A deep neural network classifier for P300 BCI speller based on Cohen's class time-frequency distribution

Hamed GHAZIKHANI[1], **Modjtaba ROUHANI**[2,*]
[1]Department of Computer Science and Software Engineering, Concordia University, Montreal, Canada
[2]Department of Computer Engineering, Ferdowsi University of Mashhad (FUM), Mashhad, Iran

Abstract: This paper presents a new method of predicting the P300 component of an electroencephalography (EEG) signal to recognize the characters in a P300 brain-computer interface (BCI) speller accurately. This method consists of a deep learning model and the nonlinear time-frequency features. It is believed that the combination of the deep model network and extracting the nonlinear features of the EEG led this research to a better prediction of the P300 and, therefore, character recognition. Cohen's class distribution is used in order to extract the nonlinear features of the EEG. Evaluating all of the kernels, Butterworth found to be more informative and it produced better results. Based on the differences observed between time-frequency responses of target and nontarget signals, specific subbands are selected to extract seven features. A deep-structured neural network, namely stacked sparse autoencoders, is applied for BCI character recognition. This deep network reduces the dimension of feature space by extracting unsupervised features. Then, the features are fed to a Softmax classifier. Afterward, the whole network passes a fine-tuning phase by a supervised backpropagation algorithm. For evaluating the work, Dataset II of BCI Competition III is utilized. Based on the results, this approach would improve the accuracy in both P300 detection and character recognition. This research results in 82.7% and 93.5% accuracy for P300 classification and character recognition, respectively.

Key words: Cohen's class distribution, P300, brain-computer interface (bci), stacked autoencoders, event-related potential (ERP)

1. Introduction

Brain-computer interface (BCI) is a method to connect the brain to a device [1–7]. This connection could be done through a series of components and signals [8]. The neural activity that is used in the BCI systems could be attained using either noninvasive or invasive techniques [6]. Noninvasive methods are easier for subjects, because there is no surgeries [8]. A common way to record the brain signals and use it for the BCI is EEG (electroencephalography), which is shown in Figure 1. In this figure, there are highlighted channels which are being used in this paper for detecting the P300 component. It is more advantageous to use this method of recording signal for a BCI system [2, 9].

There are some brain signals which appear in specific times, namely event-related potentials (ERPs). We can get these signals via an external stimulator. ERPs have several components consist of a series of positive and negative ones that have a specific meaning and are identified by their occurrence time and the polarity [10, 11]. One of the most important components of an ERP is P300. Its name comes from a positive signal

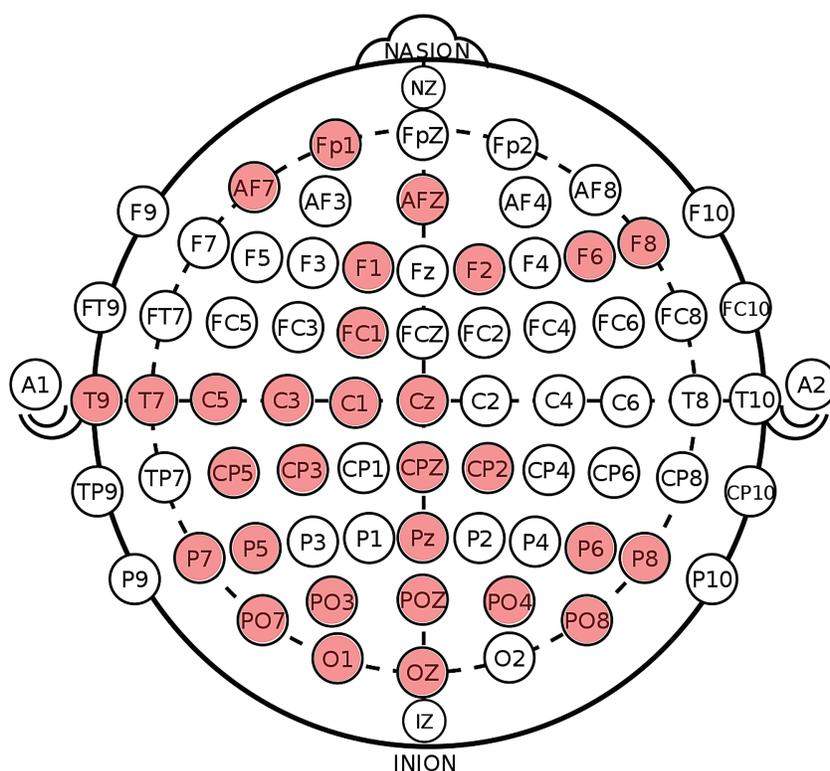---

*Correspondence: rouhani@um.ac.ir

Figure 1. 64 EEG channels based on the 10–20 system. We have used the highlighted channels in our research in order to extract the P300 component. [15].

peak and the delayed duration that appears in a brain signal of around 300 ms. As it is shown in Figure 2, P300 has a positive peak that can be distinguished from nonP300(nontarget) component. There are numerous research on the BCI systems based on P300 component [12–14].

It could be defined that P300 detection is a machine learning problem. There are essential steps to extract the P300, which are: preprocessing, extract/select features, and classification algorithm. Most of the methods have focused on those parts of the signal, which are significant for the preprocessing step. A standard method of preprocessing is bandpass filtering. Also, in order to reduce the noise of the signal and improve the signal-to-noise ratio, a general way is to take the average from the signal [16].

Many different features such as time, frequency, wavelet, and nonlinear features are proposed in various publications [17–19]. Producing substantial numbers of features normally leads to an overfitting problem and requires the reduction of the dimensions of the feature space. This can be done by either feature extraction or feature selection methods. Various feature extraction methods are presented in time and frequency space, such as PCA and genetic algorithms [4]. Some of the publications have used the nonlinear Cohen's class distribution features of EEG in the detection of deception [20] and extracting the visual evoked potential P300 [21]. However, based on our knowledge, there is a lack of research on using this method for the P300 BCI speller.

