

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/310953136>

EEGNet: A Compact Convolutional Network for EEG-based Brain-Computer Interfaces

Article · November 2016

CITATIONS

12

READS

405

6 authors, including:



[Vernon Lawhern](#)

Army Research Laboratory

32 PUBLICATIONS 285 CITATIONS

[SEE PROFILE](#)



[Nicholas Waytowich](#)

Columbia University

10 PUBLICATIONS 57 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Machine learning for brain-computer interface [View project](#)

EEGNet: A Compact Convolutional Network for EEG-based Brain-Computer Interfaces

Vernon J. Lawhern^{1,2}, Amelia J. Solon^{1,3}, Nicholas R. Waytowich^{1,4}, Stephen M. Gordon^{1,3},
Chou P. Hung^{1,5}, and Brent J. Lance¹

¹Human Research and Engineering Directorate, U.S. Army Research Laboratory, Aberdeen Proving Ground, MD

²Department of Computer Science, University of Texas at San Antonio, San Antonio, TX

³DCS Corporation, Alexandria, VA

⁴Department of Biomedical Engineering, Columbia University, New York, NY

⁵Department of Neuroscience, Georgetown University, Washington, DC

November 28, 2016

Abstract

Objective: Brain-Computer Interface technologies (BCI) enable the direct communication between humans and computers by analyzing brain measurements, such as electroencephalography (EEG). These technologies have been applied to a variety of domains, including neuroprosthetic control and the monitoring of epileptic seizures. Existing BCI systems primarily use *a priori* knowledge of EEG features of interest to build machine learning models. Recently, convolutional networks have been used for automatic feature extraction of large image databases, where they have obtained state-of-the-art results. In this work we introduce EEGNet, a compact fully convolutional network for EEG-based BCIs developed using Deep Learning approaches. *Methods:* EEGNet is a 4-layer convolutional network that uses filter factorization for learning a compact representation of EEG time series. EEGNet is one of the smallest convolutional networks to date, having less than 2200 parameters for a binary classification. *Results:* We show state-of-the-art classification performance across four different BCI paradigms: P300 event-related potential, error-related negativity, movement-related cortical potential, and sensory motor rhythm, with as few as 500 EEG trials. We also show that adding more trials reduces the error variance of prediction rather than improving classification performance. *Conclusion:* We provide preliminary evidence suggesting that our model can be used with small EEG databases while improving upon the state-of-the-art performance across several tasks and across subjects. *Significance:* The EEGNet neural network architecture provides state-of-the-art performance across several tasks and across subjects, challenging the notion that large datasets are required to obtain optimal performance.

1 Introduction

A Brain-Computer Interface (BCI) is a mechanism for communicating with a machine via brain signals, bypassing normal neuromuscular outputs by using neural activity [1]. Traditionally, BCIs

leverage implanted electrodes for medical applications, such as neural control of prosthetic artificial limbs [2]. However, recent research has opened up the possibility for novel BCIs focused on enhancing performance of healthy users, often focused on noninvasive approaches based on electroencephalography (EEG) [3–6]. In both the medical and non-medical domains, there has been a proliferation of BCI paradigms. These paradigms range from directly controlling devices such as mouse cursors [7, 8], to passive brain monitoring for estimating states such as alertness or fatigue [9–11] or the occurrence of epileptic seizures [12–14], to paradigms for spelling [15] or image analysis [16].

Despite this wide range of paradigms, the general approach to designing BCI technologies remain the same. A BCI consists of five main processing stages [17]: a data collection stage, where neural data is recorded; a signal processing stage, where the recorded data is preprocessed and cleaned; a feature extraction stage, where meaningful information is extracted from the neural data; a classification stage, where a decision is interpreted from the data; and a feedback stage where the result of that decision is provided to the user. Each distinct BCI paradigm relies on different aspects of the neural signal obtained from the user, meaning that each paradigm uses distinct and specialized methods for signal processing [18], feature extraction [19] and classification [20]. Many feature extraction and classification methods have been used for EEG-based BCI, including spatially filtering the EEG data (e.g. Common Spatial Patterns, or CSP), spectrally filtering the data (e.g. Fourier or autoregressive methods), dimension reduction methods such as Principal Component Analysis (PCA), blind source separation methods such as Independent Components Analysis (ICA), and source localization methods such as LORETA [21–25]. Unfortunately, the optimal combination of feature extraction and classification approach has to be defined manually for each BCI application, a process which often requires significant subject-matter expertise and a priori knowledge about the expected EEG signal [20]. There is also no guarantee that the same combination of feature extraction and classifier will provide acceptable performance across BCI applications, limiting the overall utility of these systems.

Deep Learning has largely alleviated the need for manual feature extraction, achieving state-of-the-art performance in fields such as computer vision and speech recognition [26, 27]. The term Deep Learning refers to a class of multi-layered artificial neural network models aimed at learning abstract feature representations and classifiers from raw data, with the derived feature representation depending on the architecture used. The use of convolutional neural networks (CNNs) in particular has increased significantly, due in part to the success of the first large-scale CNN for image classification that won the ImageNet 2012 competition [28]. This model obtained gains of over 10% in top-5 accuracy, where gains in previous years managed only a few percentage points. Subsequent models published by various research groups have improved upon this approach, with current classification performance rivaling that of human labelers [27, 29, 30]. The main advantage of this approach lies in the ability of the network to automatically extract relevant features for the problem at hand, rather than relying on manual feature extraction approaches.

Recently, Deep Learning approaches have been used to model neurophysiological data, in functional magnetic resonance imaging (fMRI) [31] as well as EEG [32–35]. For example, Långkvist et al. used deep belief networks (DBNs) and restricted boltzmann machines (RBMs), together with hidden Markov models, to temporally model different stages of sleep using EEG [36]. Mirowski et al. used a temporal convolutional network on EEG signals for predicting epileptic events [37]. Bashivan et al. used convolutional recurrent networks for classification of mental workload by temporally stacking EEG “images” using frequency transforms to form a “movie”, where they then used

recurrent neural network approaches, originally developed for video classification, to predict mental workload [32]. Stober et. al. investigated using convolutional networks on frequency-transformed EEG signals for classification among different classes of listened music [34, 38]. Temporal convolutional networks have also been used for the detection of visual-evoked potentials in [39] and in [33]. However, a comprehensive analysis of the performance of a single model architecture across multiple EEG modalities has not been previously explored.

In this work we introduce *EEGNet*, a compact fully convolutional network for EEG-based BCIs developed using Deep Learning methodologies. EEGNet differs from previous convolutional models for EEG in several ways. First, our model removes the dependence on the channel layout through the use of spatial filtering in the first layer [21]. This operation serves to both improve the signal-to-noise (SNR) ratio and reduce the dimensionality of the EEG signal of interest. Second, we focus on spatiotemporal convolutions in the spatial filter space to capture both spatial and temporal relationships in the EEG. Finally, our model omits fully-connected layers in an effort to reduce the total number of parameters, a strategy inspired by the work of [40]. We evaluate our model against the current state-of-the-art approaches for four data collections representing four different BCI paradigms: P300 visual-evoked potential (P300), error-related negativity (ERN), movement-related cortical potential (MRCP) and the sensory motor rhythm (SMR). For binary classification problems, EEGNet has approximately 2200 parameters, which is half the total number of parameters compared to the model proposed by [41] and has much fewer parameters than the $\sim 10K$ parameters of the smallest model proposed by [32]. We fit several models, all with the same number of free parameters, to statistically control for the effect of model capacity versus performance. Our models are also trained through a cross-subject cross-validation procedure, implying that they are user-independent, without requiring any test subject information [42, 43], whereas previous approaches have mainly trained within subject [33, 39, 41]. We provide preliminary evidence suggesting that our model can be used with small EEG databases while still obtaining, and in some cases improving upon, the state-of-the-art performance across several tasks and across subjects, challenging the notion that large datasets are required to obtain optimal performance.

