# Deep Transfer Learning for Subject-Independent ERP-based BCIs

Kassymzhomart Kunanbayev[*] Dias Azhigulov[*] Berdakh Abibullaev[†] Amin Zollanvari[*]

[*]Electrical and Computer Engineering Department, School of Engineering and Digital Sciences, Nazarbayev University, Kazakhstan
[†]Robotics and Mechatronics Department, School of Engineering and Digital Sciences, Nazarbayev University, Kazakhstan
Emails: {kassymzhomart.kunanbayev, dias.azhigulov, berdakh.abibullaev, amin.zollanvari}@nu.edu.kz

2021 9th International Winter Conference on Brain-Computer Interface (BCI) | 978-1-7281-8485-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/BCI51272.2021.9385311

***Abstract*—Designing subject-independent Brain-Computer Interfaces remains to be an open question for developing systems that can capture the inter-subject intrinsic brain features and classify them with reasonable accuracy. This paper presents the application of the state-of-the-art deep transfer learning architectures on classifying ERP signals. We report 66.87%, 67.64%, 65.58%, and 71.93% test classification accuracy for DenseNet121, DenseNet201, Xception, and VGG-16 models, respectively. The experimental results demonstrate the viability of our approach in subject independent ERP-signals classification and suggest the better performance of models with fewer layers in classifying ERP signals.**

***Index Terms*—Brain-Computer Interfaces, ERPs, P300, subject-independent, deep transfer learning**

## I. Introduction

BRAIN Computer Interfaces (BCI) is a promising field of the current neuroscience and engineering technology, which would allow people to communicate and control external devices by decoding neural activities [1], [2]. This is especially of great importance for people affected by amyotrophic lateral sclerosis (ALS), in which the neurons responsible for voluntary muscular movements are affected. Therefore, a potential technology for establishing the communication with the external world for such people is considered to be a BCI-based speller device.

One of the first and still most prominent BCI spellers is a P300-based speller, a 6 by 6 matrix consisting of alphanumeric characters [3]. It utilizes P300 Event-Related Potentials (ERP), one of the most studied component of electroencephalography (EEG) signals [4]. Although measuring the P300 signals may seem trivial, there are specific research challenges that affect the accuracy of the speller. Detecting ERPs is challenging due to their low signal to noise ratio, large variability, and high dimensionality of these observations [5]. Due to the psychological and physiological distinctions among subjects, classification of ERP-based P300 signals becomes even harder, thus urging for the design of robust subject-independent BCIs that are robust to data variability in EEG signals. Therefore, this work is devoted to utilizing various signal processing techniques and deep learning models to construct a reliable decoding pipeline.

Deep transfer learning – a novel line of work in deep learning–involves fine-tuning pre-trained deep architectures developed for specific applications. CNN-based deep neural networks are pre-trained on large-scale datasets such as ImageNet in order for models to capture generic features that could be useful in other problems. In this regard, the pre-trained model are transferred to other applications and, subsequently, retrained to make the extracted features more relevant to the application at hand. This approach is usually beneficial in reducing time for training, higher accuracy, and computational efficiency. The commonly known architectures include models such as VGGNet, DenseNet, and Xception. In some BCI applications, these models were already deployed: in [6], authors suggest VGG-16 based model that improves the classification accuracy and efficiency on a motor imagery BCI, while authors in [7] report a superior performance of deep transfer learning models on music imagery dataset. The primary focus in the present study is to apply deep transfer learning on ERP-based data to achieve subject-independent predictive models. The architectures under consideration are Xception, VGG-16, and DenseNet based family of architectures.

This work was partially supported by the Faculty Development Competitive Research Grants Program of Nazarbayev University under grant numbers 021220FD2051, 021220FD1151, and SOE2018008.

978-1-7281-8485-2/21/$31.00 ©2021 IEEE

## II. Material and Methods

### A. Dataset

We used the P300-based dataset, which was collected on an experiment using a Farwell and Donchin styled speller by P.Arico *et al.* [8], which is accessible via MOABB toolkit [9]. The dataset is collected on 16 electrodes, sampled at 256 Hz, and high- and low-pass filtered at 0.1 and 20 Hz cutoff frequencies. The following data processing methods were applied to enhance the accuracy of the final model:

*1) Data upsampling.* To avoid the effects of dataset class imbalance in the model selection process, we apply upsampling the observations of the minority class to be the same size as the majority class.

