Cross-Subject and Cross-Paradigm learning using Convolutional Neural Network for P300 Event-Related Potential Detection

Amir Mohammad Mijani[1*], Aref Einizade[1,2], Mohammad Bagher Shamsollahi[1] and Behrad Taghi Beyglou[1]

[1]BiSIPL, Department of Electrical Engineering, Sharif university of Technology, Tehran, Iran

[2]HMIL, Department of Electrical Engineering, Sharif university of Technology, Tehran, Iran

[*]**Corresponding author:** Amir Mohammad Mijani, BiSIPL, Department of Electrical Engineering, Sharif university of Technology, Tehran, Iran, Tel: +989106763930; E-mail: amirmohamad_mijani@yahoo.com

Abstract

Background: P300 Speller systems conventionally are using an oddball pattern which results in P300 component. By using the P300 component, mentally disabled can spell different characters.One of the most disadvantages of ERP-based BCI systems, especially P300 spellers, is the need for a large amount of training data which is time-consuming and exhausting for users.

New method: The goals are to evaluate the possibility of Transfer Learning (TL) implementation using the fine-tuning technique on convolutional neural networks (CNN) by two different approaches: 1) Cross-Subject, and 2) Cross-Paradigm. In cross-subject, data of individual paradigm cross different subjects and in cross-paradigm, data of individual subject cross different paradigms were applied.

Results: The final results illustrate that the amount of network training data is reduced up to 75 percent. Furthermore, the average of character detection accuracy using CNN is increased 11.76%, 13.95% and 13.51% in cross-subject TL in comparison to LDA classifier for single, dual and triple paradigms respectively. Also, such accuracy is increased by 6.76%, 10.95% in cross-paradigm TL in comparison to LDA classifier for dual and triple paradigms respectively.

Comparison with existing methods: Cross-subject and the novel cross-paradigm suggested in this study reduce the amount of needed training data in comparison to existing subject-dependent methods. In addition, the performance overall was improved against LDA in TL condition.

Conclusions: Overall in cross-subject and cross-paradigm TL approach using CNN, character detection accuracy was improved in comparison to LDA and the amount of training data was decreased significantly.

Introduction

The Brain-Computer Interface (BCI) is a direct communication channel between the human brain and the external device. These systems allow people to communicate with the outside world without any movement and only through measuring brain activities [1-3]. BCIs may be the only possible way for people with disabilities, especially those suffering from spinal cord injury [3,4]. In recent years, BCIs have also been significantly developed for healthy people in areas such as the environment [5], lie detection [6] and computer games [7]. For each BCI system, two main tasks are defined: 1) first, it must detect the purpose of the subject from brain signals; 2) it must translate the detected target to the executable commands for the output of a device [8]. Among non-invasive methods for recording brain activity, Electroencephalography (EEG) is attractive for many researchers in the field of BCI, due to the available capacity of recording, cheapness and, most importantly, the brilliant temporal [9,10].

As mentioned, one of the crucial parts in each BCI system is the proper detection of the subject's purpose from the brain signals. Therefore, according to different paradigms of neural systems, BCI systems can be categorized to various groups, such as systems based on SSVEP [11], ERP [12] and Motor Imagery [13]. In this paper, we focus on BCI systems based on EPR. In such systems, detection of subject's purpose turns to an EEG signal classification problem using algorithms and techniques of pattern recognition. Among the traditional methods used for classification of the EEG signal, we can mention SVM [14], LDA [15], and Hidden Markov Model [16]. Unfortunately, manual extraction and selection of features and the use of traditional methods have been accompanied by problems. For example, since the EEG signals have cross-subject variations (sometimes severe), then methods that lead

to good results for a subject may not perform well for other subjects, and this is one of the limitations of traditional algorithms and manual feature extraction [10].

In recent years, deep neural networks, especially convolutional neural networks, have been considerably used, and have shown high performance in the field of pattern recognition such as image recognition [17] and Natural Language Processing (NLP) [18]. Convolutional neural networks (CNNs), can automatically learn and extract features from raw data, these networks, in addition to the knowledge we have from the data, can give us unknown information about the data. The advantage of automatic learning has made the results of deep CNNs much better than traditional methods.

The use of neural networks for the first time in order to detect the readiness potential shows that they can also be used to classify EEG signals. Therefore, in recent years, researchers have been very interested in designing such networks to use in BCI, because CNNs are capable of improving identification and classification of EEG signals significantly and overcome the difficulties of traditional machine learning methods [19-23].

There is evidence which proves that deep learning can be very successful in EEG signal processing: 1) convolutional neural networks provide the ability to detect spatial relationships within data, 2) filtering and classification in neural networks can be combined together to create a discriminatory structure; and 3) improvement in the architecture of the recurrent neural networks (RNN) and the appearance of the LSTM model have provided the ability to detect temporal relationships of the data [24]. One of the significant advantages of convolutional neural networks is the automatic feature extraction that is performed without the need for specific knowledge about data. In other words, features extraction is a part of the learning process and is not necessary to classify the extracted features like the traditional classifier.