2 Background

BCIs are generally categorized into two types, depending on the EEG feature of interest [42]: event-related and oscillatory. *Event-Related Potential* (ERP) BCIs are designed to detect an EEG response to a known, time-locked external stimulus. They are generally robust across subjects and contain well-stereotyped waveforms, enabling the exact time course of the ERP to be modeled through machine learning efficiently [44]. BCI systems can also leverage desynchronization/synchronization of EEG oscillations, such as those which might occur during a self-paced mental task or a change in user mental state. In contrast to ERP-based BCIs, which rely mainly on the detection of the ERP waveform from some external event or stimulus, *Oscillatory* BCIs use the signal power of specific EEG frequency bands for external control and are generally not time-locked to an external stimulus [45]. When oscillatory signals are time-locked to an external stimulus, they can be represented through event-related spectral perturbation (ERSP) analyses [46]. Oscillatory BCIs are more difficult to train, generally due to the lower SNR as well as greater variation across subjects [45]. Oscillatory BCIs are also more susceptible to external noise sources than ERP BCIs, and thus require more data and/or more advanced signal processing approaches to design effective systems [42]. An overall summary of the datasets used to evaluate EEGNet can be found in Table

Paradigm	Feature Type	# of Subjects	Trials per Subject	Train/Test/Validation Subjects	# of Classes
P300	ERP	15	~ 670	13/1/1	2
Feedback ERN	ERP	16	~ 200	14/1/1	2
MRCP	ERP/Oscillatory	13	~ 1100	11/1/1	2
SMR	Oscillatory	109	~ 90	64/15/30	2

Table 1: Summary of Data Collections

1.

2.1 Datasets: Event-Related Potentials

2.1.1 Paradigm 1: P300 Event-Related Potential (ERP)

The P300 event-related potential is a stereotyped neural response to novel visual stimuli [47]. It is most commonly elicited with the visual oddball paradigm, where participants are shown repetitive “nontarget” visual stimuli that are interspersed with infrequent “target” stimuli at a fixed presentation rate (for example, 1 Hz). Observed over the parietal cortex, the P300 waveform is a large positive deflection of electrical activity observed approximately 300 ms post stimulus onset, the strength of the observed deflection being inversely proportional to the frequency of the target stimuli. The P300 ERP is one of the strongest neural signatures observable by EEG, especially when the target presentation rate is infrequent [47]. When the image presentation rate increases to 2 Hz or more, it is commonly referred to as rapid serial visual presentation (RSVP), which has been used in several BCIs for large image database triage [43, 48, 49].

The EEG data used here has been previously described in [49]; a brief description is given below. 18 participants volunteered for an RSVP BCI study. Participants were shown images of natural scenery at 2 Hz rate, with images either containing a vehicle or person (target), or with no vehicle or person present (nontarget). The target/nontarget ratio was 20%/80%. Data from 15 participants (9 male and 14 right-handed) who ranged in age from 18 to 57 years (mean age 39.5 years) were further analyzed. EEG recordings were digitally sampled at 512 Hz from 64 scalp electrodes arranged in a 10-10 montage using a BioSemi Active Two system (Amsterdam, The Netherlands). Continuous EEG data were referenced offline to the average of the left and right earlobes, digitally bandpass filtered to 1-40 Hz and downsampled to 128 Hz. EEG trials of target and nontarget conditions were extracted at $[0, 1]$ s post stimulus onset, and used for a two-class classification.

2.1.2 Paradigm 2: Feedback Error-Related Negativity (ERN)

Error potentials are perturbations of the EEG following an erroneous or unusual event in the subject’s environment or task. They can be observed in a variety of tasks, including time interval production paradigms [50] and in forced-choice paradigms [51, 52]. Here we focus on the feedback error-related negativity (ERN), which is an amplitude perturbation of the EEG following the perception of an erroneous feedback produced by a BCI. The feedback ERN is characterized as a large negative deflection approximately 300ms after feedback, followed by a positive component 500ms to 1s after feedback (see Figure 7 of [53] for an illustration). The detection of the feedback ERN provides a mechanism to infer, and to possibly correct in real-time, the incorrect output of a BCI. This two-stage system has been proposed as a hybrid BCI in [54, 55] and has been shown to improve the performance of a P300 speller in online applications [56].

Layer	Input ($C \times T$)	Operation	Output	Number of Parameters
1	$C \times T$	16 x Conv1D (Cx1)	$16 \times 1 \times T$	$16C + 16$
	$16 \times 1 \times T$	BatchNorm	$16 \times 1 \times T$	32
	$16 \times 1 \times T$	Reshape	$1 \times 16 \times T$	
	$1 \times 16 \times T$	Dropout (.25)	$1 \times 16 \times T$	
2	$1 \times 16 \times T$	4 x Conv2D (2x32)	$4 \times 16 \times T$	$4 \times 2 \times 32 + 4 = 260$
	$4 \times 16 \times T$	BatchNorm	$4 \times 16 \times T$	8
	$4 \times 16 \times T$	Maxpool2D (2,4)	$4 \times 8 \times T/4$	
	$4 \times 8 \times T/4$	Dropout (.25)	$4 \times 8 \times T/4$	
3	$4 \times 8 \times T/4$	4 x Conv2D (8x4)	$4 \times 8 \times T/4$	$4 \times 4 \times 8 \times 4 + 4 = 516$
	$4 \times 8 \times T/4$	BatchNorm	$4 \times 8 \times T/4$	8
	$4 \times 8 \times T/4$	Maxpool2D (2,4)	$4 \times 4 \times T/16$	
	$4 \times 4 \times T/16$	Dropout (.25)	$4 \times 4 \times T/16$	
4	$4 \times 4 \times T/16$	Softmax Regression	N	$TN + N$
Total				$16C + N(T + 1) + 836$

Table 2: Convolutional Network Architecture, where C = number of channels, T = number of time points and N = number of classes, respectively. For Layers 1-3, the Exponential Linear Unit (ELU) activation function is used.

The EEG data used here comes from [53] and was used in the “BCI Challenge” hosted by Kaggle (<https://www.kaggle.com/c/inria-bci-challenge>); a brief description is given below. 16 healthy participants participated in a P300 speller task, a system which uses a random sequence of flashing letters, arranged in a 6×6 grid, to elicit the P300 response [15]. EEG data was recorded at 600Hz using 56 passive Ag/AgCl EEG sensors (VSM-CTF compatible system) following the extended 10-20 system for electrode placement. The EEG data was referenced offline to an electrode placed on the nose, band-pass filtered to 1-40 Hz and down-sampled to 128Hz. EEG trials of correct and incorrect feedback were extracted at $[0, 1]$ s post feedback presentation and used as features for a two-class classification.