Cecotti et al. proposed a CNN (convolutional neural network) to classify the P300 [2]. They presented two convolutional layer network to learn both spatial and temporal filters. Besides, they utilized a combination of CNNs in order to check its performance. However, the use of 64 channels and training all those features through CNN must be time-consuming. Liu et al. [22] proposes a new CNN method based on the batch
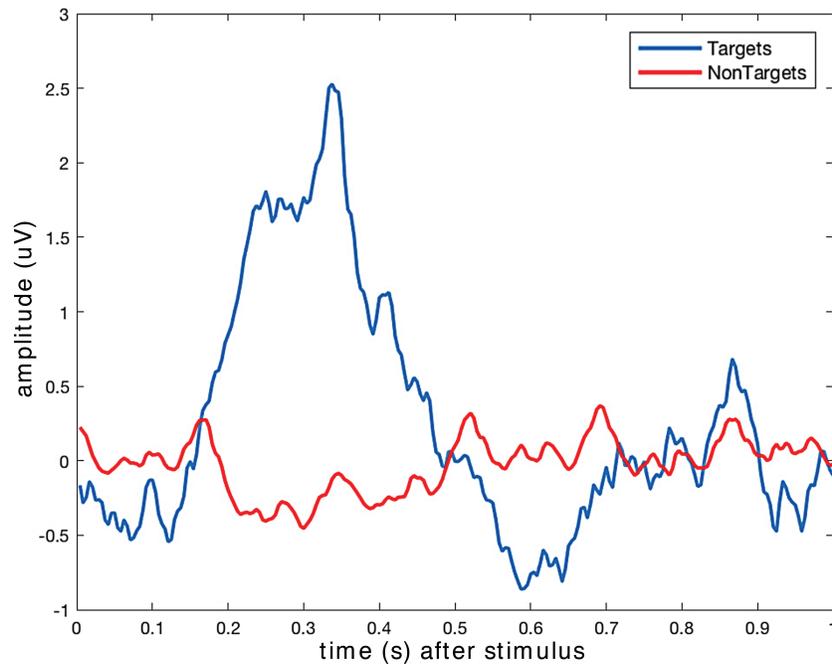
Figure 2. Averaged target and nontarget responses of P300 waveform. P300 has a positive peak which will appear 300 ms after a stimuli.

normalization. A deep belief network is presented by Sobhani to predict the P300 [23]. Also, a method based on the stacked autoencoders is presented by Vareka et al. [24]. Another publication [25] used the stacked autoencoders for the P300 BCI speller using bandpass filtering features.

In this paper, three essential tasks are presented to obtain useful features of the EEG signal. a) Nonlinear chaotic features of EEG signals based on Cohen's class distribution, are extracted. In brief, Cohen's class distribution, which will be described in Section 2.2, is a transformation of a time-based (1-D) EEG signal to a time-frequency space (2-D). Based on observing the differences between target and nontarget distributions, specific subbands of the 2D image of the time-frequency domain, which have the most useful information to extract the P300 component, are selected. b) Seven features (namely: mean, max, standard deviation, median, mode, entropy, and power) are extracted from those subbands for each EEG channel. c) Finally, due to having a lot of EEG channels and features, it is needed to extract features by sparse autoencoders. To classify the P300 component, a deep neural network which is called stacked autoencoders is used. Figure 3 illustrates three tasks by providing informative features for the P300 classification. Character recognition as the main task of this project is used via the typical Algorithm, which is explained in Section 2.4.

The novelties of this paper are:

- Employing the nonlinear chaotic Cohen's distribution to extract both of the time/frequency characteristics of the EEG signal in the P300 BCI speller.

- Selecting the specific subbands of Cohen's time-frequency space based on observing the differences between the target(contains P300) and nontarget(does not contain P300) classes.

- Using stacked autoencoders to extract suitable features among all sets of features that are available.
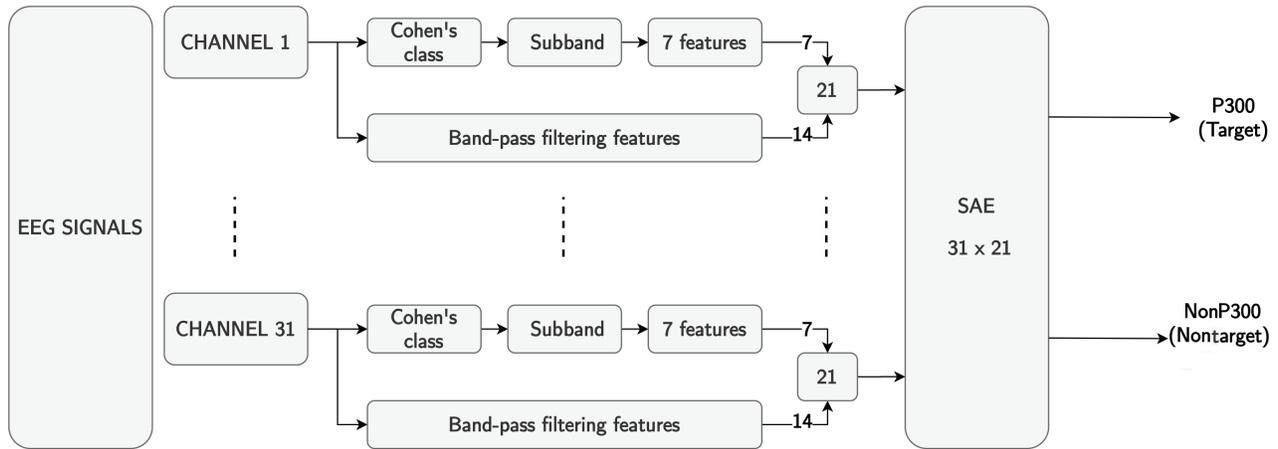
Figure 3. Illustration of P300 classification framework based on the three important feature preparation steps. This figure shows our proposed method and the way that we extract and combine features.

The structure of the paper is: Section 2 presents the proposed method. Section 3 is related to the results and discussion. Section 4 concludes this research paper.

2. Proposed method

In this section, the proposed methodology of this paper is explained.