Despite the dramatic development of neural networks and their general application in various fields, the main disadvantage of using these networks is that they require a large amount of training data to train the network appropriately. This problem has been doubled in the processing of EEG signals. Because providing a large amount of the EEG data requires several recording sessions, which is a time consuming and tedious task for the subject. On the other hand, the nature of BCI systems is generally dependent on the subject, in other words, each system must be trained using the data of specific subject and can be tested for the same subject. To overcome the mentioned problems, many transfer learning (TL) algorithms, especially in EEG-based BCI systems, have been proposed [25-27]. Our aim in this study is to use CNNs with fine-tuning technique in order to implement transfer learning. Therefore in this paper everywhere we mention TL implementation means the use of CNN with fine-tuning technique. The TL implementation used in the current study is in two scenarios: (1) a cross-subject method that has been applied a lot so far. In this case, the various subject's data are used to train the network. Therefore, the required training data for each target subject is significantly reduced, and each subject's network can be trainable with much less training data. 2) The second method of TL, introduced here for the first time, is called cross-paradigm. In this scenario, data according to three different protocols, all of which are P300-based, are available. Two protocol data are used for training the network, and the third protocol is used as test. In the end, the results are compared with LDA as a traditional classifier.

In the first part of the article, the introduction was presented. Further, in section 2, we will review the literature and previous related work and explore several deep learning networks used to classify the EEG signal. In section 3, we will become more familiar with p300 component, the speller systems, and the RSVP paradigm. In section 4, the architecture of the convolutional neural network in the fine-tuning mode for transfer learning purpose is reviewed. In section 5, the results of classification with cross-subject and cross-paradigm transfer learning scenarios have been reported using convolutional neural networks and LDA. Finally, section 6 ends with the conclusion.

Related works

CNN architecture is designed to perform well for image classification tasks (two dimensional). The use of convolutional neural networks requires a specific structure and format of the input in order to process the EEG signal. Cecotti [28] introduced a structure which EEG signal samples are arranged in a matrix form (N × C) and fed as input to the network for the first time. In the mentioned matrix's structure, N is the number of time samples, and C represents the number of electrode channels used for EEG signal recording. This architecture has three main layers: 1) The first layer acts as a spatial filter and learns to select more important channels (channel selection operation); 2) The next layer plays the role of the temporal filter. It selects the prominent temporal features (feature selection function), and 3) Fully-connected layer identifies the connection between the selected features and corresponding classes of each sample. This architecture has been tested on the BCI competition III dataset, and the average accuracy of its character detection for two subjects has been reported about 94.5%.

Also, Manor [29] implemented a CNN for classification of a single trial EEG, where the network's input was in the form of a matrix (N × C). This network has become more complex and deeper than the Cecotti architecture, in addition to increasing the number of layers; the numbers of neurons in convolutional and fully-connected layers are also increased. In such a complex network, pooling techniques in convolutional layers, Dropout in the fully-connected layers and Batch normalization layer have been used in order to reduce over fitting and increase accuracy. The data used in this network is derived from a P300-based RSVP paradigm with Image Search application.

Liu used the CNN architecture in order to detect and classify the P300 component [30]. The architecture of this network is

very similar to the Cecotti architecture; that has a layer as a spatial filter; one filter as a temporal filter and two fully-connected layers. Batch normalization and Dropout have been used in order to prevent over fitting. BCI competition III dataset II is used as data.

In work had been done in [10], the neural network architecture was also used to classify the EEG signal and detect the P300 component. The proposed network is called the One Convolutional Layer Neural Network (OCLNN). They showed that with just one Convolutional layer, they could accurately detect the P300 and the character on three datasets, which are all the P300-based speller systems. They claim that the accuracy of the character detection using their network has been increased by around 19.35% in comparison to the others.

In article [24], the combination of CNN and RNN architectures (such as LSTM) has been used to classify the P300 component. In this study, Deep Learning Networks have also been used to implement cross-subject transfer learning scenario using the fine-tuning technique in CNN and in comparison to commonly used methods such as LDA, results have been significantly improved. The dataset used in this network is derived from a RSVP paradigm using the P300 speller.

Ori Tal has applied CNN [31], RNN and their combinations onthe data obtained from [15] (data corresponding to RSVP paradigm based on P300 speller) and in comparison to the results with LDA classifier, CNN leads to higher accuracy. Also, applying various Deep network architectures, have shown that combining CNN and LSTM with the implementation of the transfer learning scenarios improved the results compared to the LDA.

In most of the mentioned articles, like our paper, the data was derived from RSVP-based paradigms.

Background

P300 Speller:

Event-Related Potentials (ERP) are the brain's response to external stimuli [32]. Among the ERP components, p300 is known as the most prominent component of ERP [33,34]. Its amplitude and latency characteristics characterize the P300 component [35]. This component appears with a positive deviation at latency of about 300-500ms after the onset of target stimulation, and this component is detectable using EEG. The P300 component is appeared in an oddball paradigm and is more clearly visible in the electrodes Cz, Pz, and Fz with a stronger amplitude [36,37]. There had been introduced several ERP-based BCI systems which P300 speller is one of the most popular ones.