2.2 Datasets: Oscillatory

2.2.1 Paradigm 3: Movement-Related Cortical Potential (MRCP)

Some neural activities contain both an ERP component as well as an oscillatory component. One particular example of this is the movement-related cortical potential (MRCP), which can be elicited by voluntary movements of the hands and feet and is observable through EEG along the central and midline electrodes, contralateral to the hand or foot movement [57]. The oscillatory component of the MRCP can be seen both before movement onset (an early desynchronization in the 10-12Hz frequency band) as well as after movement onset (a late synchronization of 20-30Hz activity approximately 1s after movement execution). The ERP component of the MRCP occurs at the start of the movement, with a duration of approximately 800ms. The MRCP has been used previously to develop motor control BCIs for both healthy and physically disabled patients [58].

The EEG data used here has been previously described in [59]; a brief description is given below. In this study, 13 subjects performed self-paced finger movements using the left index, left middle, right index, or right middle fingers. This produced the well-known alpha and beta synchronizations (i.e. increases in power) and desynchronizations (i.e. decreases in power), most clearly observed over the contralateral motor cortex [60–62]. The data was originally recorded using a 256 channel BioSemi Active II at 1024 Hz. Due to extensive signal noise present in the data, the EEG data were first processed with the PREP pipeline [63]. The data was referenced to linked mastoids, bandpass filtered between 0.3 Hz and 50 Hz, and then downsampled to 128 Hz. We further downsampled the

channel space to the standard 64 channel BioSemi montage. The index and middle finger blocks for each hand were combined for binary classification of movements originating from the left or right hand. The classes are approximately balanced, with each subject having about 500 trials per class. EEG trials of left and right hand finger movements were extracted at $[-.5, 1]$ s around finger movement onset and used for a two-class classification.

2.2.2 Paradigm 4: Sensory Motor Rhythm (SMR)

A common control signal for oscillatory-based BCI is the sensorimotor rhythm (SMR), wherein mu (8-12Hz) and beta (18-26Hz) bands desynchronize over the sensorimotor cortex contralateral to an actual or imagined movement. The SMR is very similar to that of the oscillatory component of the MRCP. While SMR-based BCIs can facilitate nuanced, endogenous BCI control (enabling high dimensional control of cursors or even prosthetic limbs) they are not without their practical challenges. As signals, SMRs tend to be weak and highly variable across and within subjects, conventionally demanding user-training (neurofeedback) and long calibration times (20 minutes) in order to achieve reasonable performance [42].

The EEG data used here is from the PhysioNet EEG motor movement/imagery dataset (N=109) [64,65]. The EEG data was recorded using 64 channels following the aforementioned 10-10 system for electrode placement, at a sampling rate of 160 Hz. The EEG data were band-pass filtered at 0.1-40Hz. A visual cue appeared to the subjects for 4.1s indicating which task to perform. Periods of activity were separated by 4.1s of rest. To avoid confounding neural activity due to the onset of the visual cue, we omitted from analysis the data 1s after the onset of the cue box. For the remaining 3.1s of data, we extracted two trials of 2 seconds length, with a 0.9 second overlap. The goal of the analysis is to predict the imagined movements of the right or left hand.

3 Methods

3.1 Convolutional Network Architecture

Our EEGNet model is shown in Table 2, for EEG trials having C channels and T time samples. All layers used the Exponential Linear Unit (ELU) non-linear activation function [66]. The model was estimated using Adam, a stochastic optimization algorithm using adaptive moment estimation [67], optimizing the binary cross-entropy criterion. All models were trained on an NVIDIA Quadro K6000 GPU, with CUDA 7.5 and cuDNN v5, in Theano [68], using the Keras API [69].

- In Layer 1, we learn 16 convolutional kernels of size $(C, 1)$. This operation estimates a set of spatial filters over the entire period of the trial and is similar to that of previous approaches such as Common Spatial Patterns (CSP) [21], xDAWN spatial filtering [70] and independent component analysis (ICA) [24]. Note that these approaches are specifically designed to produce spatial filters that either maximize the variance difference among two classes (CSP), enhance the signal to signal-plus-noise ratio of the EEG signal of interest (xDAWN) or produce spatial filters that are as mutually independent as possible (ICA). In contrast, the spatial filters used here are optimized to minimize the categorical cross-entropy of the predicted outputs and are not required to achieve either optimal variance separation or mutual independence. The spatial filters are regularized with an elastic-net ($L_1 + L_2$)

constraint, with $L_1 = 0.01$ and $L_2 = 0.01$. We apply Batch Normalization [71] together with Dropout [72] as this improved model robustness.

- In Layer 2, we learn 4 2-dimensional convolutional kernels of size (2×32) . Here, we use zero-padding to preserve the original dimension of the data after the convolution. 2D max-pooling is applied to the data; while this operation is traditionally done to induce invariance to image transformation in computer vision [73], we find that for processing these EEG tasks this operation is beneficial primarily for dimension reduction. The total reduction in parameter size due to max-pooling is by a factor of 8 (2×4). As in Layer 1, Batch Normalization and Dropout are used.
- In Layer 3, we learn 4 2-dimensional convolutional kernels of size (8×4) . All the operations in Layer 2 are also applied here.
- In Layer 4, the features are passed to a softmax classification with N units, N being the number of classes in the data. We omit the use of a fully connected layer prior to the softmax classification layer to reduce the number of free parameters in the model, as was done in [40].

3.2 Convolutional Network Model Comparison

In order to statistically separate the effects of the number of parameters of the model versus classification performance, we generated 12 different models, all having the same number of parameters, using the model framework described in Table 2. We generated the 12 models by setting the convolutional kernel sizes in Layer 2 to be one of 4 different values: $(16, 4)$, $(8, 8)$, $(4, 16)$ and $(2, 32)$ and the kernel sizes in Layer 3 to be one of 3 different values: $(8, 4)$, $(4, 8)$ and $(2, 16)$, respectively. Each configuration emphasizes different spatial/temporal aspects of the EEG signal.

We also compare EEGNet to a recent convolutional network model for EEG presented by Manor and Geva in [33] with some minor modifications. First, since the model in [33] was designed using EEG data sampled at 64Hz, we appropriately scaled the length of the temporal kernels to match the rate of our data (128Hz for the P300, Feedback ERN and MRCP, 160Hz for SMR) to model the same amount of time temporally. Second, [33] uses Dropout after fully-connected layers only, however the dropout fraction was not specified. Therefore, we set it to 0.5 for our comparison. Also note that the rectified linear unit (ReLU) activation function was used in [33], in contrast to the ELU that is used in EEGNet. Note that this model has approximately 11.8M parameters, which represents more than a 5000-fold increase in parameter size when compared to EEGNet. This model will be referred to as the “MG” model for the remainder of the manuscript.

3.3 Comparison with Traditional Approaches

We compare the performance of our CNN model to that of the best performing reference algorithm for each individual paradigm. For all event-related data analyses (P300 VEP, ERN, MRCP) the reference algorithm is xDAWN Spatial Filtering + Bayesian LDA [70]. For oscillatory-based classification of SMR, the reference algorithm is a CSP with covariance matrix regularization [74] with a Regularized LDA as used in [42].

