Subject-Independent P300 Speller Classification using Time-Frequency Representation and Double Input CNN with Feature Concatenation

Zangar Ermaganbet[1], Ayana Mussabayeva[2], Prashant Kumar Jamwal[1], and Muhammad Tahir Akhtar[1]

[1]*Department of Electrical and Computer Engineering, School of Engineering & Digital Sciences,* Nazarbayev University, Kabanbay Batyr Ave. 53, Astana, Kazakhstan.
[2]*Department of Mathematics, University of Manchester,* Oxford Rd, Manchester M13 9PL, United Kingdom
Emails: zangar.ermaganbet@nu.edu.kz, ayana.mussabayeva@manchester.ac.uk, prashant.jamwal@nu.edu.kz, muhammad.akhtar@nu.edu.kz, akhtar@ieee.org

***Abstract*—This study proposes a Double Input Convolutional Neural Network with Feature Concatenation (DiCNN-FC) for the classification task of the P300 speller. Two time-frequency representations of electroencephalography (EEG); namely, power and phase spectrograms; have been employed as input for the DiCNN-FC. Each spectrogram has been processed separately using convolutional layers and concatenated with each other for decision-making in the dense layers. The use of DiCNN-FC produces reliable results as the decision is made based on two sets of features. Two P300 datasets, one from amyotrophic lateral sclerosis (ALS) patients, and another from healthy subjects have been used to evaluate the performance of the proposed method. The performance comparison has been performed for two classical methods for P300 classification, namely Support Vector Machine (SVM) and Linear Discriminant Analysis (LDA). The achieved results show the DiCNN-FC model's ability to perform subject-independent P300 component identification based on single-trial data on both datasets.**

***Index Terms*—Brain-Computer Interface, P300 Speller, Time-Frequency Representation, Convolutional Neural Network.**

I. Introduction

The main objective of any Brain-Computer Interface (BCI) system is to acquire brain activity data and interpret it into instructions for an external device. Any BCI system comprises three main steps: pre-processing brain data, extracting features from the pre-processed data, and classification using extracted features. Then, classified data is applied to control the external device. One of the main applications of BCI techniques is in the field of assisting devices, such as wheelchairs [1] and prostheses [2]. In addition, recently, the entertainment and gaming industries started applying BCI for hands-free control [3]. Apart from that, another popular application of BCI is spelling systems. They are aimed at helping people with severe motor neuron disorders (e.g., amyotrophic lateral sclerosis (ALS), cerebral palsy, etc.) to communicate with the external world by interpreting EEG signals of their brain.

One of the well-known spelling systems is the P300 speller [4]. The P300 speller uses one of the event-related potentials (ERP), a positive peak that appears 300 ms after the stimulus - P300 component. As a stimulus, the P300 speller uses a flashing matrix of characters, from which the subject should focus on the needed character. The P300 component in electroencephalography (EEG) data can be detected when the target character is intensified.

Application of classical deep learning model, Convolutional Neural Network (CNN), in BCI for P300 component detection has been proposed in [5]. The results showed that the application of CNN outperformed other methods in limited channel number and epoch scenarios of the P300 detection. To avoid information loss between CNN layers and decrease the complexity of the training process, the application of CNN with batch normalization and residual blocks for P300 classification has been suggested [6]. Classification of ERP signals using dual input encoder CNN has been proposed for the analysis of time and spatial characteristics [7]. ERP signal classification using this model outperformed Support Vector Machine (SVM), CNN and other methods by several performance metrics. The usage of two different CNNs for feature extraction and fusion of the scores of each network for P300-based character recognition has been investigated in [8].

Methods of representing time and frequency information of EEG, Event-Related Spectral Perturbation (ERSP) and Inter-Trial Coherence (ITC) have been introduced for the feature extraction process [9]. Spectrograms created using Short-Time Fourier Transform (STFT) and CNN with extreme learning machines have been applied for seizure prediction [10]. A similar approach that involves spectrograms created using STFT, CNN and adaptive filtering has been used for the P300 component detection in order to improve the classification accuracy [11].