Farwell and Donchin introduced the p300 speller first time, known as Matrix Speller [38]. In their protocol, they have 36 characters in a 6 × 6 matrix. Each row and column is intensified randomly, and the user is asked to focus on the character which he wants to spell. When the row and column containing the target character is intensified, the P300 component is

appeared and using the recognition of the P300 component, one can specify the row and column that the user has focused on and according to this recognition, target character is detected. Even though the Matrix Speller was the first speller system and was highly usable after, researchers have shown that selecting the target character in this system depends on the focus and the gaze of the subject. According to this fact, this system is not suitable for patients with eye movement problems (oculomotor control) [15]. Several solutions have been proposed to solve the mentioned problem. One of them is changing the display paradigm. One of the paradigms that have been widely preferred rather than Matrix Speller is called RSVP.

RSVP:

Unlike the Matrix Speller, in RSVP paradigms the stimuli appear one by one in the middle of the screen randomly. The selection of the target character in this paradigm is not gaze-dependent [9,15,37,39,40]. Although the proposed RSVP model could overcome the Gaze dependency in a matrix speller well, displaying characters one by one results in increasing the experiment period and sharp decreasing in information transmission rate (ITR). Let us call the First RSVP paradigm, which is introduced in [39], single RSVP paradigm. The multi RSVP patterns were introduced [41-44] to overcome Single RSVP problem. In this study, we used the dataset of two new paradigms called Dual and Triple RSVP as recording paradigm in addition to single RSVP paradigm, and the results were compared with the single RSVP paradigm. In the Dual RSVP pattern, there are two characters at a time slot, and in triple RSVP there are three characters appear together in the middle of the screen, at the same time. In fact, in the Dual and Triple RSVP pattern, two and three character strings are shown simultaneously, with the difference that the second and third strings are arranged with a specific delay of the first string. For each of the stimuli shown in the Dual RSVP pattern, the subject should look at the left side (left character), and when he sees the target, he should focus to the right side to find target character on the right-side strings. Finally, the subject turns his look back on the left side character, and the process continues to show all the stimuli. Similar to this paradigm, the Triple RSVP pattern is implemented (see [44] for further details).

EEG Dataset:

In this article, the dataset of the three protocols containing: Single, Dual, and Triple RSVP paradigm has been used. For each of these protocols, the signals were recorded from three participants. All subjects voluntarily took part in these experiments, and data was recorded in The National Brain Mapping Lab (NBML). We present a short description of the three datasets in the following.

Each of the three protocols was tested for spelling 45 different characters. These 45 characters are spelled in 15 distinct runs. In each run, three characters are spelled and between the two characters a few seconds and between two runs few minutes are devoted to subject as relaxation time.

In the Single RSVP protocol, 26 letters of the English alphabet are used as stimuli, and displaying these 26 stimuli repeats ten times for spelling each character. Therefore, in Single RSVP protocol, for each character, 260 stimuli are displayed, among which ten stimuli are target, and the rest are non-target. In this case, the duration of each stimulus is 187.5ms.

In the Dual RSVP protocol, 29 characters containing 26 letters of the English alphabet and three punctuation marks (.?!) have been used as stimulus. Twenty-nine stimuli are displayed with five repetitions. Out of 145 stimuli, ten stimuli are targets, and the rest are non-target, and each stimulus period is 250ms.

In the Triple RSVP protocol, 35 characters consist of 26 English alphanumeric characters and nine numbers (from 1 to 9). For spelling each character, these 35 stimuli are displayed three times. Therefore, out of 105 stimulates, nine stimuli are target, and the rest are non-target. The time to display any stimulus in this protocol is 250ms. The specifications of these three protocols are presented in **Table 1**.

Table 1: Specification of different paradigms.

Paradigm	Number of all spelled letter	Number of all stimuli	Number of P300	Number of all non-P300
Single RSVP	45	45×260	45×10	45×250
Dual RSVP	45	45×145	45×10	45×135
Triple RSVP	45	45×105	45×9	45×96

Methods

In this paper, CNN structure was used for classification. In Section 4.1, Input structure of the network is described. Section 4.2 presents the overall structure according to different scenarios and implementing the convolutional neural networks with the approach of transfer learning. In section3, application of CNNs in the field of character detection is investigated.

Data Preparation for CNN:

Input data is a matrix with the size of C × N. N is the number of temporal samples, and C is the number of EEG recording channels. Also, $N = T\_s × F\_s$, where T_s and F_s are the time interval from 0 to 1 second after stimulus onset and sampling frequency, respectively. The overall structure of the network, in which the input size is specified, is shown in **Figure 1**.

Input tensor is filtered with a band pass filter with passing frequencies 0.2 to 40 Hz, in order to eliminate the high-frequency noise. Then the signals of each electrode changed to zero-mean and unit variance during the normalization process.
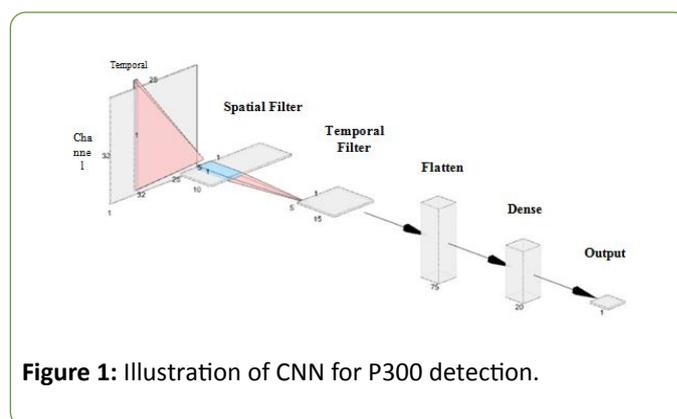

Figure 1: Illustration of CNN for P300 detection.