*2) Normalization.* We apply min-max normalization by first determining the mean and standard deviation of the training dataset, and then applying these parameters to normalize each set individually (training, validation, and test).

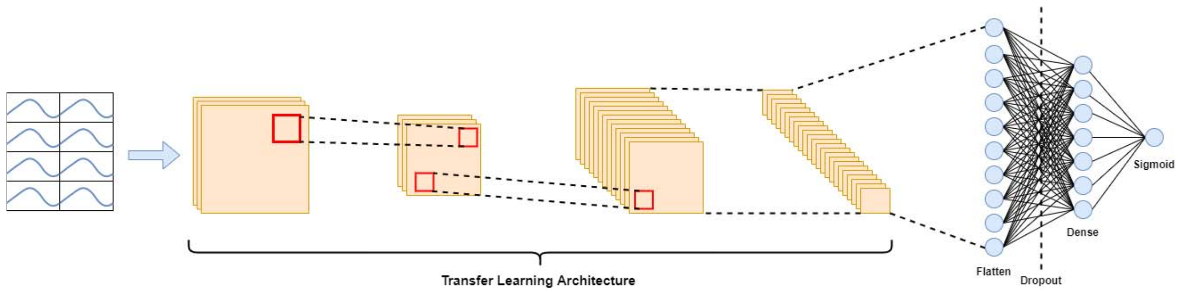


Fig. 1. The schematics of proposed framework. Identical fully connected layers are added on top of each architecture that is being transferred (i.e. DenseNet121, DenseNet201, Xception, VGG-16)

*3) Signal replication.* The architectures used here for the deep transfer learning expect larger dimensions of input data (e.g. $3\times224\times224$). Therefore, we replicate the initial ERP signal dimension of $1\times16\times78$ repeatedly in all three dimensions until the desired shape is obtained.

### B. Model selection

Let $\mathcal{S}_i$ denote measurements collected for subject $i$, $i = I, II.., X$; therefore, the full dataset, which is denoted by $\mathcal{S}$, is obtained as $\mathcal{S} = \bigcup_{i=\mathrm{I}}^{\mathrm{X}} \mathcal{S}_i$. First, we use the data collected for the last subject $\mathcal{S}_X$ for fine-tuning our models (generally referred to as validation set). Then, among the remaining nine subjects, we perform cross-validation holding one subject $\mathcal{S}_i$ out at a time for evaluating the model's performance (generally referred to as test set). The training of deep transfer learning models is performed by pooling the data for remaining subjects; that is to say, $\mathcal{S} - \mathcal{S}_i - \mathcal{S}_X$.

Let $\mathbf{x}_{l,i}$ denotes a single observation, therefore $\mathcal{S}_i = \bigcup_{l=1}^{n_i}$ is the observation data set within one single subject. Thus, both validation and test accuracies can be found assuming the general definition of accuracy for any subject $\mathcal{S}_j$:

$$\mathrm{acc}_{\mathcal{S}_j} = \frac{1}{|\mathcal{S}_j|} \sum_{\mathbf{x}_{l,i}} I_{\{y_{e,i}=\hat{y}_{e,i}\}} \tag{1}$$

where $|\mathcal{S}_j|$ is the cardinality of $\mathcal{S}_j$, $I$ is 1 if the true label $y_{e,i}$ and the predicted label $\hat{y}_{e,i}$ match, otherwise 0.

TABLE I
TEST ACCURACIES OF DEEP TRANSFER LEARNING MODELS (%)

| Test Subject | DenseNet121 | DenseNet201 | Xception | VGG-16 |
|---|---|---|---|---|
| $\mathcal{S}_I$ | 66.07 | 54.55 | 57.48 | 67.42 |
| $\mathcal{S}_{II}$ | 70.08 | 69.07 | 62.03 | 54.55 |
| $\mathcal{S}_{III}$ | 74.37 | 74.68 | 72.47 | 75.54 |
| $\mathcal{S}_{IV}$ | 73.90 | 75.13 | 70.08 | 79.10 |
| $\mathcal{S}_V$ | 71.18 | 78.79 | 76.55 | 81.63 |
| $\mathcal{S}_{VI}$ | 54.55 | 69.85 | 62.31 | 76.93 |
| $\mathcal{S}_{VII}$ | 59.79 | 60.67 | 57.32 | 75.98 |
| $\mathcal{S}_{VIII}$ | 54.55 | 45.45 | 64.08 | 54.55 |
| $\mathcal{S}_{IX}$ | 77.34 | 80.59 | 67.87 | 81.66 |
| **Average** | 66.87 | 67.64 | 65.58 | 71.93 |