This paper proposes a Double Input Convolutional Neural Network with Feature Concatenation (DiCNN-FC) for the subject-independent P300 component classification. Two different spectrogram representations of EEG have been used as input for the proposed model. The performance of the proposed method is evaluated using two different datasets and compared with classical classification techniques used for P300 speller, SVM [12], Linear Discriminant Analysis (LDA) [13], and with CNNs based on the single spectrogram representation as input. The presented results demonstrate that the proposed method is well able to identify the P300 component across subjects based on single trial information.

The rest of the paper is organized as follows. Description of datasets, previous work and proposed methodology are given in Section II. Section III describes metrics used for evaluation and compares the proposed model with classical methods used for P300 classification, while Section IV gives concluding remarks.

II. Materials and Methods

A. Datasets

For training and testing of the proposed method, two EEG datasets of P300 BCI have been used [14], [15]. First is an open-access dataset that contains EEG recordings of 8 subjects, five male and three female, all suffering from ALS [14]. Recordings have been obtained using g.MOBILAB (g.tec, Austria) equipment from 8 channels according to the 10-10 standard (Fz, Cz, Pz, Oz, P3, P4, PO7, PO8), using left mastoid and right earlobe for grounding and referencing, respectively. During the EEG recording, subjects have been asked to

979-8-3503-3959-8/23/$31.00

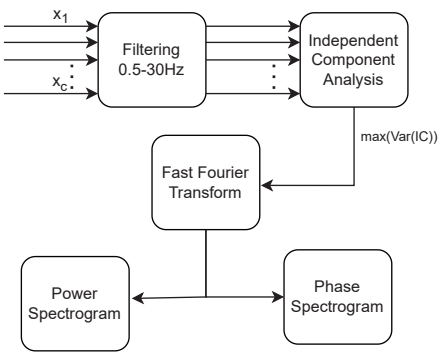


Fig. 1. Block diagram of pre-processing and spectrogram generation.

spell a predefined word using matrix GUI while focused on characters randomly intensified for 125 ms.

The second dataset used in this study is EEG recordings of healthy subjects - the Akimpech dataset [15]. This dataset contains EEG recordings of 30 subjects, recorded in 4 separate sessions. During the first two sessions, subjects have been asked to copy-spell predefined words. While on the rest of the sessions, they have been required to free-spell. EEG data has been recorded using g.USBamp (g.tec, Austria) from 10 channels (Fz, C3, Cz, C4, P3, Pz, P4, PO7, PO8, Oz).

B. Spectrograms Generation

Data pre-processing and spectrogram generation have been completed using the EEGLAB MATLAB toolbox [16]. The overall procedure of pre-processing and spectrogram generation is presented in Fig. 1. Raw EEG data has been bandpass filtered to eliminate the frequencies below 0.5 Hz and above 30 Hz using zero-phase digital filtering. Independent Component Analysis (ICA) has been applied to the filtered data, and the component with the highest variance has been used to generate a spectrogram containing essential information for classification. Examples of raw EEG data and obtained independent components are given in Fig. 2. The approach of computing the independent components by maximizing the entropy - Infomax ICA has been used for component extraction [17].

After applying ICA to the EEG, phase and power spectrograms have been created. Examples of spectrogram representations are given in Fig. 3. Each spectrogram has been created using the time frame that included information before and after the stimuli, making the duration of each image 825 ms (100 ms before, 600 ms after the stimuli). Created spectrograms of each type can be divided into two classes: first representing the target stimuli, and second representing non-target stimuli.

The power spectrogram indicates the change of power in different frequencies over time. A method used for ERSP spectrogram creation has been used to create the power spectrogram, with its limited version representing a single event in each spectrogram. It can be described as:

$$\mathrm{S_{PWR}}(f,t) = \frac{|\mathrm{FFT}(x(t))|^2}{\mu(f)}, \tag{1}$$

where FFT($x(t)$) is the Fast Fourier Transform of the signal $x(t)$, $|\cdot|^2$ denotes magnitude square of the quantity inside and $\mu(f)$ denotes the baseline, which is computed as:

$$\mu(f) = \frac{1}{n}\sum |\mathrm{FFT}(x(t'))|^2, \tag{2}$$

where n represents the number of samples for the baseline period, t' represents the baseline period.