Network training:

Since the general structure of the network, especially the convolutional neural networks, is susceptible, several cases have been tested to achieve an optimal structure. For this purpose, the overall structure of the network is presented parametrically, and in the results section, the numerical values of these parameters will be mentioned. Each layer has some kernel (filter). Size of first layer filters according to the number of EEG channels and the second layer kernels size also, if any, determine the number of down-sample rate of the input signal.

In order to train the network, the cost function for our proposed network is considered as binary cross-entropy. This function uses the Adam optimizer (with learning rate equal to 0.001) to solve the optimization problem. Batch size and training epochsare set to 128 and 10, respectively. Weights and biases of all neurons of the convolutional layer are regularized using L2-Regularizer.

In this paper, in order to detect the P300 component, CNN has been trained and evaluated with the TL approach in two general modes: cross-subject and cross-paradigm. In the cross-subject scenario, we trained the network on two subjects in a particular protocol and evaluated network with the data of the third subject in the same protocol. In the cross-paradigm scenario, we considered data from two protocols of a specific subject as training data and then evaluate the network with the data from the third protocol of the same subject. Details of network training in each of these scenarios are mentioned below.

Cross-subject:

Training of the network with the approach of transfer learning is done in two steps. First, we train the network with data from two subjects of a specific paradigm. Then, in the second step of the training, the convolutional layer weights (initial layers) are frozen, but we keep all weights of the fully-connected layer trainable. After the division of Third subject's data into k parts, we train the weights of the fully-connected layer of the previous network using one of these k parts in the same protocol and evaluate it on the k-1 sections (fine-tuning technique) and this process will continue as long as all k parts are used once as training data. Thus, it can be claimed that the

amount of training data for the third subject was reduced dramatically, and in some way, the training of the network was transferred to the data of the third subject (TL approach).

Cross-paradigm

Similar to the previous section, network training is done in two steps, with the difference that in the first step, data from the two protocols of a specific subject was used for training, the second step of training is done on one part of k sections related to the same subject in the third protocol. Finally, we evaluate the network similar to the previous section (4.3.1). In the structure of network presented in the previous section, the goal is to transfer learning between the data of different subjects from a common protocol (cross-subject Transfer Learning), but in this case it is desirable to transfer training of the network to data from different protocol in a specific subject (cross-paradigm Transfer Learning). The main goal of the present paper is the implementation of the TL approach in the cross-paradigm scenario.

Results

First, we will describe the experimental setup, then in Section 5.2, the details of the implementation of the cross-subject transfer learning approach will be described, and the results of two classifiers; CNN and LDA will be compared. In Section 5.3, the details of the cross-Paradigm transfer learning approach will be expressed, and the results will be compared for the two classifiers. Finally, we analyze the network and report the results (Section 5.4).

Experimental setup:

Here we use CNNs inspired from the structure provided by Cecotti [28]. Two relatively similar architectures are used for both the cross-subject and the cross-paradigm transfer learning scenarios. The first network architecture has a convolutional layer as a spatial filter, another convolutional layer is used as a temporal filter, and the two fully-connected layers are used to generate outputs and labels for each input sample. The number of electrode channels used is 32, and the sampling frequency is 512 Hz. Also, the period from stimulus onset till 1000ms after that is considered as an epoch. Therefore, each epoch has N = 1s × 512 = 512 temporal samples, which subsamples to 25 temporal samples. Therefore each input sample is given to the CNN in the form of a 32 × 25 matrix. The second network architecture is like the first one, with the exception that the second convolutional layer temporal filter is not included. The characteristics of the layers used in these networks are given in **Table 2**.

To train and implement LDA classifier, in the same way, applied to CNN, we first arrange the samples into a 32 × 25 matrix form, then we vectorize the feature matrix, which results in 800 features. Finally, by performing PCA and selecting the features that maintain 99% of the variance, the features are reduced.

The neural network and the LDA classifier used in this paper are trained and tested separately for the various modes described below. Character detection accuracy is derived from equation:

$$Acc_{char} = \frac{N_{true\_predict}}{N_{all}}$$

where $N_{true\_predict}$, the number of characters that are correctly detected and N_{all} represents the total number of characters. Also, the reported accuracies in Single, Dual, and Triple RSVP paradigms was calculated over ten, five, and three repetitions, respectively.

Table 2: CNN architecture specification.