2.1. Data preprocessing

As mentioned, P300 appears around 300 ms after a stimulus, but it is better to consider the duration of 0 ms to 1000 ms of the signal. Using this knowledge, 31 channels, shown in Figure 1, are selected for the duration of 0 ms and 600 ms. This amount of signal is sufficient in order to classify the P300 [26, 27]. For filtering the EEG signal, 4th order of bandpass filtering, Chebyshev Type I is employed.

The number of nontarget samples in the speller data set is much more than the targets (5:1); therefore, the data are highly imbalanced. Some imitations of the target stimulus are added to the dataset to fix this issue.

Prior to the training step, the features are normalized using the zero mean method as follows:

$$x' = \frac{x - \bar{x}}{\sigma} \tag{1}$$

Where x is the original features, and $\bar{x}$ is the mean of x. σ defines the standard deviation of x. Then, x' can be used to train the network. The same normalization method is used for both training and testing samples.

2.2. Cohen's class distribution

Standard Fourier transform has the power to decompose the signal into frequency components. However, this energy spectrum does not help the researchers to analyze the exact time of the occurrence of those signals. Hence, it is needed to have time in conjunction with the spectrum.

As it is stated, the spectrum is always a powerful way to analyze the nonstationary signals. Time-frequency representation (TFR) is widely used for nonstationary signals to capture exact signal properties in

frequency/time domains. The distribution of the Cohen's class are general frameworks for nonlinear time-frequency distributions. They can generate many different time-frequency distributions by selecting among different kernel functions [28, 29]. In fact, this joint function at first was proposed by Wigner, Ville, Page, and Choi–Williams. Cohen provided a unified formulation in which an infinite number of distributions can be established [30, 31]. This joint function is calculated as follows:

$$P(t,\omega) = \frac{1}{4\pi^2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-j\theta t - j\tau\omega + j\theta u} \phi(\theta,\tau) \cdot s^*\left(u - \frac{1}{2}\tau\right) s\left(u + \frac{1}{2}\tau\right) du\, d\tau\, d\theta \tag{2}$$

Where $\phi(\theta,\tau)$ is an arbitrary function called the *kernel* function and $s(\cdot)$ is the signal. $P(t,\omega)$ is a function that transform the time based signal into time-frequency domain. With a time-frequency representation the frequency contents of a signal at different times can be understood [32]. So, the $\phi(\theta,\tau)$ function tells that how the energy of signal is distributed in the time and frequency domain. Different kernel functions in order to produce various distributions of the signal can be used [30, 31]. Table 1 shows some important kernel functions of Cohen's class distribution.

Table 1. The kernel functions for some Cohen distributions

Distribution	Kernel function
Butterworth	$\frac{1}{1+(\frac{\theta}{\theta_1})^{2N}(\frac{\tau}{\tau_1})^{2M}}$
Choi–Williams	$\exp\left[\frac{-(\pi\theta\tau)^2}{2\sigma^2}\right]$
Born–Jordan	$\frac{\sin(\pi\theta\tau)}{\pi\theta\tau}$
Wigner–Ville (WV)	1
Rihaczek	$\exp[j\pi\theta\tau]$
Margenau–Hill	$\cos(\pi\theta\tau)$
Page	$\exp[-j\pi\theta|\tau|]$

According to Cohen's notes [31], for specific problems, it is needed to examine different distribution to choose the best one. Accordingly, a very well-matched distribution that fits the EEG data is chosen. This distribution is called Butterworth, which will be described in the next part.

2.2.1. Butterworth distribution
The Butterworth distribution (BUD) is a kernel of Cohen's class that is given in Eq. 2. The BUD kernel is as follows:

$$\phi_{BUD}(\theta,\tau) = \frac{1}{1 + (\frac{\theta}{\theta_1})^{2N}(\frac{\tau}{\tau_1})^{2M}} \tag{3}$$

Where N and M are positive order parameters and θ_1 and τ_1 are positive spectral, temporal scaling constants. Note that $\phi_{BUD}(\theta_1,\tau_1) = \frac{1}{2}$ for any value of $\theta_1, \tau_1, N,$ or M.

2.2.2. Feature extraction
A key role in every machine learning task is to provide the most informative and independent features for the classifier. In this paper, The combination of linear and nonlinear features of EEG signals based on Cohen's class distribution for the P300 detection problem is employed. Based on Cohen's class distribution method, we can

simultaneously see the changes in frequency over time. There is much information in Cohen's class distribution, and it is not practical to have an appropriate process on the data. Therefore, by representing the Cohen's time-frequency distribution of the EEG signal, specific parts of the signal which contain useful information can be observed. These specific parts are sub-bands that are selected for further feature extraction process. Averaged target and nontarget data from each channel of EEG were examined in order to find the differences. Therefore, these varieties in time-frequency subbands between the P300 and non-P300 were scrutinized, which have the most fruitful information. This issue is significant for classification tasks. Seven features (namely: mean, max, standard deviation, median, mode, entropy, and power) are extracted from those subbands that were selected. In addition to those seven features, fourteen features [27] obtained from 4th-order bandpass Chebyshev Type I. So, by adding these features together, $21(14 + 7)$ features are available for each channel. These features $(21 \times 31 = 651)$ are entered into the deep neural network.

2.3. Classifying the P300 component

Stacked autoencoders are an important and popular topic in the deep neural network field, which has been used in many various machine learning problems. The idea of this phenomenon was originally initiated to improve the performance of deep neural networks [33]. Stacked autoencoders are a group of autoencoders (often sparse autoencoder) that can solve a machine learning problem [34].