TABLE II
AUC SCORES OF DEEP TRANSFER LEARNING MODELS (%)

| Test Subject | DenseNet121 | DenseNet201 | Xception | VGG-16 |
|---|---|---|---|---|
| $\mathcal{S}_I$ | 66.32 | 50.00 | 60.45 | 68.58 |
| $\mathcal{S}_{II}$ | 71.26 | 70.67 | 64.69 | 50.00 |
| $\mathcal{S}_{III}$ | 73.21 | 73.98 | 72.93 | 74.66 |
| $\mathcal{S}_{IV}$ | 74.13 | 74.54 | 71.82 | 78.51 |
| $\mathcal{S}_V$ | 72.23 | 79.26 | 77.31 | 80.80 |
| $\mathcal{S}_{VI}$ | 50.00 | 70.58 | 63.91 | 77.20 |
| $\mathcal{S}_{VII}$ | 62.50 | 63.32 | 60.32 | 76.00 |
| $\mathcal{S}_{VIII}$ | 50.00 | 50.00 | 65.14 | 50.00 |
| $\mathcal{S}_{IX}$ | 77.37 | 80.82 | 70.02 | 81.33 |
| **Average** | 66.34 | 68.13 | 67.40 | 70.79 |

## III. RESULTS

Each model was trained for 50 epochs. Table I summarizes the test accuracies obtained from the four models for each subject treated as a test in the process of leave-one-subject-out. These results are further substantiated by Table II, where one can see AUC scores by each model. The results show that the deeper networks (DenseNets and Xception) noticeably exhibit a lower classification accuracy. Interestingly, it was determined that VGG-16, although much lighter than other networks, is the best performing model with an average test accuracy and AUC of 71.93% and 70.79%, respectively.

## IV. CONCLUSION

This study investigated an approach for subject-independent BCIs, where signal processing techniques are employed along with deep transfer learning to reach over 70% accuracy for classification of ERP signals. The results suggest that very deep neural networks such as DenseNets (121 and 201 layers) and Xception (71 layers) are inferior compared to shallower networks such as VGGNets (16 layers) for classification of ERP-based P300 signals.

## REFERENCES

[1] 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, March 2011.

[2] D. Nurseitov, A. Serekov, A. Shintemirov, and B. Abibullaev, "Design and evaluation of a p300-erp based bci system for real-time control of a mobile robot," in *2017 5th International Winter Conference on Brain-Computer Interface (BCI)*. IEEE, 2017, pp. 115–120.

[3] L. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials," *Electroen-*

*cephalography and Clinical Neurophysiology*, vol. 70, no. 6, pp. 510–523, Dec. 1988.

[4] R. Ramele, A. Villar, and J. Santos, "EEG waveform analysis of p300 ERP with applications to brain computer interfaces," *Brain Sciences*, vol. 8, no. 11, p. 199, Nov. 2018.

[5] B. Abibullaev and A. Zollanvari, "Learning discriminative spatiospectral features of ERPs for accurate brain–computer interfaces," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 5, pp. 2009–2020, Sep. 2019.

[6] G. Xu, X. Shen, S. Chen, Y. Zong, C. Zhang, H. Yue, M. Liu, F. Chen, and W. Che, "A deep transfer convolutional neural network framework for EEG signal classification," *IEEE Access*, vol. 7, pp. 112 767–112 776, 2019.

[7] C. Tan, F. Sun, and W. Zhang, "Deep transfer learning for EEG-based brain computer interface," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, Apr. 2018.

[8] P. Aricò, F. Aloise, F. Schettini, S. Salinari, D. Mattia, and F. Cincotti, "Influence of P300 latency jitter on event related potential-based brain–computer interface performance," *Journal of Neural Engineering*, vol. 11, no. 3, p. 035008, Jun. 2014.

[9] V. Jayaram and A. Barachant, "MOABB: trustworthy algorithm benchmarking for BCIs," *Journal of Neural Engineering*, vol. 15, no. 6, p. 066011, Sep. 2018.