Layer	Operation	Kernel Size	Feature maps/ neuron	Activation function
1	Convolution	(32,1)	10	ReLu
2	Convolution	(1,5)	15	ReLu
3	Fully-connected	-	20	ReLu
4	Fully-connected	-	1	Sigmoid

Cross-subject Transfer learning:

As stated previously, in this article, the dataset is related to three paradigms, and each one is available for three subjects. To implement the transfer learning approach in the cross-subject mode for each paradigm, the data of 2 other subjects (out of three existing subjects) with 20 percent of the third subject's data are considered as the train and after training, we use remaining amount of third subject's data as test(it means that data is divided to 5 parts). For each subject in each protocol, we repeat this operation five times, so that each part of the data is used once as train data. Finally, the average of 5 accuracies obtained from these 5 test parts is reported as the final character recognition accuracy for each subject.

It is worth noting that train and test phase of the LDA algorithm is similar to the one aforementioned for CNN. In this case, there is no specific operation to be referred to as transfer learning, and only a standard classification procedure is performed, with this difference, rather than classifying with only one subject's data the data from other subjects is also used. The main purpose of using the LDA classifier is to better comparison with the results of CNN. The results of the implementation of transfer learning approach using CNN with the fine-tuning technique are presented in **Table 3** the classification results using LDA are presented in **Table 4**.

Cross-paradigm Transfer learning:

We have a dataset of three different protocols, all of which are oddball patterns for triggering to excite the P300 based on RSVP Paradigm, with some minor differences. This idea came to our minds to examine the transfer learning algorithm on different protocols (cross paradigm). There are also three subjects from each data protocol here. For each subject, the data for the two protocols and 20% of the third protocol are

considered as the train data, and the remaining 80% are used as the test.

Table 3: Cross-subject transfer learning using CNN (fine-tuning).

Paradigm	Subject 1	Subject 2	Subject 3	Mean
Single	0.93	0.93	0.91	0.92
Dual	1	0.95	1	0.98
Triple	0.75	0.88	0.88	0.84

Table 4: Cross-subject transfer learning using LDA.

Paradigm	Subject 1	Subject 2	Subject 3	Mean
Single	0.84	0.88	0.84	0.85
Dual	0.88	0.77	0.93	0.86
Triple	0.71	0.68	0.84	0.74

However, in CNN, the model will first be trained with two protocol data, then we use the fine-tuning model, similar to that in Section 5.3. The initial layers are frozen, and the fully-connected layers are trainable and train the fully-connected layer weights of the network, with 20% of the third protocol data. The results of LDA and CNN have been reported in **Table 5 and Table 6,** respectively.

Table 5: Cross-paradigm transfer learning using CNN.

Test Paradigm	Subject 1	Subject 2	Subject 3	Mean
Single	0.92	0.95	0.86	0.91
Dual	1	0.97	1	0.99
Triple	0.75	0.86	0.88	0.83

Table 6: Cross- paradigmtransfer learning using LDA.

Test Paradigm	Subject 1	Subject 2	Subject 3	Mean
Single	0.97	0.88	0.88	0.91
Dual	0.93	0.93	0.95	0.93
Triple	0.68	0.73	0.84	0.75

In addition to the above results, another case has been investigated to use only one protocol data as a train, and for one target protocol same as before, the target subject data is divided to 20% for train and 80% for test. In this case, the results of the CNN classifier are shown in **Table 7**.

Network Analysis:

The use of CNN with the architecture provided by Cecotti [28] and similar works have a definite step for classification of the EEG signal, which is using a convolutional layer as a spatial filter. The network used in this paper is similar to the previous related work, on the first hidden layer has a convolutional layer with a C × 1 kernel size, where C = 32 and shows the number of electrode channels. This convolutional layer carries a spatial filter and performs some automatic channel selection. Once a kernel (32 × 1) is convolved with inputs in size of 32 × 25, each output sample resulted as a linear combination of different channels. This operation will occur in the learning process in such a way that electrodes that contain more information and have a higher discrimination ability will get higher weight and the other electrodes get a lower weight (near 0). Perhaps the most critical layer in such networks is the first layer (the spatial filter) which optimally combines channels. This is an important feature extraction process that can further simplify the classification with less error. Similarly, H.Shan [10] implemented a network in which the role of the first convolution layer was to subsample and transform the signal in the temporal dimension as well as spatial filtering.

Table 7: Cross- paradigm transfer learning using CNN only with one paradigm dataset for training.

Test paradigm	Train paradigm	Subject 1	Subject 2	Subject 3	Mean
Single	Dual	0.91	0.91	0.93	0.91
Single	Triple	0.67	0.81	0.75	0.74
Dual	Single	0.91	0.88	0.61	0.8
Dual	Triple	0.95	0.95	1	0.96
Triple	Single	0.53	0.64	0.4	0.52
Triple	Dual	0.73	0.86	0.88	0.82

In the following, topologies obtained from the feature maps of the first layer of the trained network on the dataset of the triple protocol of subject A are shown in **Figure 1**. In these topographies, the dark areas correspond to larger weights, and the light regions correspond to smaller ones. The darker regions represent the position of the electrodes that are more discriminable ability and probably have a stronger p300 component because here the significant distinction is related to the presence and absence of p300 in the samples. On the contrary, lighter areas exhibit less important electrodes positions.

In Figure, the Grand averages of ERPs corresponding to the Triple RSVP protocol for the subject A are plotted. The blue curve is the mean of samples with ERP response to non-target stimuli, and the red curve is the result of the samples obtained from the ERP response to the target stimuli. The brain topography of the Cz electrode has also been drawn at different times, and it is clear that about 400-600ms, P300 component is excited and the topography of the brain at this time in the middle of the head is the most active.