Connected autoencoders (as it is called stacked autoencoders) is employed to reduce the dimension of a feature space and to extract a new representation of the data. The input is fed to the first autoencoder, and the first autoencoder's code would be passed to the next autoencoder and so far to the last one. In a stacked autoencoder, the encoder (first layer) part of each autoencoder is used. This process will continue until the desired feature size is reached and it is ready to pass to the Softmax classifier. Stacked autoencoders have two learning phases: a pretraining and a fine-tuning. Each autoencoder is pretrained separately based on the output of the previous one. In the latter phase, the parameters of the whole network are fine-tuned via the error backpropagation algorithm. Figure 4 demonstrates a structure of stacked autoencoder.
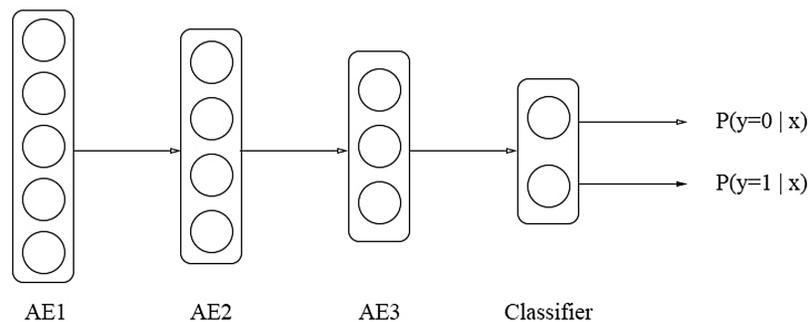

Figure 4. A scheme of a stacked autoencoders. This Figure demonstrates a binary classification problem.

Accordingly, the features are reduced from 651 to 10, by autoencoders. Then, the final obtained ten features were fed to a Softmax. The structure of the proposed stacked autoencoders is as follows: 651-440-220-100-50-10-2. Next, backpropagation was used in order to fine-tune the whole network. Figure 5 illustrates the proposed stacked autoencoders.
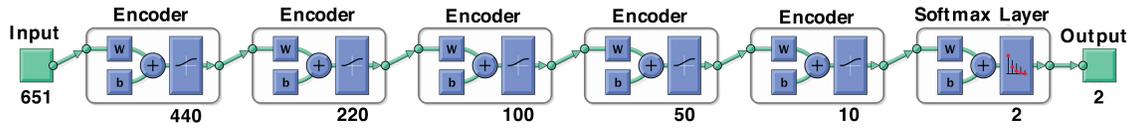
Figure 5. The proposed stacked sparse autoencoder deep structure for our research.

2.4. Character recognition

The main purpose of classifying the P300 is to recognize the characters in a speller system. In the BCI competition III, there are 30 targets out of 180 trials in each character. This process has been repeated 15 times for improvement in the character recognition rate. The feature vectors for each trial are fed to the trained network to classify the targets. By accumulating each predicted score correspondence to the row or the column, the desired character would be found. Eq. 4 shows the process of accumulating the predicted scores:

$$C_i = \sum_{i=1}^{15} S(f_{r,c}) \tag{4}$$

Where $S(f_{r,c})$ is the output of the proposed stacked autoencoders corresponding to the presence of P300. Consequently, the focused character would be chosen:

$$column = \text{argmax}_{1 \leq i \leq 6} C_i \tag{5}$$

$$row = \text{argmax}_{7 \leq i \leq 12} C_i$$

Figure 3 demonstrates the whole process of the proposed method, which is an algorithm to predict a character in a P300 BCI speller. The overall process is also detailed in Figure 6.

3. Experiments

3.1. Dataset

The BCI competition III and its second dataset[1] with two different subjects is used. There are training and test sets for both of the subjects. Each of the training sets contains 85 characters, and there are 100 characters in every test set. To have more accurate predictions, the process of character recognition is repeated 15 times for all of the 12 rows and columns. As a result, the number of trials is 180 ($15 \times 12 = 180$), containing only 30 targets.

3.2. Application of proposed method

Cohen's class method with Butterworth kernel is applied to transform and extract the important features. Other methods and kernels to determine which one of the kernels would be suitable for this problem have been evaluated. The Butterworth kernel, as the best one among others to do the process of feature extraction, is discovered. Figure **??** represents three averaged target and nontarget for different channels. As it can be considered, there are differences in specific subbands (highlighted with red rectangles) of Cohen's distribution

---

[1]Berlin Brain-Computer Interfac (2005). BCI Competition III – Data set II: [online]. Website http://www.bbci.de/competition/iii.
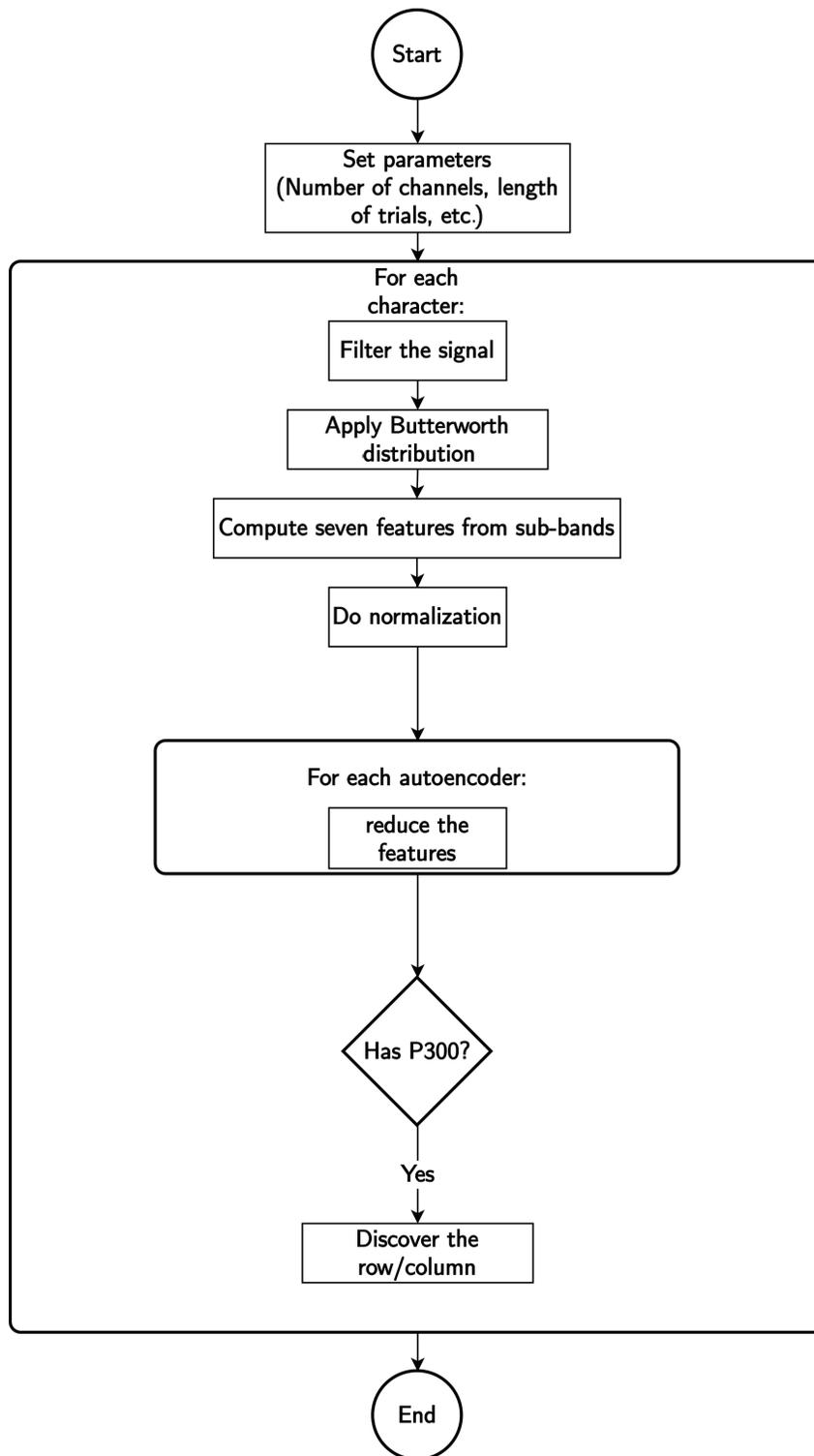
Figure 6. The overall process of proposed method of character recognition.

