Fusion of Convolutional Neural Networks for P300 based Character Recognition

Sourav Kundu
Department of Electronics and Communication Engineering
National Institute of Technology, Rourkela
Odisha-769008, India
Email: souravkundu90@gmail.com

Samit Ari
Department of Electronics and Communication Engineering
National Institute of Technology, Rourkela
Odisha-769008, India
Email: samit@nitrkl.ac.in

***Abstract*—In this work, a brain-computer interface (BCI) system for character recognition is proposed based on P300 signal. Signal classification is the most challenging task in electroencephalography (EEG) signal processing as the amplitude of the EEG signal is low and it can be affected by the surrounding noise. The feature extraction is an important step for any classification task. The manually designed features are not sufficient to represent the signal properly due to the subject and surrounding environment variability. There is a need to develop feature extraction techniques which will extract high-level feature from the raw data automatically. In this work, two parallel CNN model with different kernel size is proposed to extract multi-resolution feature from the dataset. The proposed CNN model extracts spatial and temporal feature from the dataset. To mitigate the over-fitting problem, dropout is used before the fully-connected layer of the CNN architecture, which improves the network performance. The scores of these two CNN architectures are fused together for P300 detection. Also, ensemble of CNN (ECNN) architecture is proposed to reduce the variation between the classifiers and enhance the character recognition performance. The proposed method is tested on BCI Competition III dataset and the results are fairly comparable and better with the earlier methods.**

***Index Terms*—Brain-computer interface (BCI), convolutional neural network (CNN), electroencephalogram (EEG), P300.**

I. INTRODUCTION

Brain-computer interface (BCI) is a system that connects human brain with the machine. BCI system analyses the brain activity and generates the control command for external device [1]. A BCI system can assist elder, handicap people for their daily work. In general electroencephalogram (EEG) is acquired in non-invasive way, where the electrodes are place on the scalp of the subject. Non-invasive EEG signal acquisition is preferred as it is inexpensive and safe for the subject. The signal-to-noise-ratio (SNR) of the acquired EEG signal is low as the signal is collected in non-invasive manner [2]. Low SNR of EEG signal is the main challenge in BCI application. Signal averaging is a well established method to improve the SNR of the EEG signal. In this work, a P300 based BCI speller is proposed for character recognition. Due to random visual stimulus, a P300 wave appears in the EEG signal. This visual stimulus is presented by a BCI paradigm, which consist of several characters as shown in Fig. 1. The rows and columns of the character matrix intensified randomly, which generate P300 signal. For character recognition, P300 detection is an important task. From the detected P300 signal, row and column information is extracted, and the cross-section of row and column provides the position of the desired character.

For P300 based character recognition, feature extraction and classification are crucial steps. Several feature extraction and classification techniques are reported for P300 classification in [3]–[6]. It is important to extract efficient and relevant features from the dataset. Irrelevant features reduce the classifier performance and high dimensionality of the features create over-fitting problem. In [3], principal component analysis (PCA) based feature reduction technique has been reported for P300 classification. Temporal feature which provides the structural information about the EEG signal are used for P300 classification in [7]. For P300 signal detection, a feature reduction technique based on tensor is reported in [5] and it is referred as higher order spectral regression discriminant analysis (HOSRDA). A wavelet based feature extraction and sequential floating forward search for channel selection are reported in [4]. For P300 detection, different machine learning techniques like ensemble of support vector machine (ESVM), ensemble of weighted support vector machine (EWSVM) and ensemble of Fisher's linear discriminant (EFLD) are used in [3]–[5], [7], [8]. The above mentioned techniques use hand-crafted features, which can not represent the signal properly. To overcome the limitation of hand-crafted features, deep learning techniques are introduced as it extracts feature automatically from the dataset. Over the past few years, convolutional neural network (CNN), batch normalization neural network (BN^3), stacked autoencode (SAE) based techniques are reported for BCI application in [2], [9]–[12]. In [9], seven different CNN based models are reported and among them, CNN-1 and multi-classifiers convolutional neural network (MCNN-1) show superior performance compared to other models. The amplitude and latency of the P300 signal varies due to the physical condition of the subject. Fixed length of the CNN kernel is insufficient to tolerate these changes.