As we know, the P300 component occurs in the cognitive cortex, and we expect electrodes located in this area to be more active. This activity is seen in the brain topographies **(Figure 2)**.
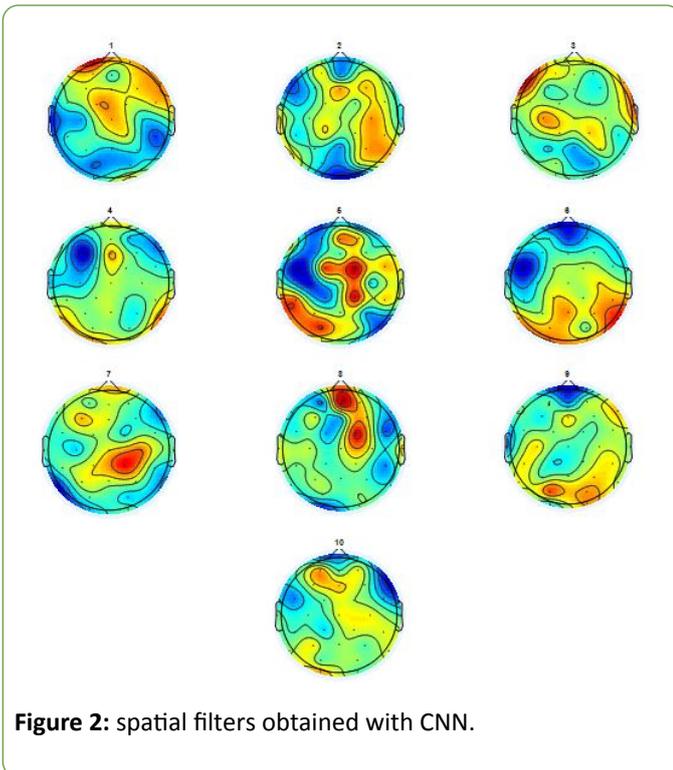
Figure 2: spatial filters obtained with CNN.

Also, the weights obtained from the first convolutional layer show the ability of the network to identify the position of the optimal electrodes, which in the topologies 5, 7 and 8 of the cognitive region electrodes weighs more **(Figure 3)**. Although all of the feature maps have not met our expectation in detecting optimal electrodes, the final classification results indicate that this layer has been very efficient, especially when only one convolutional layer is used and this single layer has been able to create distinctive features for the classifier's layer.
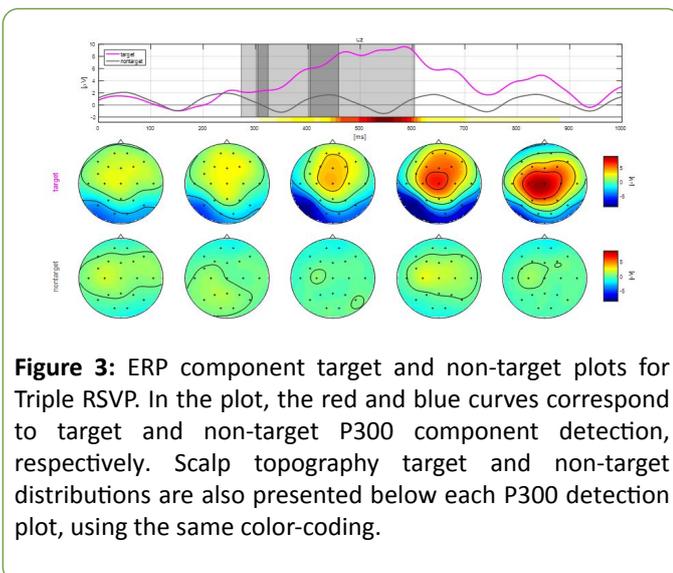

Figure 3: ERP component target and non-target plots for Triple RSVP. In the plot, the red and blue curves correspond to target and non-target P300 component detection, respectively. Scalp topography target and non-target distributions are also presented below each P300 detection plot, using the same color-coding.

Discussion

In this paper, CNN has been used to detect P300 on a dataset derived from three P300-based speller paradigms (Single, Dual, and Triple RSVP). The use of CNNswith fine-tuning technique in the TL approach has been done with two scenarios of cross-subject and cross-paradigm. The advantage of using the transfer learning algorithms is that the amount of data obtained from the target subject for training the network is significantly reduced. Also, in EEG analysis, since the classifiers are highly sensitive to the variations between subjects, the role, and importance of transfer learning is bolded with the presentation of generalizable within-subject algorithms.

For each subject in the Single, Dual, and Triple RSVP protocols, we have 11700, 6525, and 4725 samples, respectively. If we want to use the usual data of a subject related to a specific protocol for training and testing a network, we should devote roughly 80% of the data used as a train and 20% as a test. However, in order to implement the transfer learning approach, in this case, only 20% of the target subject data is used as training data, and the rest of 80% is used as test data. Therefore, the amount of training data required is reduced by about 75%. We used CNNs with fine-tuning technique. Therefore network training takes place in two stages. In the first step, the training data of other subjects are used, and because of the large amount of training data needed to train deep learning networks, the training process goes well. The above explanation illustrates the main reasons for transfer learning implementation using the fine-tuning technique.