3.3.1 ERP: xDAWN Spatial Filtering + Bayesian LDA

Here we provide a summary of the xDAWN algorithm; more details can be found in [70,75]. xDAWN is a spatial filtering algorithm designed to enhance the ERP through maximizing the signal to signal plus noise ratio (SSNR). The result of this algorithm are the spatial filters U , which are ranked in terms of their SSNR. The EEG signal X is decomposed into three parts as:

$$XU = (D_1A_1 + D_2A_2 + N)U \quad (1)$$

where $X \in \mathbb{R}^{T \times C}$, where T and C denote the number of time samples and channels, respectively. D_1A_1 denotes the ERP neural response, D_2A_2 denotes the neural response that is confounded with the ERP neural response whenever any stimulus is presented, and N is a residual noise term. xDAWN optimizes the SSNR, given as:

$$SSNR(U) = \underset{U}{\operatorname{argmax}} \frac{\operatorname{Tr}(U^T \hat{A}_1^T D_1^T D_1 \hat{A}_1 U)}{\operatorname{Tr}(U^T X^T X U)} \quad (2)$$

where \hat{A}_1 is the solution to the least squares problem

$$\hat{A} = \begin{bmatrix} \hat{A}_1 \\ \hat{A}_2 \end{bmatrix} = \left(\begin{bmatrix} D_1 \\ D_2 \end{bmatrix}^T \begin{bmatrix} D_1 \\ D_2 \end{bmatrix} \right)^{-1} \begin{bmatrix} D_1 \\ D_2 \end{bmatrix}^T X \quad (3)$$

where $[D_1; D_2] \in \mathbb{R}^{T \times F}$, F denotes the number of filters, and $\operatorname{Tr}(\cdot)$ denotes the trace operator. The spatially filtered EEG signals are then classified using Bayesian Linear Discriminant Analysis (BLDA) (see Appendix B of [76] for more details).

3.3.2 Oscillatory (SMR): Covariance Shrinkage Regularized CSP

The best performing user-independent model for SMR-based BCI, which was evaluated on the BCI Competition IV competition dataset and compared against 11 other approaches, is the approach outlined in [77], wherein all available subjects are pooled to train CSP filters and a linear discriminant analysis (LDA) classifier. In addition to the pooled design, the covariance matrices in both the CSP and LDA algorithms are regularized by diagonal loading, which shrinks the covariance matrix towards the identity matrix [77]. A single parameter, λ , must be identified, for which there is an automatic method outlined in [74].

3.4 EEG Data Processing

We split each data collection into a training, testing, and validation set. For Paradigms 1, 2 and 3, we sample, without replacement, $N - 2$ subjects for the train set and 1 for the test and validation sets, respectively, where N is the total number of subjects the dataset (see Table 1). Due to the relatively few trials per subject in Paradigm 4, we sample, without replacement, 64 subjects for the train set, 15 for the test and 30 for the validation set. In all cases, the data were balanced such that the number of trials in each class is equal; this was done by randomly downsampling the size of the larger class to match the size of the smallest class. For each paradigm we generated 30 unique datasets of training/testing/validation with this procedure. The train, test and validation sets were z-score normalized at each (channel x time) pair relative to the training set; if the EEG

data is arranged in a 3-dimensional matrix as (channels, time, trials), then the z-score is calculated along the third dimension.

3.5 Statistical Analysis

We calculate ΔAUC , the difference between the AUC performance between EEGNet and the reference algorithm for each paradigm for each of 30 folds, where positive values of Δ denote that EEGNet outperformed the reference algorithm. For each EEGNet model configuration a one-way t-test was performed, testing if the model significantly outperformed the reference algorithm. We use the False Discovery Rate (FDR) [78] procedure to correct for multiple comparisons.

We also conduct two additional analyses where model performance is compared relative to (1) the number of subjects in the training set and (2) the number of EEG trials in the training set. For the first analysis, we sample, without replacement, N subjects to be in the training set and increased N up to the total number of training subjects available (see Table 1). For the P300, ERN and MRCP analyses we started with $N = 1$ subject and increased by 1 subject up to the total number of training subjects available, while for SMR we started with $N = 8$ subjects and increased by 8 subjects up to the total number of training subjects available. This procedure was repeated 10 times for each value of N . For the second analysis, from the full training set we randomly sample, without replacement, K trials and set this random sample to be the training set. For the P300 analysis we start with $K = 500$ trials, then increase by 2000 trials up to the total number of available trials; the ERN analysis, we start with $K = 500$ trials, then increase by 500 ; the MRCP analysis we start with $K = 500$ trials, then increase by 2000; finally for the SMR analysis we start with approximately $K = 800$ trials, then increase by approximately 700 trials. This procedure was repeated 10 times for each value of K .

4 Results

Figure 1 shows the AUC difference in mean classification performance between the EEGNet model and the reference algorithm for each paradigm. For classification of both the P300 and SMR, we see that all models, regardless of kernel configuration, produced statistically significant improvements over that of the reference algorithm. For classification of the ERN, we see that only a few of the models were significantly better, with the $(2, 32) \times (8, 4)$ configuration having the highest performance improvement. While gains were observed for classification of the MRCP, these gains were not statistically significant ($p > .05$).

Of the 12 models tested, either the $(4, 16) \times (8, 4)$ or the $(2, 32) \times (8, 4)$ model configurations were among the top four models across all data paradigms (see Figure 1). This is true for the ERN paradigm, whose best performing model is the $(2, 32) \times (8, 4)$ configuration. While the best model for P300 was the $(8, 8) \times (4, 8)$ configuration, the next best configuration is the $(4, 16) \times (8, 4)$, which is the best configuration observed in MRCP. For SMR, the fourth-best model is the $(2, 32) \times (8, 4)$ configuration. This trend suggests that a general model configuration can be used when lack of apriori knowledge of model architecture exists for a particular analysis, although specific configurations are better for each individual paradigm.

Figure 2 illustrates the classification performance relative to the number of trials in the training set when using the $(2, 32) \times (8, 4)$ model configuration, with different colors denoting the number of subjects available in the training set. As expected, when the size of the training set increases,