between the target and nontarget. Accordingly, those differences in the same area are considered and calculated through the implementations. For example, in channel 11, which is the CZ electrode of the 10–20 system, the subband that is different between the averaged target and nontarget in the time domain (x-axis) is in the range of 590 ms to 666 ms, and, in the frequency domain, (y-axis) is in the range of 0 Hz to 12 Hz. Figure 3.2 refers to an averaged target stimulus of channel 11 (CZ) and Figure 3.2 mentions the averaged nontarget stimulus that does not contain the P300 component. Equally, other figures have the same explanation. Figure 3.2 and Figure 3.2 demonstrates the time-frequnecy distribution of channel 58 and Figure 3.2, Figure 3.2 posits the time and the spectrum both for the channel number 60.

These subbands of the signal are considered for feature extraction. The total number of available features is 21. This number stems from the 14 features from the filtering procedure and seven features extracted by selected subbands of the Butterworth kernel. These seven features are mean, max, standard deviation, median, mode, entropy, and power. These parameters are calculated after applying the Butterworth distribution. So, 21 features $(14 + 7)$ for each channel are achieved, therefore, for 31 channels a $21 \times 31 = 651$ data are obtained. After that, the feature vectors are normalized via scaling to zero mean and unit variance, which is shown in Eq. 1. Next, the problem's imbalance issue is solved by duplicating the signals that include the target components. Eventually, the feature vectors are passed through the deep neural network in order to train the classifier.

We define a dataset D consists of all the trials $D = \{x_i, y_i\}_{i=1}^N$. x_i refers to each feature vector and $y_i \in \{-1, +1\}$ refers to correspondence label, which is either +1 (Target) or –1 (Nontarget).

Matlab[2] is used to implement the stacked autoencoders. Some of the parameters (number of neurons/layers) of the network were computed empirically. Other parameters, such as L2 weight regularization, sparsity regularization, and sparsity proportion, are chosen via the cross-validation algorithm, and they are 0.004, 4, and 0.2, respectively.

3.3. Discussions

In this section, the results that have been achieved would be explained. P300 detection is a two classes (binary) problem that it is (+1) when the signal contains P300 component and (–1) when there is an absence of the P300 component. For evaluation of the suggested method, accuracy, precision, recall and F1-score is calculated. These measures are formulated as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

$$F1 - score = \frac{2 \times (Recall \times Precision)}{(Recall + Precision)} \tag{9}$$

While (TP) is the number of true positive, (TN) true negative, (FP) false positive and (FN) false negative. Table 2 demonstrates the results of the P300 detection evaluation for all measurements. According