In the scenario of cross-subject transfer learning, the mean accuracy of character recognition for CNN in three single, dual, and triple RSVP modes has increased 11.76%, 13.95%, and 13.51%, respectively in comparison to LDA classifier (compare the results of **Table 3 and Table 4**). The results show that in this scenario we have been able to minimize the training data needed for training the network, while the accuracy of character recognition in single subject mode reported in [44], did not decrease a lot.

Conclusion

In the cross-paradigm transfer learning scenario, which is the innovation of this paper, the goal is to use the dataset for various but related protocols corresponding to the target protocol for training network. Since the dataset of the three different protocols are available, two types of network training are performed using auxiliary data: 1) only data from one protocol for training the second protocol's network, 2) using two protocol's data for training the third protocols' network. In the case of two-protocol data, the average accuracy character recognition of CNN has increased by 6.45% and 10.66% respectively in comparison to LDA for single and dual RSVP (**Table 5 and Table 6**). Comparing the results of the networks that are trained with the data of a protocol against the networks that uses the data of both protocols, the results were interesting. The single protocol is well-trained with Dual protocol data, but the nature of data in the triple and single protocol shows a massive difference, which makes the average of the accuracy of the character's recognition decreased to 74%. Also, the Dual protocol is not well-trained through the Single protocol, while using triple-protocol data as training

lead to an average accuracy of 96% in character spelling. Eventually, considering the triple protocol as test data, Dual protocol has played the role of Train data well for it, but for the single protocol average character recognition accuracy decreases to about 52% and confirms the difference in the nature of the two protocols. According to results mentioned in **Table 5** and **Table 7**, it can be said that combination of two paradigms leads to a better result than using the only one paradigm data for character detection of the third paradigm.

References

1. Kostov A, Polak M (2000) Parallel man-machine training in development of EEG-based cursor control. IEEE Trans Rehabil Eng 8: 203-205.

2. Allison BZ, Wolpaw EW, Wolpaw JR (2007) Brain−computer interface systems: progress and prospects. Expert Rev Med Devices 4: 463-474.

3. Birbaumer N, Cohen LG (2007) Brain−computer interfaces: communication and restoration of movement in paralysis. JPhysiol 579: 621-636.

4. Birbaumer N, Ghanayim N, Hinterberger T, Iversen I, Kotchoubey B, et al. (1999) A spelling device for the paralysed. Nature 398: 297.

5. Gilroy SW, Porteous J, Charles F, Cavazza M, Soreq E, Raz G, et al. (2013) A Brain-Computer Interface to a Plan-Based Narrative. IJCAI, Citeseer.

6. Abootalebi V, Moradi MH, Khalilzadeh MA (2009) A new approach for EEG feature extraction in P300-based lie detection.Comput Meth Prog Bio 94: 48-57.

7. Chumerin N, Manyakov NV, Combaz A, Robben A, van Vliet M, et al. (2011) Steady state visual evoked potential based computer gaming−The Maze. Int Conf Intell Technol Interact Entertain, Springer.

8. Mijani AM, Shamsollahi MB, Hassani MS, Jalilpour S (2018) Comparison between Single, Dual and Triple Rapid Serial Visual Presentation Paradigms for P300 Speller. IEEE Int Conf Bioinformatics Biomed.

9. Orhan U, Hild KE, Erdogmus D, Roark B, Oken B, et al. (2012) RSVP keyboard: An EEG based typing interface. ICASSP, IEEE.

10. Shan H, Liu Y, Stefanov TP(2018) A Simple Convolutional Neural Network for Accurate P300 Detection and Character Spelling in Brain Computer Interface. IJCAI, pp. 1604-1610.

11. Wu Z, Lai Y, Xia Y, Wu D, Yao D (2008) Stimulator selection in SSVEP-based BCI. Med Eng Phys 30: 1079-1088.

12. Zhang Y, Zhao Q, Jin J, Wang X, Cichocki A (2012) A novel BCI based on ERP components sensitive to configural processing of human faces. J Neural Eng 9: 026018.

13. Pfurtscheller G, Neuper C (2001) Motor imagery and direct brain-computer communication. Proceed IEEE 89: 1123-1134.

14. Simbolon AI, Turnip A, Hutahaean J, Siagian Y, Irawati N (2015) An experiment of lie detection based EEG-p300 classified by SVM algorithm. ICACOMIT, IEEE.

15. Acqualagna L and Blankertz B (2013) Gaze-independent BCI-spelling using rapid serial visual presentation (RSVP). Clin Neurophysiol 124: 901-908.

16. Rastjoo A, Arabalibeik H (2009) Evaluation of Hidden Markov Model for p300 detection in EEG signal. Stud Health Technol Inform 142: 265-267.

17. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification withdeep convolutional neural networks. Adv Neural Inf Process Syst 2012: 1097-1105.

18. Zhang Y, Meng JE, Venkatesan R, Wang N, Pratama M (2016) Sentiment classification using comprehensive attention recurrent models. IJCNN, IEEE.