In the proposed method, a novel fusion of CNN network is proposed for P300 based character recognition. In the developed model, two different CNN networks with two different dimension of kernels are used for multi-resolution feature extraction. The score of these two models are added together for P300 detection. Ensemble of CNN model is used as ensemble of classifier reduce the classifier variability and enhanced the classification performance. Due to the BCI paradigm, the training dataset is unbalanced which reduce the classifier performance. The training dataset is divided into five equal parts, to balance the dataset. The proposed model combines the feature extraction and classification step into a single process, which reduces the computational complexity of the network.

The rest of the paper is organized as follows: Section II describes about P300 speller paradigm and the BCI Competition dataset which is provided by BCI Laboratory of the Wadsworth Center, NYS Department of Health. The proposed framework for character recognition is explained in Section III. Section IV shows the experimental results and discussions about the developed model. Finally, the work is concluded in Section V.

II. The Data Set

A BCI paradigm and a data acquisition system are required to collect EEG data from the scalp. Random intensification of BCI paradigm generates a positive peak in the EEG signal. In this work, the dataset is provided by the organizer of the BCI III competition.

A. BCI Paradigm

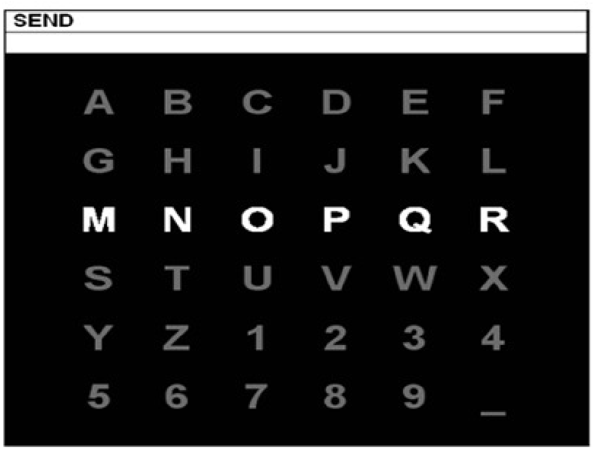

Fig. 1. P300 speller Paradigm [13]

In 1988, a row-column paradigm is reported by Farwell and Donchin for P300 based character recognition [14] In BCI Competition III, the BCI speller paradigm used for visual stimulus is shown in Fig. 1. It has 6 rows and 6 columns which are intensified randomly. The intensification rate of the BCI paradigm is 5.7 Hz. The matrix is on for 100 ms and remain blanked for 75 ms. In one round, there are 12 intensifications, out of which only two intensifications contained the desired character.

B. Database Used

The data set II of BCI competition III is used here [13]. It consists of two subjects data. Each subject has 85 training characters and 100 testing characters. In one round, there are 12 intensifications and it is repeated 15 times for single character. This one round of flashing is referred as epoch. For one character of 15 epochs, there are $12 \times 15 = 180$ flashing. There is a gap of 2.5 s after each character data is collected. A 64 channels data acquisition system is used for EEG data recording. The collected EEG signal is bandpass-filtered with cut-off frequency of 0.1 - 60 Hz and the signal is samples at the rate of 240 Hz.

III. Proposed Framework

A. Preprocessing

In preprocessing stage, 667ms of data samples are collected form the each channel after the stimulus. It is postulated that a time windows of 667ms is enough to capture the necessary information about the P300 signal. A 8^{th} order Chebyshev band-pass filter is used to filter the each channel data. The cut-off frequency of the filter is between 0.1 and 10 Hz. These post-stimulus signal consists of 160 samples from each channel. Thus the feature dimension of the sample is 64×160. Due to the design of BCI paradigm, the number of not-P300 signal is five times more than the P300 signal. A binary classifier performs well when the distribution of P300 signals and not-P300 signals are equal. To solve this issue, the P300 signals are cloned four times. Then, the training data is divided into five equal parts. In each part, there is all the P300 signals and 1/5 of not-P300 signals.