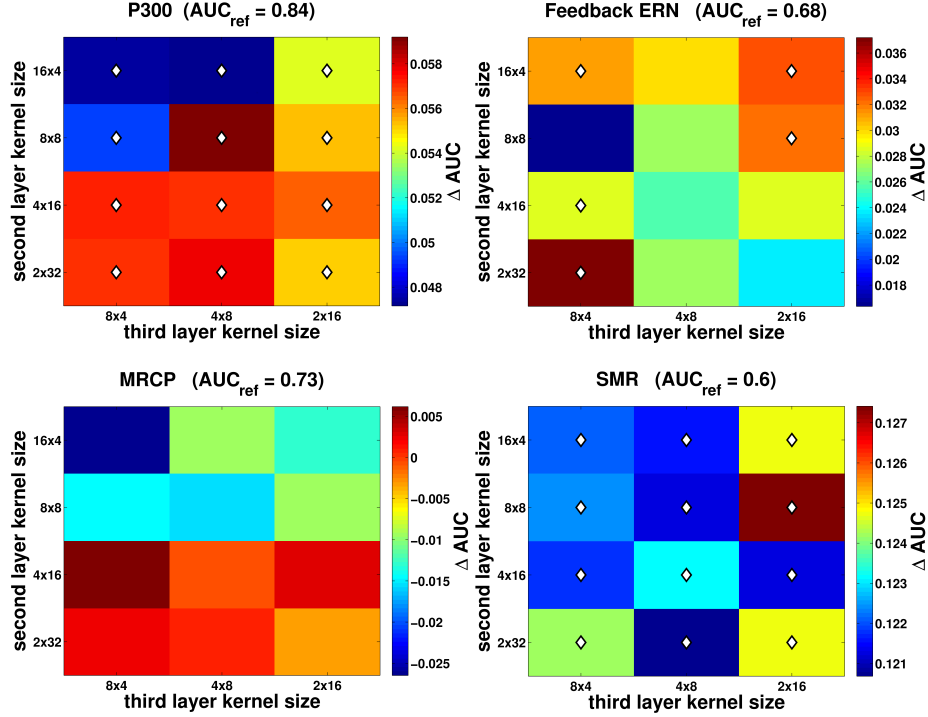


Figure 1: Classification performance for all 12 EEGNet models across all paradigms. White diamonds at the center of each (row,column) square denote a statistically significant improvement of the EEGNet model configuration relative to the reference classification model. AUC_{ref} denotes the mean classification performance for the reference algorithm averaged across all 30 folds.

the classification performance also increases while also having a smaller error variance in the prediction. The classification performance with less than 20% of the training set is often times very competitive when compared to having the full training set, suggesting that our compact model can accurately capture the dynamics of EEG across subjects with very little data. For example, for P300 classification, having 2000 training trials (slightly more than 20% of all available trials) produces a classification performance equivalent to having all trials available, albeit with a slightly larger variance in the prediction. For SMR, the classification performance of the model trained with approximately 2200 trials, which is about 33% of the training set, produces a classification performance within 5% of the model using all available training data. This trend is also observed when sampling a subset of data from all available subjects (orange open circles) and using this subset as the training set. For the Feedback ERN classification, having only 500 trials sampled across 14 subjects (which constitutes less than 20% of all available trials) produces a classification performance of approximately 0.78 AUC, which is within 5% of the classification performance from having all trials available (0.8 AUC, with ~ 2700 trials). For the MRCP, however, we found that we needed significantly more trials (at least 2500) to obtain reasonable classification performance. We believe this is due primarily to the MRCP having lower SNR than that of the P300 and the Feedback ERN. Across all paradigms we found that having at least 1 EEG trial per parameter produced consistently good performance, and that one could have fewer trials depending on the strength of the expected neural response. Beyond this threshold, increases in the training data

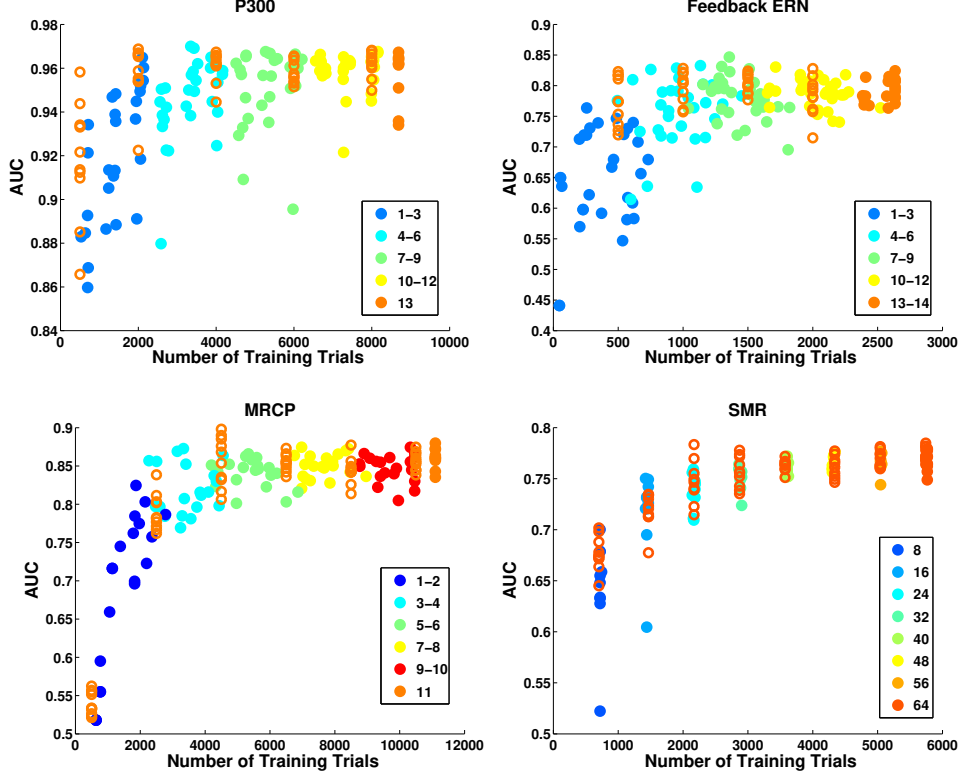


Figure 2: Classification performance relative to the number of training trials in the training set across all paradigms with the $(2, 32) \times (8, 4)$ model configuration for one randomly chosen fold. Open orange circles are the classification performance by randomly sampling a subset of trials from all subjects and using this subsample as the training set (see Section 3.5).

size primarily has the effect of reducing the error variance as opposed to improving classification performance.

Figure 3 shows the difference in AUC performance of EEGNet when compared to both the MG model [33] and the reference algorithm (xDAWN+BLDA for ERP and Regularized CSP for SMR) for each of 10 folds across several different training set sizes. EEGNet outperforms both the MG model and the reference algorithm in nearly all cases across all training set sizes. For the P300 paradigm, xDAWN has both larger errors as well as larger error variances at the smaller training set sizes, both of which decrease as the training set size increases, converging to an average of approximately 0.05 AUC difference. Generally speaking, the MG model outperforms xDAWN at all training set sizes while underperforming EEGNet. This is interesting to note, as MG contains more than 5000 times the parameters than EEGNet ($\sim 11.8\text{M}$ for MG vs. ~ 2200 for EEGNet). In contrast, for classification of the SMR, the MG model performance does not exceed chance level (0.5 AUC) at any training set size, whereas EEGNet was able to classify better than chance with only 800 trials (see Figure 2). We suspect this is due to having insufficient data given the larger model size of the MG model.