19. Hiraiwa A, Shimohara K, Tokunaga Y (1990) EEG topography recognition by neural networks.IEEE Eng Med Biol 9: 39-42.

20. Tjepkema-Cloostermans MC, de Carvalho RC, van Putten MJ (2018) Deep learning for detection of focal epileptiform discharges from scalp EEG recordings. ClinNeurophysiol 129: 2191-2196.

21. Hussein R, Palangi H, Ward RK, Wang ZJ (2019) Optimized deep neural network architecture for robust detection of epileptic seizures using EEG signals. Clin Neurophysiol 130: 25-37.

22. Van Leeuwen KG, Sun H, Tabaeizadeh M, Struck AF, Van Putten MJ, et al. (2019) Detecting abnormal electroencephalograms using deep convolutionalnetworks. ClinNeurophysiol 130: 77-84.

23. Abou Jaoude M, Jing J, Sun H, Jacobs CS, Pellerin KR,et al. (2020) Detection of mesial temporal lobe epileptiform discharges on intracranial electrodes using deep learning. ClinNeurophysiol 131: 133-141.

24. Maddula R, Stivers J, Mousavi M, Ravindran S, de Sa V(2017)Deep recurrent convolutional neural networks for classifying P300 BCI signals. GBCIC.

25. Wang P, Lu J, Zhang B, Tang Z(2015) A review on transfer learning for brain-computer interface classification. ICIST, IEEE.

26. Jayaram V, Alamgir M, Altun Y, Scholkopf B, Grosse-Wentrup M (2016) Transfer learning in brain-computer interfaces. IEEE Comput Intell M 11: 20-31.

27. Jayaram V and Grosse-Wentrup M (2016) A transfer learning approach for adaptive classification in P300 paradigms. Proceedings of the Sixth International BCI Meeting.

28. Cecotti H and Graser A (2011) Convolutional neural networks for P300 detection with application to brain-computer interfaces. IEEE Trans Pattern Anal Mach Intel l33: 433-445.

29. Manor R, Geva AB (2015) Convolutional neural network for multi-category rapid serial visual presentation BCIFront Comput Neurosci 9: 146.

30. Liu M, Wu W, Gu Z, Yu Z, Qi F, Li Y (2018) Deep learning based on Batch Normalization for P300 signal detection. Neurocomputing 275: 288-297.

31. Tal O, Friedman D (2019) Recurrent Neural Networks for P300-based BCI. arXiv preprint arXiv 2019: 1901.

32. Cremone-Caira A, Vaidyanathan A, Hyatt D, Gilbert R, Clarkson T, et al. (2019) Test-Retest Reliability of the N2 Event-Related Potential in School-Aged Children with Autism Spectrum Disorder (ASD). Clin Neurophysiol 131: 406-413.

33. Kajosch H, Gallhofer B, Corten P, From L, Verbanck P, et al. (2016) The bimodal P300 oddball component is decreased in patients with an adjustment disorder: An event-related potentials study. ClinNeurophysiol 127: 3209-3216.

34. Nan C, Wang G, Wang H, Wang X, Liu Z, et al. (2018) The P300 component decreases in a bimodal oddball task in individuals

with depression: An event-related potentials study. Clin Neurophysiol 129: 2525-2533.

35. Hünerli D, Emek-Savaş DD, Çavuşoğlu B, Çolakoğlu BD, Ada E, et al. (2019) Mild cognitive impairment in Parkinson's disease is associated with decreased P300 amplitude and reduced putamen volume. Clin Neurophysiolo1 30: 1208-1217.

36. Gonsalvez CJ, Polich J (2002) P300 amplitude is determined by target-to-target interval. Psychophysiol 39: 388-396.

37. Acqualagna L, Blankertz B (2011) A gaze independent spelling based on rapid serialvisual presentation. Conf Proc IEEE Eng Med Biol Soc.

38. Farwell LA, Donchin E (1988) Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. J Clin Neurophysiol 70: 510-523.

39. Acqualagna L, Treder MS, Schreuder M, Blankertz B (2010) A novel brain-computer interface based on the rapid serial visual presentation paradigm. 2010 Conf Proc IEEE Eng Med Bio.

40. Mohedano E, McGuinness K, Healy G, O'Connor NE, Smeaton AF, et al. (2015) Exploring EEG for object detection and retrieval. Proceedings of the 5th ACM on International Conference on Multimedia Retrieval, ACM.

41. Cecotti H (2016) Single-trial detection with magnetoencephalography during a dual-rapid serial visual presentation task. IEEE Trans Biomed Eng 63: 220-227.

42. Lin Z, Zeng Y, Gao H, Tong L, Zhang C, et al. (2017) Multirapid serial visual presentation framework for EEG-based target detection. BioMed Res Int 2017.

43. Lin Z, Zhang C, Zeng Y, Tong L, Yan B (2018) A novel P300 BCI speller based on the Triple RSVP paradigm. Sci Rep 8: 3350.

44. Mijani AM, Shamsollahi MB, Hassani MS (2019) A Novel Dual and Triple Shifted RSVP Paradigm for P300 Speller. J Neurosci Meth 328: 108420.