Proceedings of the
9th Graz Brain-Computer Interface Conference 2024 10.3217/978-3-99161-014-4-085

NEUROPHONE: REAL-TIME BRAIN-MOBILE PHONE INTERFACE

Norhan Abdelhafez[1], Manal Tantawy[1], Abdelrahman Sayed[1], Nora Ekramy[1], Mohammed Nour-Eldin[1]

[1] Scientific Computing Department, Faculty of Computer and Information Science, Ain Shams University, Cairo, Egypt

E-mail: norhan.abdelhafez3@gmail.com

ABSTRACT: The extensively studied P300 component of the human event-related potential in cognitive neuroscience has significant applications, including constructing BCI systems for individuals with motor disabilities. However, accurately and efficiently identifying the P300 component in EEG data poses challenges due to the low signal-to-noise ratio and biological diversity among subjects. To address this, cutting-edge deep learning architectures were developed and employed. Initially, digital signal processing techniques were applied, followed by training and evaluation of DL models like Chrononet, EEGNet, DCRNN, CNNs, and RNNs. Results revealed that our lightweight CNN model, combined with K-fold cross-validation and weighted class, achieved the highest average classification accuracy of 98% surpassing other models for subject-dependent P300 classification. This high-performing CNN model facilitated the creation of NeuroPhone, a communication application grounded in the core principles of BCI systems.

INTRODUCTION

Electroencephalography (EEG) has opened a window into the human brain, allowing us to visualize its electrical activity and delve into the hidden language of its neurons. From its widespread medical applications to its growing presence in research and consumer domains, EEG offers a powerful tool for understanding and interacting with the mind. At the heart of this interaction lies the fascinating world of brainwaves, different patterns reflecting different states of consciousness. From the high-frequency beta waves associated with focused attention to the slow delta waves accompanying deep sleep, each frequency serves as a neural biomarker for specific cognitive states [1].

Among these brainwaves, the P300 event-related potential (ERP) holds a special position. This distinctive positive spike, peaking roughly 300 milliseconds after a specific stimulus, reveals much about our cognitive processes. Researchers have extensively studied the P300, recognizing its crucial role in attention, memory, decision-making, and information processing [2]. Its potential, however, extends beyond research labs, paving the way for revolutionary technology called Brain-Computer Interfaces (BCIs).

BCIs offer a direct communication channel between the brain and external devices, bypassing traditional input methods. By harnessing the power of P300 and other EEG signals, BCIs empower individuals to control computers, prosthetic limbs, and even communicate through their thoughts [3]. Yet, despite the immense promise of BCIs, their path to widespread adoption is met with two key challenges: achieving robust and accurate P300 detection and overcoming the computational limitations of existing BCI systems. Current models often struggle to extract the subtle P300 signal from the inherent noise of EEG data, and their demanding computational requirements prevent seamless integration with mobile devices, a crucial step towards accessibility for a wider population.

In this study, we present the development of NeuroPhone, an efficient BCI in the form of a communication application designed to break down these barriers. NeuroPhone leverages the P300 peak, enabling individuals with motor disabilities to control their smartphones and engage in digital communication solely through their visual attention. By employing cutting-edge deep learning techniques tailored for mobile device processing power, NeuroPhone aims to overcome the previous limitations of accuracy and accessibility. This paper delves into the development of NeuroPhone including the implemented digital signal processing (DSP) techniques, deep learning architectures used, and specifics of the development process of NeuroPhone's application software.

METHODS

We aimed to develop a comprehensive and computationally efficient classification model based on the detection of P300 event-related potential (ERP). We first went on exploring and evaluating different architectures, such as ChronoNet, EEGNet, DCRNN, and others, to determine their effectiveness and performance in a subject-dependent task. Then, we introduced a lightweight convolutional neural network (CNN) architecture that excels in capturing unique ERP features, leading to superior classification accuracy compared to existing state-of-the-art architectures.

In the subsequent sections, we first present details about the datasets we utilized in our work, including the online EPFL BCI Group dataset [4] and the data we collected offline using Emotiv EPOC headset. After that, we

Published by
Verlag der Technischen Universität Graz

 10.3217/978-3-99161-014-4-085

provide a comprehensive overview of the models we investigated. Then we present the architecture of our CNN model. Additionally, we outline the training methodology we employed, highlighting the steps and techniques utilized to optimize and fine-tune the models for optimal performance. And finally, we explain the details of NeuroPhone's application software, and the technology used.