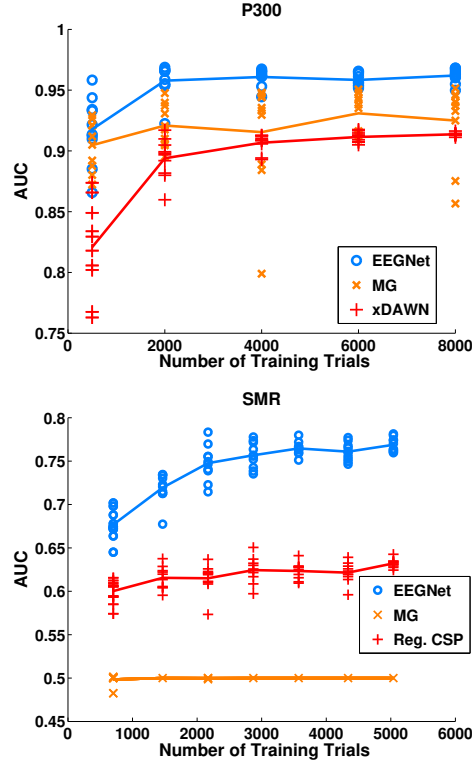


Figure 3: Difference in AUC performance of EEGNet when compared to both the MG model and the reference algorithm for the P300 and SMR paradigms for the data shown in Figure 2. Solid lines denote the mean performance at each training set size. The reference algorithm is xDAWN + BLDA for the P300 and a Regularized CSP for SMR.

5 Discussion

In this work we proposed *EEGNet*, a compact fully-convolutional network for EEG-based BCIs. We showed that our model can outperform the state-of-the-art approach over many BCI paradigms with as few as 500 training samples, and that adding more training samples primarily reduces the prediction variance as opposed to obtaining improved performance.

We note that the number of parameters in our model is directly tied to the sampling rate of the EEG (see Table 2). The EEG data used in this manuscript have a sampling rate of at least 128 Hz. This gives us a sufficient sampling rate to measure frequency content of up to 64 Hz by the Nyquist theorem, encapsulating the EEG frequency bands of interest for most BCI applications. A higher sampling rate would incur more parameters in the model, which could be offset by increasing the dimension of the maxpool used in Layers 2 and 3. In contrast, there is generally less influence on the number of channels of the EEG due to the spatial filter that is used in Layer 1; the spatial filter operation removes any channel dependence after the first layer, while the effect of the sampling rate propagates throughout all layers.

Our model employs multiple forms of regularization at each layer, which is an approach that is generally uncommon in other fields such as computer vision and natural language processing. For example, the computer vision models proposed by [29] use Dropout only after dense layers, while [71]

suggests that using Batch Normalization may remove the need for Dropout entirely. In contrast, we found that Batch Normalization with Dropout improved model performance over each approach individually, and that combining these regularizers together with elastic net regularization further improved robustness. We believe that, given the low SNR of EEG together with traditionally small datasets, that this is necessary to learn the true underlying neural signal, as opposed to learning on signal noise.

Our model comparison analysis showed that a general model configuration (shown in Table 2) can be used whenever knowledge of a particular EEG feature of interest is not known a priori. Because all the models have exactly the same number of parameters, statistically speaking the performance of the model can be primarily attributed to the kernel configuration itself. The model alternates between aggregating information along spatial (Layers 1 and 3) and temporal (Layer 2) dimensions. This makes intuitive sense in terms of efficient model parametrization using filter factorization. For example, to represent an EEG segment of size $[C, T]$, for C channels and T time points, one could model the kernel size as $[C, T]$ directly, incurring $C \times T$ total parameters, or one could model the EEG segment with two layers, with the first layer having a kernel size of $[C, 1]$ and the second layer having a kernel size of $[1, T]$, incurring a total of $C + T$ parameters. It is easy to see that both approaches model the same segment of data; however, the second approach represents a significant reduction in the total number of parameters. This parametrization trick was used by [29] as a way to efficiently build deep convolutional networks for image analysis; by replacing a 7×7 kernel size with 3 layers of 3×3 kernels, one can represent the same amount of data with about a 44% reduction in parameters. We essentially use the same strategy here to represent EEG time series in a compact manner by aggregating information over the spatial and temporal dimensions iteratively.

Because our model was validated using a cross-subject cross-validation procedure, our model is essentially *user-independent*; the model requires no observations from the test subject to achieve reasonable classification performance. As a result, this model has immediate implications for practical BCI systems, as the development of reliable user-independent methods is still an active area of research interest [42, 43]. One question that arises in this situation is how to effectively update a model using trials obtained from the test subject, especially when the training dataset used to train the model is large. We will explore how to update our model as EEG trials are recorded in real time in future research.

The first three layers of EEGNet can potentially be used as a preprocessing step prior to using recurrent neural networks for modeling long temporal sequences in the EEG. Recently, a recurrent convolutional network was proposed by [32] to predict mental workload. Their approach first used fourier transforms to convert the EEG time series into a time-frequency representation. This representation was then used in a recurrent neural network to model temporal dependencies in the time-frequency space. In our future work we will explore a similar approach, where we will replace the manually derived representation (time-frequency space) with the automatically-derived representation from the EEGNet model as the initial feature extraction, with the goal of analyzing non-time-locked EEG events temporally. We will also test the generalizability of this approach for describing other kinds of EEG phenomena, for example performance or alertness monitoring, in future research.

6 Conclusion

In this manuscript we present *EEGNet*, a compact convolutional network for EEG-based BCIs. The model is computationally efficient, and for some paradigms, can be estimated with as few as 500 EEG trials. We showed that the model can outperform the current state-of-the-art across several BCI paradigms. Because our model is validated using a cross-subject cross-validation procedure, our model is user-independent, which has immediate applications for practical BCI technologies. We show significant gains in classification accuracy of EEGNet across multiple paradigms when compared to a recent CNN model proposed by [33]. Finally, this model can potentially serve as the initial feature extraction to more advanced neural network architectures such as recurrent neural networks to efficiently model long EEG sequences.

Acknowledgment

This project was supported by the Office of the Secretary of Defense Autonomy Research Pilot Initiative program MIPR DWAM31168. The views and the conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Government.

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 : official journal of the International Federation of Clinical Neurophysiology*, vol. 113, no. 6, pp. 767–91, jun 2002. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/12048038>
- [2] A. B. Schwartz, X. T. Cui, D. Weber, and D. W. Moran, “Brain-controlled interfaces: Movement restoration with neural prosthetics,” *Neuron*, vol. 52, no. 1, pp. 205 – 220, 2006.
- [3] B. J. Lance, S. E. Kerick, A. J. Ries, K. S. Oie, and K. McDowell, “Brain-computer interface technologies in the coming decades,” *Proceedings of the IEEE*, vol. 100, no. Special Centennial Issue, pp. 1585–1599, May 2012.
- [4] S. Saproo, J. Faller, V. Shih, P. Sajda, N. R. Waytowich, A. Bohannon, V. J. Lawhern, B. J. Lance, and D. Jangraw, “Cortically coupled computing: A new paradigm for synergistic human-machine interaction,” *Computer*, vol. 49, no. 9, pp. 60–68, Sept 2016.
- [5] J. van Erp, F. Lotte, and M. Tangermann, “Brain-Computer Interfaces: Beyond Medical Applications,” *Computer*, vol. 45, no. 4, pp. 26–34, Apr. 2012.
- [6] B. Blankertz, M. Tangermann, C. Vidaurre, S. Fazli, C. Sannelli, S. Haufe, C. Maeder, L. E. Ramsey, I. Sturm, G. Curio, and K. R. Müller, “The berlin brain-computer interface: Non-medical uses of bci technology,” *Frontiers in Neuroscience*, vol. 4, no. 198, 2010.
- [7] F. Cincotti, D. Mattia, F. Aloise, S. Bufalari, G. Schalk, G. Oriolo, A. Cherubini, M. G. Marciani, and F. Babiloni, “Non-invasive braincomputer interface system: Towards its application as assistive technology,” *Brain Research Bulletin*, vol. 75, no. 6, pp. 796 – 803, 2008, special Issue: Robotics and Neuroscience.