---

[2]MathWorks Inc. (2017). MATLAB [online]. Website http://www.mathworks.com/products/neuralnet.
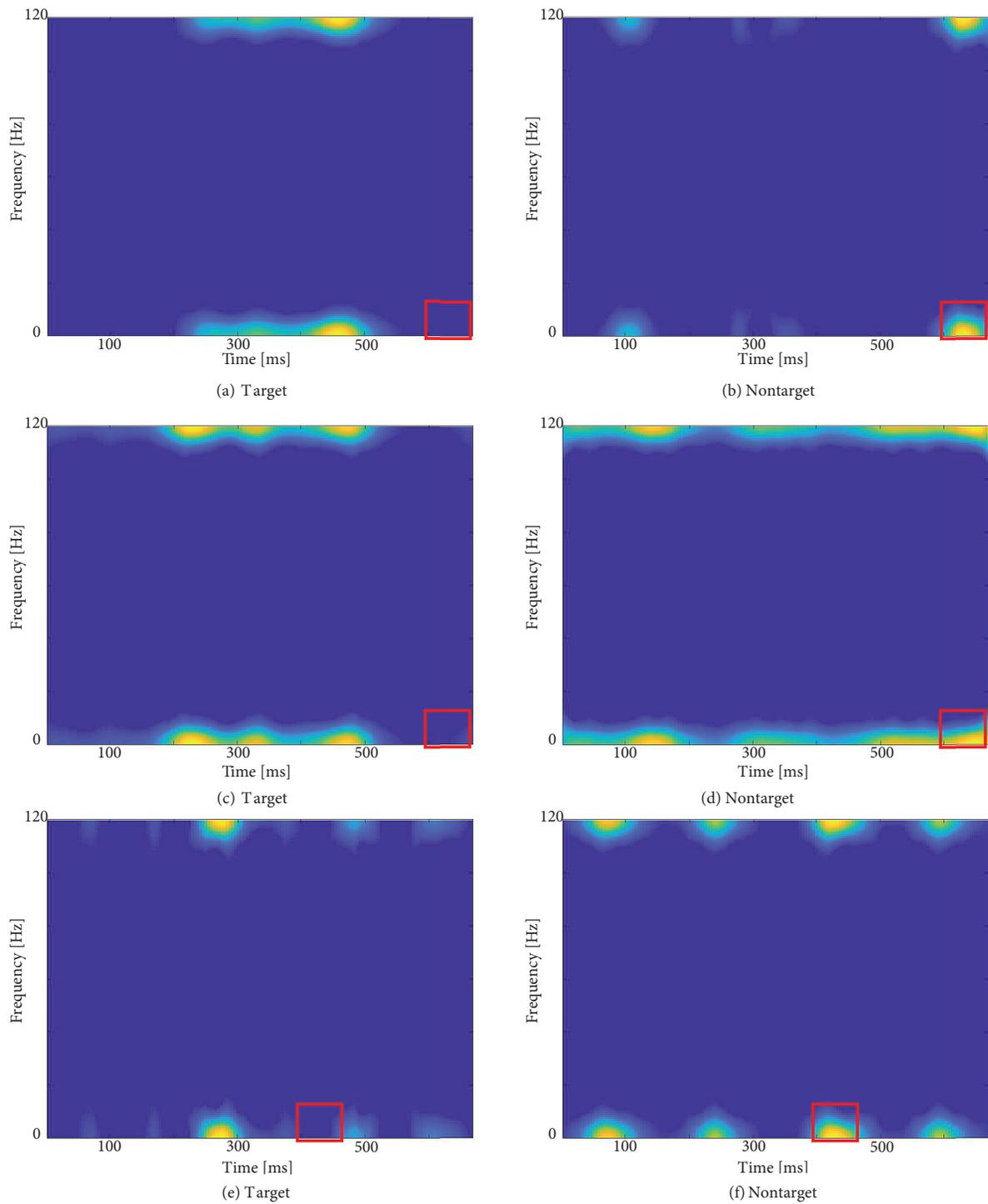
Figure 7. A depiction of a Butterworth's distributions for averaged targets and nontargets of three channels. In these figures, converted signal to time-frequency domain are shown. This is the way that we compare targets and nontargets areas.

to the results, for subjects A and B, an accuracy of 81.2% and 84.2% are obtained for the P300 detection, respectively. Consequently, Table 3 shows the proposed method of performance comparing to other state-of-the-art publications in this field. The proposed method using Cohen's class distribution produces much better results in comparison with other methods in the P300 detection rate.

Table 2. Results of P300 detection evaluation for subject A and B.

Subject	TP	TN	FP	FN	Accuracy	Precision	Recall	F1-score
A	7200(40.3%)	7287(40.8%)	1635(9.2%)	1728(9.7%)	81.2%	81.4%	80.6%	80.9%
B	7385(41.4%)	7651(42.9%)	1274(7.1%)	1540(8.6%)	84.2%	85.2%	82.7%	83.9%
Mean	7292(40.85%)	7469(41.8%)	1454(8.1%)	1634(9.1%)	82.7%	83.3%	81.6%	82.4%

Table 3. Performance comparison of the P300 detection.

Method	Accuracy(%)
MCNN-3 [2]	74.2
CNN-1 [2]	74.2
Stacked autoencoders [24]	69.2
Stacked autoencoders based on wavelet* [16]	79.3
DBN** [23]	60-90
LDA*	75.6
BLDA*	77.1
SVM*	73.7
MLP*	76.7
Human detection*	64.4
Proposed method	**82.7**

*Reported and tested by [16].

**Various reports by the author.

Our final and most crucial step is recognizing the exact characters. The complete process of character detection is clarified in Section 2.4. Table 4 represents the total accuracy of the proposed method in 15 repetitions of the character recognition process. As it is evident, the development is apparent while the cycles are repeated, and the more repetitions are performed, the better the results for recognizing the characters are achieved. Figure 8 reveals the performance of the proposed method for character recognition rate of 15 repetitions. To assess the presented technique, the results are compared to the previous works. Table 5 shows the performance of the proposed method in comparison with others. References [2, 23–25] have used the deep learning methods for P300 recognition. In [25] the character recognition accuracies are 55.5% and 91.5% for 5 and 15 epochs, respectively, while for the proposed method, those accuracies are raised to 61% and 93.5%. This achievement is because of using features extracted employing Cohen's class distribution in addition to filtering features. Indeed, Cohen's class distribution helped the learner to distinguish the target and nontarget more precisely, so fruitful features are extracted. In brief, Cohen's class distribution can extract chaotic, nonlinear features with much more information for nonstationary signals.

Nevertheless, because there are vast amounts of Cohen's data, it is not suitable to use all of them. So, by averaging on the target and nontarget signals, sections of 2D time-frequency representation have been

Table 4. Accuracy of character recognition in percent.

Subject	Epochs														
	1	2	3	4	**5**	6	7	8	9	**10**	11	12	13	14	**15**
A	14	29	45	49	**52**	58	68	66	72	**82**	82	84	85	89	**94**
B	27	35	53	60	**70**	74	73	83	88	**89**	93	91	91	92	**93**
Mean	20.5	32	49	54.5	**61**	66	70.5	74.5	80	**85.5**	87.5	87.5	88	90.5	**93.5**

Table 5. Performance comparison of the character recognition rate in percent.