The phase spectrogram represents phase shifts in different frequencies over time. It has been created using a method used for ITC spectrogram representation, limited by the single event representation which can be described as:

$$\mathrm{S_{PHS}}(f,t) = \tan^{-1}\frac{\mathrm{Im\{FFT(x)\}}}{\mathrm{Real\{FFT}(x)\}}, \tag{3}$$

where Im $\{\cdot\}$ and Real$\{\cdot\}$ denote real and imaginary parts of the quantity inside.

C. CNN with Ensemble Voting

The method of ensemble voting allows the combination of several algorithms for a single classification [18]. There are different approaches to combining the algorithms: an example is the weighted majority algorithm [19]. It proposes the solution for the pool of algorithms in a case when the algorithms' effectiveness on the given task is unknown. The weighted majority algorithm has been proven efficient with data that contained anomalies and remained robust while being used in the data with errors. The type of ensemble voting used in the study has been based on ensemble-averaged voting (EAV). Classification probabilities using EAV have been computed as follows:

$$\mathrm{P_{avg}}(X|y=1) = \frac{\sum_{i=1}^{N} P_i(X|y=1)}{\mathrm{N}}, \tag{4}$$

where P_{avg} is the average prediction result or EAV prediction, P_i is the prediction results of separate classifiers, and N is the total number of classifiers used for voting.

Spectrograms have been classified using CNN-based EAV, and two identical CNN models have been used to classify each time-frequency representation of EEG data. Combining two CNNs can provide high accuracy in P300 recognition [8]. Due to the availability of two representations that allowed the processing of additional features, CNN-ENS has been assumed to provide more stable and trusted classification results compared to CNN models that used only one of the given spectrograms.

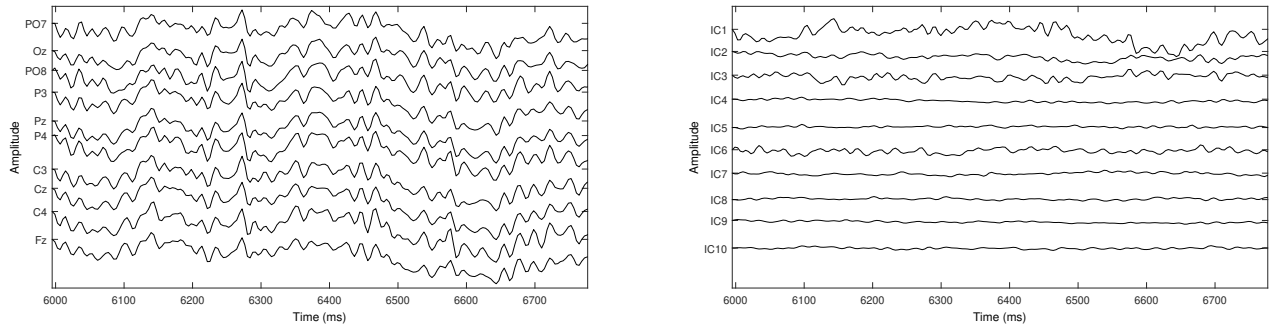
D. Proposed Dual-Input CNN with Feature Concatenation

This study proposes an extension of double input classification for the P300 speller using DiCNN-FC. The structure of the model used in the proposed approach is given in Fig. 4. Two different spectrogram representations of EEG signals are obtained and fed into the classifier model. The first step is EEG pre-processing and spectrogram generation, as explained earlier with reference to Fig. 1. The model consists of two branches for inputting two different time-frequency representations of EEG, and each branch consists of two pairs of convolution and max-pooling layers. In addition, at the end of each branch, resulting feature maps are flattened. Spectrograms fed into the model are compressed to 51×46×3 dimensions. The first convolutional layer consists of 3×3-sized 32 kernels. The resulting feature map of size 49×44×32 is fed into a max-pooling layer of size 2×2 for dimensionality reduction. The second convolutional layer with 32 kernels of size 3×3 in combination with the second max-pooling layer results in a feature map of size 11×10×32, which after flattening, becomes a feature vector of size 3520. Feature vectors from each branch are concatenated, and then concatenated feature vector is fed into the fully connected layer of size 256.