B. Fusion of Convolutional Neural Network

In general, convolution layers and the fully connected (FC) layers are the main block of CNN architecture. The proposed CNN architecture is discussed in this section. Two different CNN models with two different kernels size are proposed in this work to extract multi resolution feature from the dataset. There is two convolutional kernels in the proposed CNN model.

TABLE I
Layers description for the proposed CNN architecture of CNN-20

Layer Name	Kernel/ Stride	Input	Feature map
Batch normalization	-	64×160	-
Convolution	$[64 \times 1]/[1 \times 1]$	64×160	16
Convolution	$[1 \times 20]/[1 \times 20]$	1×160	16
Batch normalization	-	1×8	-
ReLU	-	-	-
Dropout	-	-	-
Fully connected	-	128	-
Dropout	-	-	-
Softmax	-	128	-

The convolution is performed across spatial domain in first layer and in second layer, the convolution is performed across time domain. Both the spatial and temporal information of the EEG data are extracted and fed to FC layer. In the last layer, a softmax layer is used for classification. Batch normalization (BN) [15] layer is adopted in CNN framework as it reduces the internal covariant shift which accelerates the training process. The BN layer is used before first convolution layer and after

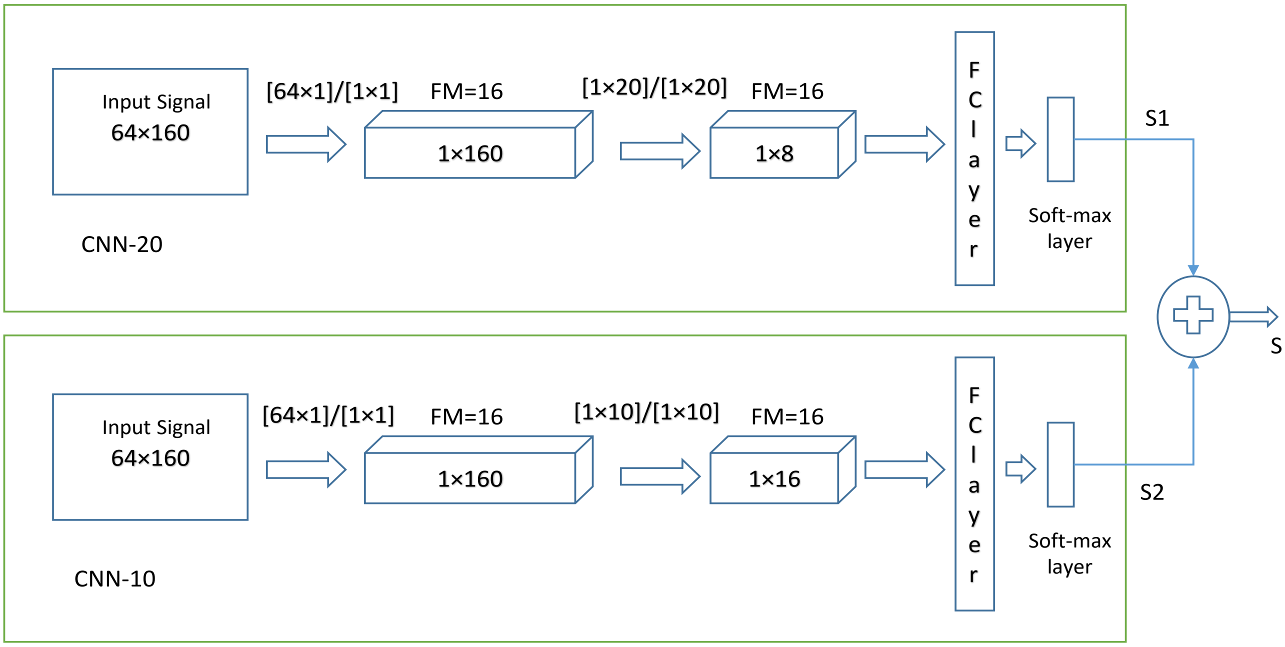

Fig. 2. Flowchart of proposed fusion of multi kernel convolutional neural network. The scores of CNN-10 and CNN-20 are fused together for character recognition. FM represents the feature map.

second convolution layer. To introduce the non linearity in the CNN, rectified linear unit (ReLU) is used as activation function after the second BN layer as it is transferring the negative value of the neuron to zero. The ReLU function is defined as $ReLU(x) = max(x, 0)$. To address the over-fitting problem in CNN, dropout [16] technique is adopted which randomly drops the nodes of neural network during training.

TABLE II
LAYERS DESCRIPTION FOR THE PROPOSED CNN ARCHITECTURE OF CNN-10

Layer Name	Kernel/ Stride	Input	Feature map
Batch normalization	-	64×160	-
Convolution	$[64 \times 1]/[1 \times 1]$	64×160	16
Convolution	$[1 \times 10]/[1 \times 10]$	1×160	16
Batch normalization	-	1×16	-
ReLU	-	-	-
Dropout	-	-	-
Fully connected	-	128	-
Dropout	-	-	-
Softmax	-	128	-