- [8] B. Z. Allison, D. J. McFarland, G. Schalk, S. D. Zheng, M. M. Jackson, and J. R. Wolpaw, "Towards an independent braincomputer interface using steady state visual evoked potentials," *Clinical Neurophysiology*, vol. 119, no. 2, pp. 399 – 408, 2008.
- [9] T.-P. Jung, S. Makeig, M. Stensmo, and T. J. Sejnowski, "Estimating alertness from the eeg power spectrum," *IEEE Transactions on Biomedical Engineering*, vol. 44, no. 1, pp. 60–69, Jan 1997.
- [10] C.-T. Lin, R.-C. Wu, S.-F. Liang, W.-H. Chao, Y.-J. Chen, and T.-P. Jung, "Eeg-based drowsiness estimation for safety driving using independent component analysis," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 52, no. 12, pp. 2726–2738, Dec 2005.
- [11] T. O. Zander and C. Kothe, "Towards passive braincomputer interfaces: applying braincomputer interface technology to humanmachine systems in general," *Journal of Neural Engineering*, vol. 8, no. 2, p. 025005, 2011. [Online]. Available: <http://stacks.iop.org/1741-2552/8/i=2/a=025005>
- [12] F. Mormann, T. Kreuz, C. Rieke, R. G. Andrzejak, A. Kraskov, P. David, C. E. Elger, and K. Lehnertz, "On the predictability of epileptic seizures," *Clinical Neurophysiology*, vol. 116, no. 3, pp. 569 – 587, 2005.
- [13] C. C. Jouny, P. J. Franaszczuk, and G. K. Bergey, "Signal complexity and synchrony of epileptic seizures: is there an identifiable preictal period?" *Clinical Neurophysiology*, vol. 116, no. 3, pp. 552 – 558, 2005.
- [14] H. Ocak, "Automatic detection of epileptic seizures in {EEG} using discrete wavelet transform and approximate entropy," *Expert Systems with Applications*, vol. 36, no. 2, Part 1, pp. 2027 – 2036, 2009.
- [15] D. J. Krusienski, E. W. Sellers, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "Toward enhanced p300 speller performance," *Journal of neuroscience methods*, vol. 167, no. 1, pp. 15–21, 2008.
- [16] A. D. Gerson, L. C. Parra, and P. Sajda, "Cortically coupled computer vision for rapid image search," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 14, no. 2, pp. 174–179, June 2006.
- [17] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review," *Sensors*, vol. 12, no. 2, p. 1211, 2012.
- [18] A. Bashashati, M. Fatourechi, R. K. Ward, and G. E. Birch, "A survey of signal processing algorithms in braincomputer interfaces based on electrical brain signals," *Journal of Neural Engineering*, vol. 4, no. 2, p. R32, 2007. [Online]. Available: <http://stacks.iop.org/1741-2552/4/i=2/a=R03>
- [19] D. J. McFarland, C. W. Anderson, K. R. Muller, A. Schlögl, and D. J. Krusienski, "Bci meeting 2005-workshop on bci signal processing: feature extraction and translation," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 14, no. 2, pp. 135–138, June 2006.

- [20] F. Lotte, M. Congedo, A. Lcuyer, F. Lamarche, and B. Arnaldi, “A review of classification algorithms for eeg-based braincomputer interfaces,” *Journal of Neural Engineering*, vol. 4, no. 2, p. R1, 2007. [Online]. Available: <http://stacks.iop.org/1741-2552/4/i=2/a=R01>
- [21] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K. r. Muller, “Optimizing spatial filters for robust eeg single-trial analysis,” *IEEE Signal Processing Magazine*, vol. 25, no. 1, pp. 41–56, 2008.
- [22] G. Pfurtscheller, C. Neuper, A. Schlogl, and K. Lugger, “Separability of eeg signals recorded during right and left motor imagery using adaptive autoregressive parameters,” *IEEE Transactions on Rehabilitation Engineering*, vol. 6, no. 3, pp. 316–325, Sep 1998.
- [23] V. Lawhern, W. D. Hairston, K. McDowell, M. Westerfield, and K. Robbins, “Detection and classification of subject-generated artifacts in {EEG} signals using autoregressive models,” *Journal of Neuroscience Methods*, vol. 208, no. 2, pp. 181 – 189, 2012.
- [24] S. Makeig, A. J. Bell, T.-P. Jung, and T. J. Sejnowski, “Independent component analysis of electroencephalographic data,” *Advances in Neural Information Processing Systems*, pp. 145–151, 1996.
- [25] R. Pascual-Marqui, C. Michel, and D. Lehmann, “Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain,” *International Journal of Psychophysiology*, vol. 18, no. 1, pp. 49 – 65, 1994.
- [26] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A. r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, and B. Kingsbury, “Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups,” *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, Nov 2012.
- [27] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *CoRR*, vol. abs/1512.03385, 2015. [Online]. Available: <http://arxiv.org/abs/1512.03385>
- [28] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds., 2012, pp. 1097–1105. [Online]. Available: <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>
- [29] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *CoRR*, vol. abs/1409.1556, 2014. [Online]. Available: <http://arxiv.org/abs/1409.1556>
- [30] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” *CoRR*, vol. abs/1409.4842, 2014. [Online]. Available: <http://arxiv.org/abs/1409.4842>
- [31] S. M. Plis, D. R. Hjelm, R. Salakhutdinov, E. A. Allen, H. J. Bockholt, J. D. Long, H. J. Johnson, J. S. Paulsen, J. A. Turner, and V. D. Calhoun, “Deep learning for neuroimaging: a validation study,” *Frontiers in Neuroscience*, vol. 8, p. 229, 2014. [Online]. Available: <http://journal.frontiersin.org/article/10.3389/fnins.2014.00229>