Each convolutional layer uses rectified linear unit (ReLU) as an activation function to eliminate negative values, being computed as:

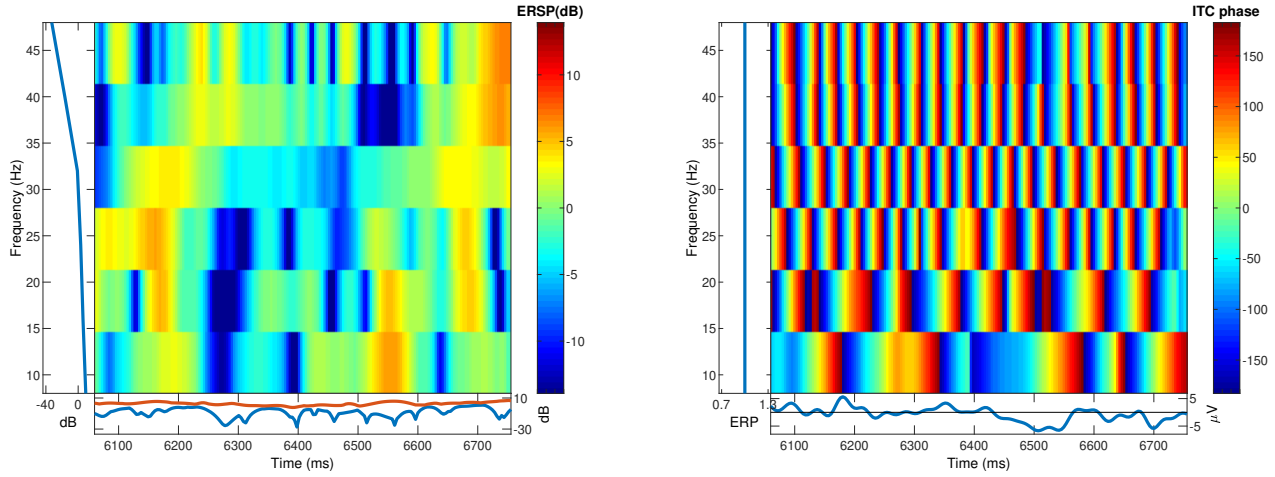
$$\mathrm{ReLU}(\tilde{X}_i) = \max(0, \tilde{X}_i), \tag{5}$$

where $\tilde{X}_i$ is the output data point of the convolution layer and ReLU is used to eliminate the negative values.



(a) (b)

Fig. 2. Time domain information used for spectrogram creation: (a) raw EEG portion with target stimuli, and (b) independent components obtained using ICA and sorted by variance.



(a) (b)

Fig. 3. Spectrogram representations used for classification: (a) **Power spectrogram**: The left panel displays the baseline log spectrum, while the red and blue lines at the bottom panel indicate the most positive and negative power values, respectively. (b) **Phase spectrogram**: the left panel displays the frequency means ITC, while the bottom panel displays the time-domain average of the input data.

The final layer of the model uses softmax activation that turns the output results of the final layer into a probability distribution. The softmax function is calculated as:

$$\text{softmax}(x_i) = \frac{e^{x_i}}{\sum_{k=1}^{K} e^{x_k}}, \tag{6}$$

where x_i is an output data point of the last layer.

The loss function used for the given task is categorical cross-entropy and is computed as:

$$\text{Loss} = \sum_{i=1}^{K} y_i \log(\hat{y_i}), \tag{7}$$

where K represents the number of classes, y_i represents true label or probability and $\hat{y_i}$ represents estimated probability. Negative logarithms of the estimated probabilities of each class are summed up to calculate how far the predicted probability is from the actual value.

III. Experiments and Results

This section gives details of the experiments carried out to demonstrate the effectiveness of the proposed method. The following methods have been selected for performance comparison, SVM [12], LDA [13], CNN-PWR and CNN-PHS, where CNN-PWR and CNN-PHS refer to the CNNs trained only using either power spectrograms or phase spectrograms, respectively. In addition, a comparison of obtained results with recent results published in different papers using the ALS dataset [14] is presented.

A. Performance Measures and Experimental Settings

The proposed model's performance has been evaluated by two metrics - accuracy and F1-score. Accuracy signifies the ratio of correct predictions to all predictions and is computed as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{8}$$

where TP and TN denote true positive and true negative, respectively, and FP and FN denote false positive and false negative predictions, respectively. The F1-score is another metric being computed as:

$$F_1\text{-score} = \frac{2 \times TP}{2 \times TP + FP + FN}, \tag{9}$$

which is better suited for cases when dealing with imbalanced data and differences in the importance of error between FP and FN.