Datasets: The EPFL BCI group dataset, which was employed in our research, played a crucial role in evaluating the performance of various models. This dataset was specifically curated by the Brain-Computer Interface (BCI) group at École Polytechnique Fédérale de Lausanne (EPFL) and is widely recognized in the field. The dataset consists of meticulously recorded electroencephalogram (EEG) signals, making it a valuable resource for investigating brain-computer interfaces. The dataset has a population of five disabled and four able-bodied subjects. Subjects were facing a laptop screen on which six images were displayed. The images were selected according to an application scenario in which users can control electrical appliances via a BCI system. The EEG was recorded at 2048 Hz sampling rate from 32 electrodes placed at the standard positions of the 10–20 international system. Each subject recorded 4 sessions and each session had 6 runs. For a single run, the images were flashed in random sequences, one image at a time. Each flash of an image lasted for 100 ms and during the following 300 ms none of the images was flashed, i.e. the interstimulus interval was 400 ms, see (Fig. 1).

For our collected dataset, we followed the same recording paradigm as EFPL dataset. We used the famous Emotiv EPOC headset with 14 channels placed at the standard positions of the 10–20. We recorded the EEG signal from a single male subject. The subject was faced by NeuroPhone's application screen which displayed 6 images (icons), each represented a certain functionality that allows the user to communicate with others, see (Fig. 2). More details are provided at the application subsection. The subject recorded 6 sessions; each session had a duration of 90 seconds with a sampling rate of 256 Hz.

Preprocessing: The data underwent several preprocessing steps to ensure optimal analysis. The re-referencing step involved utilizing the average signal from the two mastoid electrodes for re-referencing purposes. To obtain a desired signal range of 1 to 12 Hz, a band-pass Butterworth filter of order 3 was applied to filter the signal [5]. Subsequently, the signal was down sampled by 64 Hz to reduce computational load. Then, data was segmented such that each segment was corresponding to an event. A duration of 1 second was taken after each stimulus event and given that the duration of the flashing event was 400 ms, there was a 600 ms overlap. On a single segment, z-score normalization was implemented to normalize the signal. To handle extreme values, the signal underwent a Winsorizing process, where the 10th and 90th percentiles were calculated for samples from each electrode. Any amplitude values falling below the 10th percentile or above the 90th percentile was substituted with the respective 10th or 90th percentile value [4]. These preprocessing steps collectively aimed to optimize the data for deep learning models' training. The input signal shape for the models was (32 × 32) where the first dimension is the number of time samples, and the second dimension is the number of channels. And in our collected data input shape was (32 × 14) because EMOTIV dataset contained only 14 channels.

Deep Learning architectures: In our quest for the optimal deep learning architecture for subject-dependent P300 classification, we explored a diverse range of models, each offering its own set of advantages and limitations. Chrononet and EEGNet, specifically designed for EEG analysis, leverage convolutional layers to efficiently capture the temporal characteristics of the P300 component, making them well-suited for this task [6][7]. However, their deep architectures can be computationally expensive to train and might require a substantial amount of data for optimal performance. DCRNNs, combining the strengths of CNNs and RNNs, excel at capturing both the spatial and temporal information crucial for P300 detection [8]. Despite their effectiveness, DCRNNs can be more complex to design and train effectively, requiring careful hyperparameter tuning to unlock their full potential. Finally, standard RNNs, while adept at learning sequential data like EEG signals, can suffer from vanishing gradients, hindering their ability to learn from long sequences.

Our CNN model architecture: The proposed CNN architecture is designed to extract salient features from 2D EEG signals for robust P300 component classification. The model comprises two convolutional layers with 32 and 64 filters (3x3 kernel size), applying learned filters to extract spatial patterns relevant to P300 detection. ReLU activation introduces non-linearity. A max-pooling layer (2x2 pool size) down-samples feature maps, reducing dimensionality and promoting spatial invariance. A flatten layer prepares the extracted features for classification by two fully connected layers (128 neurons with ReLU activation, and 1 neuron with sigmoid activation), see (Fig. 3).

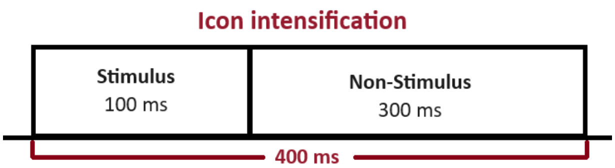

Figure 1: The interstimulus interval of each flashing event.