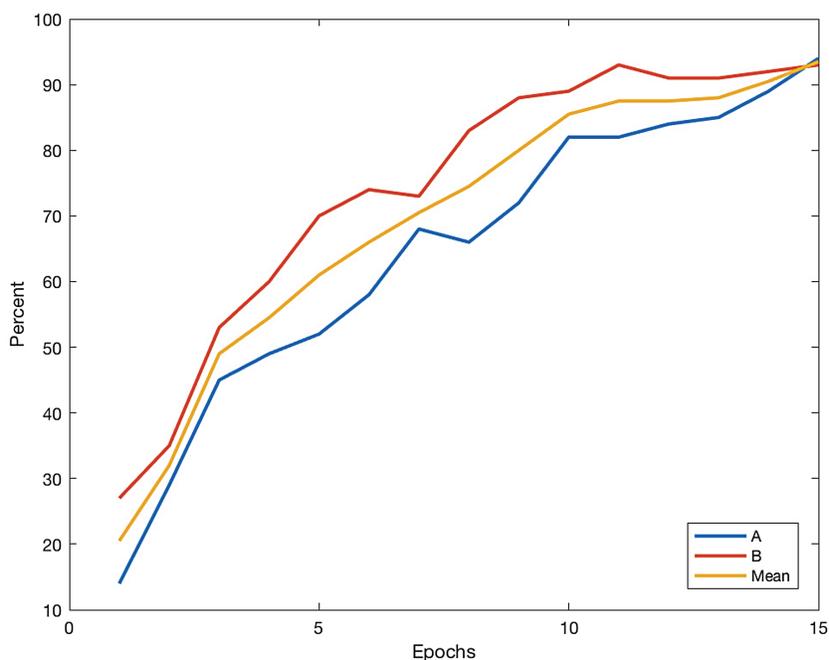| Method	Repetitions	
	5	15
MCNN-2[2]	55	90.5
MKL[26]	13.5	13
LSVM[26]	47.5	76
GSVM[26]	3	3
Hoffmann[35]	53	89.5
Zongtan[?]	59.5	90.5
Yandong[?]	55	90.5
mLVQ[2]	59.5	91.5
LDA[2]	60.5	93
Stacked Autoencoders [25]	55.5	91.5
Proposed method	**61**	**93.5**

Figure 8. The accuracy of character recognition for both subjects.

recognized, which contains the most differences between the targets and nontargets. After that, the number of features was reduced to classify the P300 component with the help of Stacked autoencoders. Finally, the

character recognition phase was performed based on the trained Stacked autoencoders.

4. Conclusion

In this project, a new method to recognize the P300 waveform in an EEG signal is formulated. As it is described, EEG signals are encircled by a substantial amount of noise. Thus, it requires a robust feature extraction method and a machine-learning algorithm to resist the outliers and sustain the current condition of the signal. This means that an algorithm demanded to act against the noisy data and preserve the useful information among signals. Therefore, Cohen's class distribution is used for extracting functional features and a deep neural network to learn the features and do the classification toward the project objectives. Cohen's class distribution is a nonlinear time-frequency representation (TFR) that captures the exact signal properties both in time and frequency domains. One of the findings of this paper is that the average power spectrum of target and nontarget signals are distinguishable in specific subbands of Cohen's class distribution. Afterward, a deep neural model that learns the representation of data in an unsupervised manner is introduced. This neural net is called stacked sparse autoencoders, which has been used in many research. To evaluate the performance of the implementation, the BCI competition III dataset II is employed. According to the results, the performance of the proposed method is comparable with the state-of-the-art methods. In the classification of the P300 component, an accuracy of 82.7% in the test set, which is 30% of the training samples, is achieved. For character recognition, the accuracy of 93.5%, which shows the robustness of the neural network, is attained. For future works, the combination of other feature extraction methods with Cohen's class distributions, signal processing algorithms, and other deep learning models that may enhance this system would be considered.

Conflict of interest

The authors declare that they have no conflict of interest.

References

[1] Amiri S, Fazel-Rezai R, Asadpour V. A review of hybrid brain-computer interface systems. Advances in Human-Computer Interaction 2013. doi: 10.1155/2013/187024

[2] Cecotti H, Graser A. Convolutional Neural Networks for P300 Detection with Application to Brain-Computer Interfaces. IEEE Transactions on Pattern Analysis and Machine Intelligence 2011; 33 (3): 433-445. doi: 10.1109/TPAMI.2010.125

[3] Elshout J, Garcia Molina G. Review of brain-computer interfaces based on the P300 evoked potential. Philips Research Europe 2009.

[4] Lotte F, Bougrain L, Cichocki A, Clerc M, Congedo M et al. A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. Journal of Neural Engineering 2018; 15 (3). doi:10.1088/1741-2552/aab2f2

[5] Vaid S, Singh P, Kaur C. EEG signal analysis for BCI interface: A Review. In: 2015 Fifth International Conference on Advanced Computing & Communication Technologies; Haryana; 2015. pp. 143-147. doi: 10.1109/ACCT.2015.72

[6] Wolpaw JR, Birbaumer N, Heetderks WJ, McFarland DJ, Peckham PH et al. Brain-computer interface technology: a review of the first international meeting. IEEE Transactions on Rehabilitation Engineering 2000; 8(2): 164-173. doi: 10.1109/tre.2000.847807

[7] Olivas-Padilla BE, Chacon-Murguia MI. Classification of multiple motor imagery using deep convolutional neural networks and spatial filters. Applied Soft Computing 2019; 75: 461-472. doi: j.asoc.2018.11.031

[8] Vallabhaneni A, Wang T, He B. Brain—Computer Interface. In: He B. (eds) Neural Engineering. Bioelectric Engineering. Boston, MA, USA: Springer2005. doi: 10.1007/0-306-48610-5_3

[9] Al-Sharhan S, Bimba A. Adaptive multi-parent crossover GA for feature optimization in epileptic seizure identification. Applied Soft Computing 2019; 75: 575-587. doi: j.asoc.2018.11.012

[10] Bressler SL, Ding M. Event-Related Potentials. Wiley Encyclopedia of Biomedical Engineering 2006. doi: 10.1002/9780471740360.ebs0455