Cross-validation techniques used during the experimental procedure can be described as "leave-one-subject-out". Seven subjects' data has been used for training with portions separated for the validation, while the testing procedure has been conducted using the unseen subject's data. The proposed model has been trained for 30 epochs with a batch size of 16. The training has been conducted using Python 3.8.10 with the virtual machine with hardware 2.2 GHz Dual-Core Intel Xeon with NVIDIA Tesla T4. The total training time of

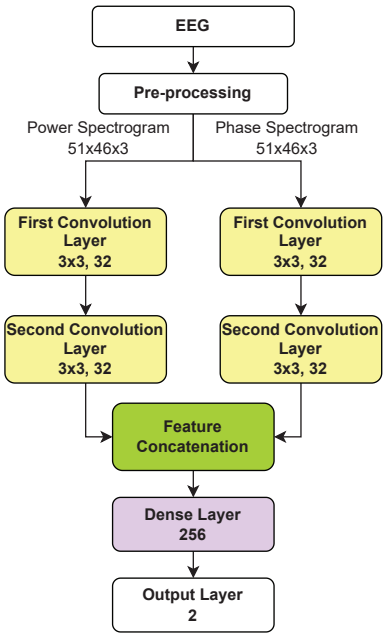


Fig. 4. Architecture of the proposed DiCNN-FC model.

the model for ALS patients' dataset reached 40.31 s for 30 epochs while the processing time of one input reached 4.75 ms.

B. Experiments for ALS patients dataset

The proposed and classical methods' validation and test results on the ALS dataset are presented in Table I. The proposed model classified P300 components with a test accuracy of 99.68 % on the dataset obtained from ALS patients. Time-series EEG data has been used as input for LDA and SVM. The best performance achieved using classical methods has been a test accuracy of 75.28% by LDA, inferior to the performance of the DiCNN with Power and Phase spectrograms. Each of the CNNs trained using one type of spectrogram performed well on the ALS dataset, CNN-PWR classifying 98.72% and CNN-PHS classifying 99.58% of spectrograms correctly. It can be observed that DiCNN-FC that combines both features produces somewhat better results.

It is worth mentioning that the proposed approach achieves better results than those reported in the literature on the same data set presented in Table II. For example, the classification of spectrograms created using STFT with CNN achieved a mean accuracy of 89.81% by averaging reference signals of samples ranging from from 500 to 2500 [11]. Weighted ensemble of LDA, SVM and kNN achieved an accuracy of 91.34% [20]. Recently, an SVM with an optimized score-based decision function has been able to achieve an accuracy of 94.4% on the same dataset of ALS patients [21].

C. Experiments for Akimpech dataset

Apart from performance evaluation using ALS data, the proposed model has been tested on the information of healthy subjects and the performance results of the proposed model on the Akimpech dataset are presented in Table III. DiCNN-FC achieved a validation accuracy of 81.60% and F_1-score 79.12% and test accuracy of 82.91% and F_1-score of 81.37%. The best performance of classical methods measured on healthy subjects has been achieved using SVM with a test accuracy of 72.37%. CNN-PWR and CNN-PHS achieved 72.94% and 80.43% test accuracy. DiCNN-FC provided the best results among compared methods in terms of test accuracy and second best in terms of test F_1-score.

TABLE I. Performance comparison for ALS patients' dataset

Model	Validation		Test	
	Accuracy (%)	F-1 score (%)	Accuracy (%)	F-1 score (%)
SVM	71.56	68.29	66.01	61.26
LDA	75.11	74.45	75.28	74.99
CNN-PWR	100.00	100.00	98.72	99.58
CNN-PHS	100.00	100.00	99.58	98.76
DiCNN-FC	**100.00**	**100.00**	**99.68**	**99.69**

TABLE II. Classification results of existing approaches, obtained using same ALS patients' dataset

Model	Data	Accuracy (%)	F-1 score (%)
Meng [11]	Time-Frequency	89.81	92.23
EML [20]	Time-series	91.34	90.40
OSBF M-SVM [21]	Time-series	94.40	-