Training Methodology: To ensure robust model performance and mitigate the effects of overfitting, a K-fold cross-validation strategy was employed during training. Specifically, a 5-fold cross-validation approach was implemented (K=5). This technique divides the dataset into five partitions (folds) while preserving the proportion of P300 and non-P300 examples in each fold. Iteratively, one fold is designated as the testing set while

Published by
Verlag der Technischen Universität Graz

 10.3217/978-3-99161-014-4-085

the remaining folds are used for training. Model evaluation metrics are computed on the held-out testing set after each training iteration.

Figure 2: NeuroPhone's Mobile Application interface which contains 6 flashing images (icons).

To address potential class imbalance within the EEG dataset, where the number of P300 events (positive class) is significantly lower compared to non-P300 events (negative class), class weights were computed and incorporated into the training process. This approach assigns higher weights to the minority class (P300 events) during training. The specific weights are calculated based on the class frequencies within the training data. By assigning higher weights, the model is effectively forced to pay closer attention to the less frequent P300 examples, leading to a more balanced learning process and improved classification performance for the minority class.

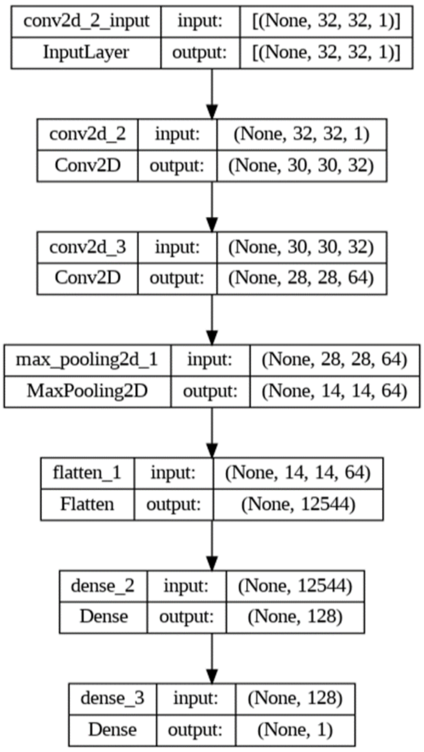

Figure 3: Architectural configuration of our CNN model.

The deep learning models were compiled with the Adam optimizer, a common choice for its adaptive learning rate capabilities, and binary cross-entropy loss, suitable for binary classification. Accuracy served as the primary evaluation metric.
Finally, average loss and accuracy scores across all folds were calculated to provide a comprehensive assessment of model performance under the K-fold cross-validation procedure.

Application: The purpose of our application is to help disabled people use their mobile phones and perform some important functions through it using only their visual attention. Our application is designed in a way that visually stimulates the user to choose the icon they desire. Each icon represents a functionality that enables them to control their smartphone. They are: Gallery, Contacts, Police, Favorite Contacts, Ambulance, and Fireman. The whole set of icons would flash in random order. The user would focus their attention on any icon they want to choose, and after some repetitions of flashing the whole set of icons, the DL model would detect the P300 peak that synchronized with the timing of the desired icon's flash, and thus, would fire the start of the execution of that icon's functionality.
We used Emotiv EPOC X 14 which consists of 14 EEG channels and 2 reference channels. The electrodes are located at AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4 according to the International 10-20 system, see (Fig. 4). The headset was connected to its software EMOTIV-PRO on the laptop then we start streaming data from Lab Streaming Layer (LSL) option in the application. The LSL feature allows efficient, two-way communication between EmotivPRO and other third-party software and devices.

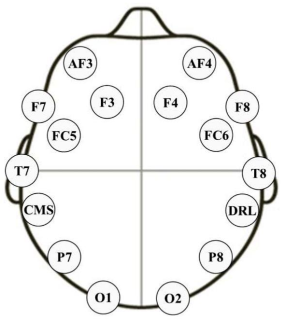

Figure 4: The location of EMOTIV electrodes.

We have created an API with Flask Library in Python language. The main functionality of this API is to synchronize between the EEG signal coming from EMOTIV and the flash timing coming from Neurophone's mobile application.
The flask API receives raw EEG signal samples from EMOTIV in addition to timestamp of each sample, so the first step it performs is to segment the raw EEG signal and preprocess it. Each segment is 1000 ms long which corresponds to 256 signal samples because EMOTIV's sampling rate is 256 HZ. After preprocessing, the segment's length would be 32 samples because of the

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-085

down sampling step in preprocessing. We also perform bandpass Butterworth filter between 1 and 12 HZ and Z-score normalization. After preprocessing, each segment is passed to the DL model to determine whether it contains P300 or not. The DL model we utilized in the API was our CNN model that we presented its architecture earlier in (Fig. 3). We chose CNN because it outperformed the others in the offline evaluation. Along with the received EEG signal samples and timestamp of each sample, Flask API receives each icon's flash timing from Neurophone's flutter mobile application.