The proposed CNN model is shown in Fig. 2 and the detail layers description of CNN models are provided in Table I and Table II. In first convolution layer, the convolution is performed across spatial domain and the dimension of the kernel size is $[64 \times 1]$ for both the models. In second convolution layer, the convolution is performed across time domain and the kernels dimension are $[1\times20]$ and $[1\times10]$. The CNN models with kernels size $[1\times20]$ and $[1\times10]$ are referred as CNN-20 and CNN-10, respectively. These two different dimension of kernels extract different resolution feature from the data. The dimension of the kernels are chosen empirically. In second convolution layer, the stride of the the kernels are 20 and 10 for CNN-20 and CNN-10, respectively. The striding of the kernels are non-overlapping, which are used to reduce the feature dimension. After these two models are prepared, the scores of the softmax layers are fused together for character recognition. Let CNN-20 and CNN-10 provide score of S1 and S2, respectively. Then the fused score (S) is defined as

$$S = S1 + S2 \tag{1}$$

C. Ensemble of Convolutional Neural Network

In Ensemble of CNN (ECNN) classifier, the score of each classifier ensembles together as it reduces the classifier variability [17]. In BCI application, a simple averaging rule shows better performance compared to the voting strategy as mentioned in [7]. The training data is divided into five equal parts as mentioned in the preprocessing stage. From each part of the data, a CNN model is designed for classification. The proposed ECNN model is shown in Fig. 3.

For testing, the test data is passed through all the five classifier as shown in Fig. 3. Let, the k^{th} classifier assigns a fused score of S_k, then the ECNN score (f_{avg}) is written as

$$f_{avg} = \frac{1}{J} \sum_{k=1}^{K} \sum_{j=1}^{J} S_k \tag{2}$$

where number of epochs and classifiers are represented as J and K. After each epoch, the desired position of row and

TABLE III
CHARACTER DETECTION PERFORMANCE (IN %) OF THE PROPOSED METHOD

	Epochs														
Subject	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A	18	34	51	57	66	68	74	76	82	88	88	92	96	96	98
B	40	55	66	75	80	86	87	87	91	93	94	96	93	94	95
Mean	29.0	44.5	58.5	66.0	73.0	77.0	80.5	81.5	86.5	90.5	91.0	94.0	94.5	95.0	96.5