[11] Koukkou M, Koenig T, Bänninger A, Rieger K, Diaz Hernandez L et al. Neurobiology of schizophrenia: electrophysiological indices. In: A. Javed & K. N. Fountoulakis (eds.), Advances in Psychiatry. Boston, MA, USA: Springer International Publishing, 2019, pp. 433–459. doi: 10.1007/978-3-319-70554-5_27

[12] Farwell LA, Donchin E. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. Electroencephalography and Clinical Neurophysiology 1988; 70 (6), 510-523. doi: 10.1016/0013-4694(88)90149-6

[13] Fazel-Rezai R, Allison BZ, Guger C, Sellers EW, Kleih SC et al. P300 brain computer interface: current challenges and emerging trends. Frontiers in Neuroengineering 2012; 5, 14. doi: 10.3389/fneng.2012.00014

[14] Amiri S, Fazel-Rezai R, Asadpour V. A review of hybrid brain-computer interface systems. Advances in Human-Computer Interaction 2013; 2013: 1. doi: 10.1155/2013/187024

[15] Song C, Jeon S, Lee S, Ha HG, Kim J et al. Augmented reality-based electrode guidance system for reliable electroencephalography. BioMedical Engineering OnLine 2018; 17(1): 64. doi: 10.1186/s12938-018-0500-x

[16] Vařeka L, Prokop T, Mouček R, Mautner P, Štěbeták J. Application of stacked autoencoders to P300 experimental data. In: International Conference on Artificial Intelligence and Soft Computing; 2017. pp. 187–198. doi: 10.1007/978-3-319-59063-9_17

[17] Cömert Z, Boopathi AM, Velappan S, Yang Z, Kocamaz AF. The influences of different window functions and lengths on image-based time-frequency features of fetal heart rate signals. In: 2018 26th Signal Processing and Communications Applications Conference (SIU); 2018. pp. 1-4. doi: 10.1109/SIU.2018.8404247

[18] Iscan Z, Dokur Z, Demiralp T. Classification of electroencephalogram signals with combined time and frequency features. Expert Systems with Applications 2011; 38 (8), 10499-10505. doi: 10.1016/j.eswa.2011.02.110

[19] Murugappan M, Nagarajan R, Yaacob S. Comparison of different wavelet features from EEG signals for classifying human emotions. In: 2009 IEEE Symposium on Industrial Electronics & Applications; Kuala Lumpur, Malaysia; 2009. pp. 836-841. doi: 10.1109/ISIEA.2009.5356339

[20] Ebrahimzadeh E, Alavi SM, Bijar A, Pakkhesal A. A novel approach for detection of deception using smoothed pseudo Wigner–Ville distribution (SPWVD). Journal of Biomedical Science and Engineering 2013; 6 (01): 8. doi: 10.4236/jbise.2013.61002

[21] Cheng W, Ren Z, Qiao X. Improved Wigner–Ville distribution for extracting the visual evoked potential P300. Journal of Shanxi University (Natural Science Edition) 2013; 2, 18.

[22] Liu M, Wu W, Gu Z, Yu Z, Qi F et al. Deep learning based on Batch Normalization for P300 signal detection. Neurocomputing 2018; 275: 288-297. doi: 10.1016/j.neucom.2017.08.039

[23] Sobhani A. P300 Classification using Deep Belief Nets. MSc, Colorado State University, Fort Collins, Colorado, USA, 2014.

[24] Vařeka L, Mautner P. Stacked autoencoders for the P300 component detection. Frontiers in Neuroscience 2017; 11: 302. doi: 10.3389/fnins.2017.00302

[25] Ghazikhani H, Rouhani M. A stacked autoencoders approach for a P300 Speller BCI. In: 8th International Conference on Computer and Knowledge Engineering (ICCKE2018); Mashhad, Iran; 2018. pp. 25-26. doi: 10.1109/ICCKE.2018.8566534

[26] Yoon K, Kim K. Multiple kernel learning based on three discriminant features for a P300 speller BCI. Neurocomputing 2017; 237: 133–144. doi: 10.1016/j.neucom.2016.09.053

[27] Rakotomamonjy A, Guigue V. BCI competition III: dataset II-ensemble of SVMs for BCI P300 speller. IEEE Transactions on Biomedical Engineering 2008; 55 (3): 1147–1154. doi: 10.1109/TBME.2008.915728

[28] Konopko K. An implementation of the Cohen's class time-frequency distributions on a massively parallel processor. Prz. Elektrot 2012; 88 (9B): 289-291.

[29] Ochoa JB. EEG signal classification for brain computer interface applications. Ecole Polytechnique Federale De Lausanne 2002; 7: 1-72.

[30] Kavallieratou E, Fakotakis N, Kokkinakis G. Skew angle estimation in document processing using Cohen's class distributions. Pattern Recognition Letters 1999; 20 (11): 1305-1311. doi: 10.1016/S0167-8655(99)00097-5

[31] Cohen L. Time-frequency distributions-a review. Proceedings of the IEEE 1989; 77(7): 941-981. doi: 10.1109/5.30749.

[32] Poyil AT, Nasimudeen KM, Aljahdali S. Significance of Cohen's class for time frequency analysis of signals. International Journal of Computer Applications 2013; 72 (12): 1-8. doi: 10.5120/12543-8960

[33] Bengio Y, Lamblin P, Popovici D, Larochelle H. Greedy layer-wise training of deep networks. In: NIPS'06: Proceedings of the 19th International Conference on Neural Information Processing; Cambridge, MA, USA; 2006. pp. 153-160.

[34] Zhao C, Zhang L. Spectral-spatial stacked autoencoders based on low-rank and sparse matrix decomposition for hyperspectral anomaly detection. Infrared Physics & Technology 2018; 92: 166-176. doi: 10.1016/j.infrared.2018.06.001

[35] Hoffmann U, Garcia G, Vesin JM, Ebrahimi T. Application of the evidence framework to brain-computer interfaces. In: The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; San Francisco, CA, USA; 2004. pp. 446-449.