We used Flutter to create the mobile application. The application was running on a Galaxy M31 Phone with an Octa-core Exynos 9611 (10nm) Processor and Android 12 operating system. The Application would continuously send each icon's flash time and icon's index to Flask API. The second step performed by the API is synchronization. If the EEG segment that was synchronized with an icon flash time contained a P300 peak, the API would send a firing response back to the flutter application to start the execution of that icon's functionality, see (Fig. 5).

RESULTS

In this section, we present the experimental results of our study focusing on the classification of P300 in EEG signals in subject-dependent task. To achieve this objective, we trained a variety of models, including Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Dynamic Convolutional Recurrent Neural Networks (DCRNN), EEGNet, and ChronoNet. Those models are evaluated on the EPFL BCI group dataset and our collected dataset. In Tab. 1, we demonstrate the results on the EPFL dataset. We provide the average k-fold accuracy and F1-score (across the 5 folds).

Among accuracy results, CNN achieved the highest average classification accuracy on EPFL data. We also explored their performance on our collected dataset. In Tab. 2, we demonstrate the results of our single male subject data. The demonstrated results on our collected are offline results, meaning that the collection and evaluation were performed offline and then the best performing model in the offline evaluation was utilized in the real-time scenario (it was the CNN model in this case).

Table 2: Results on our collected data

Model	Accuracy	F1-score
EEGNET	0.94	0.86
DRCNN	0.92	0.81
RNN	0.97	0.89
CNN	**0.98**	**0.95**

The results on our in-house data also demonstrated that CNN model outperforms the others by achieving a 98% average classification accuracy, highlighting the eligibility of CNN to be utilized in real-time.

DISCUSSION

We could notice from results in Tab. 1 that subject 8 achieved the highest average classification accuracy across most of the DL models. It may be attributed to the fact that the subject was highly motivated during the experiments as stated by the authors who collected the data [4]. Even though EPFL contained 9 subjects, they excluded the data of the fifth subject due to the difficulty of communication with him. We could also notice that CNN achieved the highest average classification accuracy and F1 score across all the subjects compared to the other models.

We acknowledge that the in-house data were small, and a larger dataset is required for the results to be generalizable. The use of parameters such as information transfer rate (ITR) is an essential metric to be utilized for an improved evaluation of the BCI system.

To compare our results with other studies implementing P300-based BCI systems, we find that Eric Sellers and Emanuel Donchin [9] achieved an average classification accuracy of 72% for ALS patients and 85% for abled subjects. We could see that Hubert Cecotti and Axel Gräser [10] achieved a classification accuracy of 95.5% using a CNN model. To compare our results to other studies that utilized the same dataset we used (EPFL dataset), we find that the authors in [11] achieved an average classification accuracy of 95.68% for the healthy subjects and 94.69% for disabled patients through their CNN model that uses 2-D EEG scalogram images. We also see Shojaedini et al. [12] reached a classification accuracy of 95.34% using a CNN model with a new adaptation method for hyperparameters. Our methods achieved a higher classification accuracy for P300 detection among the studies that used the same dataset.

While NeuroPhone provides a robust BCI system to detect P300 and control the mobile application, there are some limitations of the system. The first one is the inevitable time delay between the mobile application and the API. This delay is attributed to the quality of the connection between the phone and the API. We haven't accurately measured the delay time, but it was believed to be around a few seconds. The second limitation is the number of repetitions, in real-time, the user requires around 3 to 4 repetitions to select the desired icon. A single repetition is the flashing of all the icons, and it lasts for 2.4 seconds, so 3 repetitions would be around 8 seconds.

CONCLUSION

This research demonstrates the power of deep learning for EEG analysis and brain-computer interface development. Our lightweight CNN model, combined with K-fold cross-validation and class weighting, achieved superior P300 classification accuracy compared to other architectures. This enabled the successful creation of the NeuroPhone application. Future research will explore transfer learning for improved model generalization across subjects and investigate hybrid