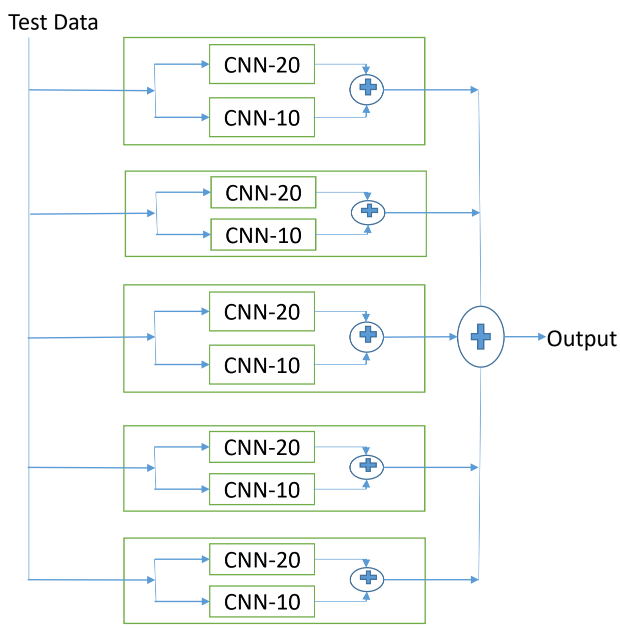

Fig. 3. Block diagram of proposed ensemble of convolutional neural network.

TABLE IV
CONFUSION IN CHARACTER RECOGNITION USING PROPOSED METHOD AFTER 15 EPOCH

Subject	Expected	Predicted
A	O	I
	Q	P
B	H	Z
	P	Q
	W	V
	T	1
	4	3

TABLE V
PERFORMANCE COMPARISON (IN PERCENTAGE) OF THE DEVELOPED METHOD WITH STATE-OF-THE-ART METHODS FOR 5, 10 AND 15 EPOCHS

Method	Epochs		
	5	10	15
ESVM [7]	73.5	87.5	96.5
CNN-1 [9]	70.0	88.5	94.5
MCNN-1 [9]	69.0	87.0	95.5
WT-EFLD [4]	71.5	87.5	95.0
HOSRDA+LDA [5]	72.5	89.0	96.5
Proposed method	73.0	90.5	96.5

column are predicted as follows:

$$\begin{aligned} C_{pos} &= \underset{1\leq jj\leq 6}{\arg\max}\ f_{avg}(jj) \\ R_{pos} &= \underset{7\leq jj\leq 12}{\arg\max}\ f_{avg}(jj) \end{aligned} \tag{3}$$

where C_{pos} and R_{pos} are the estimated position of column and row of the speller matrix, respectively. The intersection of row and column provides the character's position.

IV. RESULTS AND DISCUSSION

The character recognition performance of the proposed method is shown in Table III. From the table, it is observed that the proposed method achieves highest accuracy of 98.0% after 15 epochs for subject A and for subject B, the proposed method achieves highest accuracy of 96.0% after 12 epochs. The average highest accuracy of the proposed method is 96.5% after 15 epochs. The confusion in character recognition after 15 epochs is shown in Table IV. From the table, it is noticed that the wrongly predicted characters belong to the same row, column or near to the expected characters. Its means, one row or column of the expected character is detected correctly, or the attention of the subject shifted to the nearest character.

A comparison between proposed method and earlier reported techniques [4], [5], [7], [9] for 5, 10 and 15 epochs is shown in Table V. The developed method achieves 73.0%, 90.5% and 96.5% accuracy after 5, 10 and 15 epochs, respectively. It is noticed from the table that after 5, 10 and 15 epochs, the developed method shows better performance compared to the earlier reported techniques [4], [5], [9]. The proposed method shows better performance after 10 epochs, equal performance after 15 epochs and the performance is comparable after 5 epochs compared to [7]. In [4], wavelet based features are extracted from the dataset, which is handcrafted feature. Therefore, it is unable to represent the EEG signal perfectly. Deep learning technique is applied in [9], but it does not extract the multi-resolution feature from the dataset. In [7], temporal features are used for classification and a recursive channel selection procedure is reported for feature reduction, which is a time consuming task. In [5], HOSRDA based manual feature extraction technique is reported, whereas the proposed method extracts multi-resolution deep feature automatically from the dataset and ensemble of classifier technique is proposed to bring down the classifier variance.