TABLE III. Performance comparison for Akimpech data

Model	Validation		Test	
	Accuracy (%)	F-1 score (%)	Accuracy (%)	F-1 score (%)
SVM	73.59	69.89	72.37	67.52
LDA	73.71	70.03	68.39	70.54
CNN-PWR	71.21	73.65	72.94	76.56
CNN-PHS	79.28	77.40	80.43	80.23
DiCNN-FC	**82.91**	**81.37**	**81.60**	**79.12**

D. A Few Remarks

It is interesting to notice that the average performance of models drops for the Akimpech dataset [15] of healthy subjects, in comparison to the performance for the ALS patients' dataset [14]. Nevertheless, the performance on the second dataset shows the proposed model's ability to recognize the P300 component of a healthy subject. The recognition rate difference between the two datasets can be explained by the structural difference between the brains of the two groups. This difference is created by the progress of the disease, which can lead to widespread topological reorganization of the brain [22]. On this basis, there is a possibility that accuracy obtained on healthy subjects can be further improved by modifying the pre-processing step of the procedure. Examples of possible modifications are the application of empirical mode decomposition or simple averaging across channels. These methods have proven their efficiency on healthy data previously in literature [20], [23].

IV. Conclusion

This paper proposes a DiCNN-FC model for classification based on two spectrogram representations of EEG signals. In order to evaluate the subject-independence of the proposed model, several iterations of training and testing procedures (each time with a different test subject) have been carried out. The proposed DiCNN-FC using two spectrogram representations successfully outperformed existing methods for identifying P300 components of ALS patients. In addition, the simulations on healthy data show the proposed model's ability to classify healthy subjects' data. However, a side-by-side comparison of results achieved on ALS patients and the healthy dataset revealed a performance drop on the healthy dataset.

This work can be further extended by exploring the proposed method's efficiency in detecting P300 components of patients with cerebral palsy, peripheral neuropathy or other neuromuscular diseases. Moreover, there is a possibility for future extension of the study by focusing on processing and classification using data recorded in the real environment. In addition, the improvement of applied noise and dimensionality reduction algorithms to increase P300 recognition of healthy subjects' data can be explored in further studies.

REFERENCES

[1] N. Hasan, M. M. Hasan and M. A. Alim, "Design of EEG Based Wheel Chair by Using Color Stimuli and Rhythm Analysis," in *Proc. Int. Conf. Adv. Sci. Eng. Rob. Technol.*, 2019, pp. 1-4.

[2] H. I. Aly, S. Youssef and C. Fathy, "Hybrid Brain-Computer Interface for Movement Control of Upper Limb Prostheses," in *Proc. Int. Conf. Biomed. Eng. App.*, 2018, pp. 1-6.

[3] A. K. Singh, Y. K. Wang, J. T. King and C. T. Lin, "Extended Interaction With a BCI Video Game Changes Resting-State Brain Activity," in *Proc. IEEE Trans. Cogn. Develop. Syst.*, 2020, vol. 12, no. 4, pp. 809-823.

[4] L.A. Farwell, and E. Donchin,"Talking off the top of your head: Toward a Mental Prosthesis Utilizing Event-Related Brain Potentials," *Electroencephalogr. Clin. Neurophysiol.*, vol.70, no. 6, pp. 510-523, 1988.

[5] H. Cecotti and A. Graser, "Convolutional Neural Networks for P300 Detection with Application to Brain-Computer Interfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 433-445, March 2011.

[6] Z. Lu, Q. Li, N. Gao, T. Wang, J. Yang and O. Bai, "A Convolutional Neural Network based on Batch Normalization and Residual Block for P300 Signal Detection of P300-speller System," in *Proc. IEEE Int. Conf. Mechatron. Autom.*, 2019, pp. 2303-2308.

[7] D. Wen, Z. Wei, Y. Zhou, Z. Bian, and S. Yin," Classification of ERP Signals from Mild Cognitive Impairment Patients with Diabetes using Dual Input Encoder Convolutional Neural Network," in *Proc. IEEE Int. Conf. Comp. Intel. Virt. Env. Measur. Syst. App.*, 2019, pp. 1-4.

[8] S. Kundu and S. Ari, "Fusion of Convolutional Neural Networks for P300 Based Character Recognition," in *Proc. Int. Conf. Inf. Tech.*, 2019, pp. 155-159.

