Riedmiller M. Striving for simplicity: The all convolutional net. *arXiv preprint arXiv:14126806*. 2014.