Published by
Verlag der Technischen Universität Graz

 10.3217/978-3-99161-014-4-085

Table 1: Results on EPFL BCI group dataset

	Sub 1		Sub 2		Sub 3		Sub 4		Sub 6		Sub 7		Sub 8		Sub 9	
Model	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
EEGNET	0.89	0.70	0.84	0.64	0.93	0.80	0.90	0.74	0.89	0.71	0.90	0.77	**0.95**	0.87	0.86	0.68
ChronoNet	0.93	0.83	0.90	0.75	**0.94**	0.84	0.89	0.79	0.93	0.83	0.93	0.85	0.92	0.82	0.91	0.79
DRCNN	0.86	0.65	0.84	0.61	0.89	0.73	0.90	0.72	0.91	0.75	0.90	0.75	**0.93**	0.81	0.86	0.64
RNN	0.94	0.83	0.94	0.83	0.96	0.88	0.94	0.83	0.95	0.86	0.96	0.88	**0.97**	0.92	0.95	0.84
CNN	0.99	0.96	0.98	0.96	0.99	0.98	0.98	0.96	0.98	0.96	0.99	0.97	**0.99**	0.97	0.99	0.97

deep learning and signal processing approaches for further enhancements.

ACKNOWLEDGMENT

The authors acknowledge the support of the Information Technology Industry Development Agency (ITIDA) for their funding of this research.

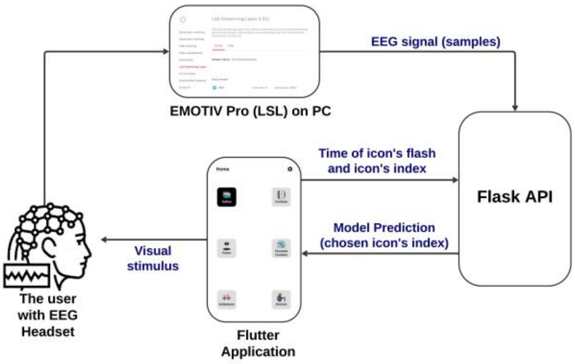

Figure 5: Using NeuroPhones' Flutter application interface, the user gets visually stimulated by the flashing, and at the same time, the EEG signal is transmitted to Flask API, the API synchronizes between the EEG signal and the flash timing and sends back a response to Flutter App to execute the desired icon's functionality.

REFERENCES

[1] Kumar JS, Bhuvaneswari P. Analysis of Electroencephalography (EEG) Signals and Its Categorization–A Study. Procedia Eng. 2012;38:2525-2536.

[2] Zhong R, Li M, Chen Q, Li J, Li G, Lin W. The P300 Event-Related Potential Component and Cognitive Impairment in Epilepsy: A Systematic Review and Meta-analysis. Front Neurol. 2019;10:943.

[3] Rashid M, Sulaiman N, P P Abdul Majeed A, Musa RM, Ab Nasir AF, Bari BS, Khatun S. Current Status, Challenges, and Possible Solutions of EEG-Based Brain-Computer Interface: A Comprehensive Review. Front Neurorobot. 2020;14:25.

[4] Hoffmann U, Vesin JM, Ebrahimi T, Diserens K. An efficient P300-based brain-computer interface for disabled subjects. J Neurosci Methods. 2008;167(1):115-25.

[5] Shojaedini SV, Morabbi S, Keyvanpour M. A New Method for Detecting P300 Signals by Using Deep Learning: Hyperparameter Tuning in High-Dimensional Space by Minimizing Nonconvex Error Function. J Med Signals Sens. 2018;8(4):205-214.

[6] Lawhern VJ, Solon AJ, Waytowich NR, Gordon SM, Hung CP, & Lance BJ. EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. J of Neural Eng. 2018;15(5):056013.

[7] Roy S, Kiral-Kornek I, Harrer S. ChronoNet: A Deep Recurrent Neural Network for Abnormal EEG Identification. ArXiv abs/1802.00308 (2018): 4.

[8] Li Q, Liu Y, Shang Y, Zhang Q, Yan F. Deep Sparse Autoencoder and Recursive Neural Network for EEG Emotion Recognition. Entropy (Basel). 2022;24(9):1187.

[9] Sellers EW, Donchin E. A P300-based brain-computer interface: initial tests by ALS patients. Clin Neurophysiol. 2006;117(3):538-48.

[10] Cecotti H, Gräser A. Convolutional neural networks for P300 detection with application to brain-computer interfaces. IEEE Trans Pattern Anal Mach Intell. 2011;33(3):433-45.

[11] Singh SA, Meitei TG, Devi ND, Majumder S. A deep neural network approach for P300 detection-based BCI using single-channel EEG scalogram images. Phys Eng Sci Med. 2021;44(4):1221-1230.

[12] Shojaedini SV, Morabbi S, Keyvanpour M. A New Method for Detecting P300 Signals by Using Deep Learning: Hyperparameter Tuning in High-Dimensional Space by Minimizing Nonconvex Error Function. J Med Signals Sens. 2018;8(4):205-214.

Published by
Verlag der Technischen Universität Graz