[9] D. Ming, X. An, B. Wan, H. Qi, Z. Zhang, and Y.Hu, "A P300-speller based on Event-Related Spectral Perturbation (ERSP)," in *Proc. IEEE Int. Conf. Sign. Proc., Comm. Comp.*, 2012, pp. 63-66.

[10] Y. Qin, H. Zheng, W. Chen, Q. Qin, C. Han, and Y. Che, "Patient-specific Seizure Prediction with Scalp EEG using Convolutional Neural Network and Extreme Learning Machine," in *Proc. 39th Chinese Cont. Conf.*, 2020, pp. 7622-7625.

[11] H. Meng, H. Wei, T. Yan and W. Zhou, "P300 Detection with Adaptive Filtering and EEG Spectrogram Graph," in *Proc. IEEE Int. Conf. on Mechatronics and Automation*, 2019, pp. 1570-1575.

[12] Q. Li, S. Ma, K. Shi and N. Gao, "Comparing the Classification Performance of Bayesian Linear Discriminate Analysis (BLDA) and Support Vector Machine (SVM) in BCI P300-speller with Familiar Face Paradigm," in *Proc. Int. Cong. Im. Sig. Proc., BioMed. Eng. Inform.*, 2016, pp. 1476-1481.

[13] X. Xiao, M. Xu, Y. Wang, T. P. Jung and D. Ming, "A Comparison of Classification Methods for Recognizing Single-trial P300 in Brain-Computer Interfaces," in *Proc. Ann. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2019, pp. 3032-3035.

[14] A. Riccio, L. Simione, F. Schettini, A. Pizzimenti, M. Inghilleri, M. O. Belardinelli, D. Mattia, and F. Cincotti, "Attention and P300-based BCI performance in People with Amyotrophic Lateral Sclerosis," *Front. Hum. Neurosci.*, vol. 7, 2013.

[15] C. Ledesma-Ramirez, E. Bojorges-Valdez, O. Yanez-Suarez, C. Saavedra, L. Bougrain, and G. Gentiletti. "An Open-Access P300 Speller Database," in *Proc. 4th Int. Brain-Computer Interface Meet.*, 2010.

[16] A. Delorme and S. Makeig, "EEGLAB: An Open Source Toolbox for Analysis of Single-Trial EEG Dynamics Including Independent Component Analysis," *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9–21, 2004.

[17] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. Wiley, 2001.

[18] A. Mussabayeva, Z. Ermaganbet, P. K. Jamwal and M. T. Akhtar, "Event-Related Spectrogram Representation of EEG for CNN-Based P300 Speller," in *Proc. Asia-Pacific Sig. Inform. Proc. Assoc. Annu. Summ. Conf.*, 2021, pp. 410-415.

[19] N. Littlestone and M. K. Warmuth, "The Weighted Majority Algorithm," *Inf. Comput.*, vol. 108, no. 2, pp. 212–261, 1994.

[20] A. Mussabayeva, P. K. Jamwal and M. T. Akhtar, "Ensemble Voting-Based Multichannel EEG Classification in a Subject-Independent P300 Speller," *Applied Science*, vol. 11, no. 23, 2021.

[21] L.Bianchi, C. Liti, G. Luizi, V. Piccialli, and C. Salvatore, "Improving P300 Speller Performance by Means of Optimization and Machine Learning," *Ann. Oper. Res.*, vol. 312, pp. 1221-1259, 2022.

[22] P. Sorrentino, R. Rucco, F. Jacini, F. Trojsi, A. Lardone, F. Baselice, C. Femiano, G. Santangelo, C. Granata, A. Vettoliere, M. R. Monsurrò, G. Tedeschi, and G. Sorrentino, "Brain Functional Networks Become More Connected as Amyotrophic Lateral Sclerosis Progresses: a Source Level Magnetoencephalographic Study," *NeuroImage. Clin.*, vol. 20, pp. 564–571, 2018.

[23] S. Shahnawaz, Z. Parveen, N. Noor and O. Farooq,"EMD Based Detection of Auditory Evoked Potential P300," in *Proc. Int. Conf. Electrical, Electronics Comp. Eng.*, 2019, pp. 1-5.