V. CONCLUSION

In this work, a novel deep learning method has been proposed for P300 based character recognition. The CNN based deep learning architecture extracts spatial and temporal features from the dataset for P300 detection. Two parallel CNN network is proposed to extract multi-resolution features from the dataset. The scores of these parallel network fused together for P300 detection. Ensemble of CNN architecture is developed as it reduces the classifier variability and improve the recognition performance. In this work, the feature extraction and classification are performed in a single step, which is effective for on-line BCI system.

REFERENCES

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain–computer interfaces for communication and control," *Clinical neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002.

[2] M. Liu, W. Wu, Z. Gu, Z. Yu, F. Qi, and Y. Li, "Deep learning based on batch normalization for p300 signal detection," *Neurocomputing*, vol. 275, pp. 288–297, 2018.

[3] S. Kundu and S. Ari, "P300 detection with brain–computer interface application using pca and ensemble of weighted svms," *IETE Journal of Research*, vol. 64, no. 3, pp. 406–414, 2018.

[4] M. Salvaris and F. Sepulveda, "Wavelets and ensemble of FLDs for P300 classification," in *2009 4th International IEEE/EMBS Conference on Neural Engineering*. IEEE, 2009, pp. 339–342.

[5] M. J. Idaji, M. B. Shamsollahi, and S. H. Sardouie, "Higher order spectral regression discriminant analysis (hosrda): A tensor feature reduction method for erp detection," *Pattern Recognition*, vol. 70, pp. 152–162, 2017.

[6] S. Kundu and S. Ari, "P300 based character recognition using sparse autoencoder with ensemble of SVMs," *Biocybernetics and Biomedical Engineering*, vol. 39, no. 4, pp. 956–966, 2019.

[7] A. Rakotomamonjy and V. Guigue, "BCI competition III: dataset II-ensemble of SVMs for BCI P300 speller," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 3, pp. 1147–1154, 2008.

[8] S. Kundu and S. Ari, "P300 Detection Using Ensemble of SVM for Brain-Computer Interface Application," in *2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*. IEEE, 2018, pp. 1–5.

[9] H. Cecotti and A. Graser, "Convolutional neural networks for P300 detection with application to brain-computer interfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 3, pp. 433–445, 2011.

[10] G. B. Kshirsagar and N. D. Londhe, "Improving Performance of Devanagari Script Input-Based P300 Speller Using Deep Learning," *IEEE Transactions on Biomedical Engineering*, 2018.

[11] S. Sakhavi, C. Guan, and S. Yan, "Learning temporal information for brain-computer interface using convolutional neural networks," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 11, pp. 5619–5629, 2018.

[12] S. Kundu and S. Ari, "P300 based character recognition using convolutional neural network and support vector machine," *Biomedical Signal Processing and Control*, vol. 55, p. 101645, 2020.

[13] B. Blankertz, K.-R. Muller, D. J. Krusienski, G. Schalk, J. R. Wolpaw, A. Schlogl, G. Pfurtscheller, J. R. Millan, M. Schroder, and N. Birbaumer, "The BCI competition III: Validating alternative approaches to actual BCI problems," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 14, no. 2, pp. 153–159, 2006.

[14] L. A. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalography and clinical Neurophysiology*, vol. 70, no. 6, pp. 510–523, 1988.

[15] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[16] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[17] Y. Grandvalet, "Bagging equalizes influence," *Machine Learning*, vol. 55, no. 3, pp. 251–270, 2004.