- [32] P. Bashivan, I. Rish, M. Yeasin, and N. Codella, “Learning representations from EEG with deep recurrent-convolutional neural networks,” *CoRR*, vol. abs/1511.06448, 2015. [Online]. Available: <http://arxiv.org/abs/1511.06448>
- [33] R. Manor and A. Geva, “Convolutional neural network for multi-category rapid serial visual presentation bci,” *Frontiers in Computational Neuroscience*, vol. 9, no. 146, 2015.
- [34] S. Stober, A. Sternin, A. M. Owen, and J. A. Grahm, “Deep feature learning for EEG recordings,” *CoRR*, vol. abs/1511.04306, 2015. [Online]. Available: <http://arxiv.org/abs/1511.04306>
- [35] D. F. Wulsin, J. R. Gupta, R. Mani, J. A. Blanco, and B. Litt, “Modeling electroencephalography waveforms with semi-supervised deep belief nets: fast classification and anomaly measurement,” *Journal of Neural Engineering*, vol. 8, no. 3, p. 036015, 2011.
- [36] M. Långkvist, L. Karlsson, and A. Loutfi, “Sleep stage classification using unsupervised feature learning,” *Adv. Artif. Neu. Sys.*, vol. 2012, pp. 5:5–5:5, Jan. 2012. [Online]. Available: <http://dx.doi.org/10.1155/2012/107046>
- [37] P. Mirowski, D. Madhavan, Y. LeCun, and R. Kuzniecky, “Classification of patterns of {EEG} synchronization for seizure prediction,” *Clinical Neurophysiology*, vol. 120, no. 11, pp. 1927 – 1940, 2009.
- [38] S. Stober, D. J. Cameron, and J. A. Grahm, “Using convolutional neural networks to recognize rhythm stimuli from electroencephalography recordings,” in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2014, pp. 1449–1457.
- [39] 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.
- [40] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. A. Riedmiller, “Striving for simplicity: The all convolutional net,” *arXiv*, vol. abs/1412.6806, 2014. [Online]. Available: <http://arxiv.org/abs/1412.6806>
- [41] H. Cecotti, M. P. Eckstein, and B. Giesbrecht, “Single-trial classification of event-related potentials in rapid serial visual presentation tasks using supervised spatial filtering,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 11, pp. 2030–2042, Nov 2014.
- [42] F. Lotte, “Signal Processing Approaches to Minimize or Suppress Calibration Time in Oscillatory Activity-Based Brain-Computer Interfaces,” *Proceedings of the IEEE*, vol. 103, no. 6, pp. 871–890, Jun. 2015.
- [43] N. Waytowich, V. Lawhern, A. Bohannon, K. Ball, and B. Lance, “Spectral transfer learning using information geometry for a user-independent brain-computer interface,” *Frontiers in Neuroscience*, vol. 10, p. 430, 2016. [Online]. Available: <http://journal.frontiersin.org/article/10.3389/fnins.2016.00430>

- [44] R. Fazel-Rezai, B. Z. Allison, C. Guger, E. W. Sellers, S. C. Kleih, and A. Kübler, “P300 brain computer interface: current challenges and emerging trends,” *Frontiers in Neuroengineering*, vol. 5, no. 14, 2012.
- [45] G. Pfurtscheller and C. Neuper, “Motor imagery and direct brain-computer communication,” *Proceedings of the IEEE*, vol. 89, no. 7, pp. 1123–1134, Jul 2001.
- [46] S. Makeig, “Auditory event-related dynamics of the {EEG} spectrum and effects of exposure to tones,” *Electroencephalography and Clinical Neurophysiology*, vol. 86, no. 4, pp. 283 – 293, 1993.
- [47] J. Polich, “Updating p300: An integrative theory of {P3a} and {P3b},” *Clinical Neurophysiology*, vol. 118, no. 10, pp. 2128 – 2148, 2007.
- [48] P. Sajda, E. Pohlmeier, J. Wang, L. C. Parra, C. Christoforou, J. Dmochowski, B. Hanna, C. Bahlmann, M. K. Singh, and S. F. Chang, “In a blink of an eye and a switch of a transistor: Cortically coupled computer vision,” *Proceedings of the IEEE*, vol. 98, no. 3, pp. 462–478, March 2010.
- [49] A. R. Marathe, V. J. Lawhern, D. Wu, D. Slayback, and B. J. Lance, “Improved neural signal classification in a rapid serial visual presentation task using active learning,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, no. 3, pp. 333–343, March 2016.
- [50] W. H. R. Miltner, C. H. Braun, and M. G. H. Coles, “Event-related brain potentials following incorrect feedback in a time-estimation task: Evidence for a generic neural system for error detection,” *Journal of Cognitive Neuroscience*, vol. 9, no. 6, pp. 788–798, 1997.
- [51] W. J. Gehring, B. Goss, M. G. H. Coles, D. E. Meyer, and E. Donchin, “A neural system for error detection and compensation,” *Psychological Science*, vol. 4, no. 6, pp. 385–390, 1993. [Online]. Available: <http://pss.sagepub.com/content/4/6/385.abstract>
- [52] M. Falkenstein, J. Hohnsbein, J. Hoormann, and L. Blanke, “Effects of crossmodal divided attention on late {ERP} components. ii. error processing in choice reaction tasks,” *Electroencephalography and Clinical Neurophysiology*, vol. 78, no. 6, pp. 447 – 455, 1991.
- [53] P. Margaux, M. Emmanuel, D. Sébastien, B. Olivier, and M. Jérémie, “Objective and subjective evaluation of online error correction during p300-based spelling,” *Advances in Human-Computer Interaction*, vol. 2012, p. 4, 2012.
- [54] T. O. Zander, C. Kothe, S. Welke, and M. Roetting, *Utilizing Secondary Input from Passive Brain-Computer Interfaces for Enhancing Human-Machine Interaction*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 759–771.
- [55] J. d. R. Milln, R. Rupp, G. Mueller-Putz, R. Murray-Smith, C. Giugliemma, M. Tangermann, C. Vidaurre, F. Cincotti, A. Kubler, R. Leeb, C. Neuper, K. R. Müeller, and D. Mattia, “Combining brain-computer interfaces and assistive technologies: State-of-the-art and challenges,” *Frontiers in Neuroscience*, vol. 4, no. 161, 2010.

- [56] M. Spüler, M. Bensch, S. Kleih, W. Rosenstiel, M. Bogdan, and A. Kübler, “Online use of error-related potentials in healthy users and people with severe motor impairment increases performance of a p300-bci,” *Clinical Neurophysiology*, vol. 123, no. 7, pp. 1328 – 1337, 2012.
- [57] C. Toro, G. Deuschl, R. Thatcher, S. Sato, C. Kufta, and M. Hallett, “Event-related desynchronization and movement-related cortical potentials on the {ECoG} and {EEG},” *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, vol. 93, no. 5, pp. 380 – 389, 1994.
- [58] E. C. Leuthardt, G. Schalk, D. Moran, and J. G. Ojemann, “The emerging world of motor neuroprosthetics: a neurosurgical perspective,” *Neurosurgery*, vol. 59, no. 1, pp. 1–14, 2006.
- [59] S. Gordon, V. Lawhern, A. Passaro, and K. McDowell, “Informed decomposition of electroencephalographic data,” *Journal of Neuroscience Methods*, vol. 256, pp. 41 – 55, 2015.
- [60] G. Pfurtscheller and A. Aranibar, “Event-related cortical desynchronization detected by power measurements of scalp {EEG},” *Electroencephalography and Clinical Neurophysiology*, vol. 42, no. 6, pp. 817 – 826, 1977.
- [61] G. Pfurtscheller and F. L. da Silva, “Event-related eeg/meg synchronization and desynchronization: basic principles,” *Clinical Neurophysiology*, vol. 110, no. 11, pp. 1842 – 1857, 1999.
- [62] K. Liao, R. Xiao, J. Gonzalez, and L. Ding, “Decoding individual finger movements from one hand using human eeg signals,” *PLoS ONE*, vol. 9, no. 1, pp. 1–12, 01 2014.
- [63] N. Bigdely-Shamlo, T. Mullen, C. Kothe, K. M. Su, and K. A. Robbins, “The prep pipeline: Standardized preprocessing for large-scale eeg analysis,” *Frontiers in Neuroinformatics*, vol. 9, no. 16, 2015.
- [64] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, “Bci2000: a general-purpose brain-computer interface (bci) system,” *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1034–1043, 2004.
- [65] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, “Physiobank, physiotoolkit, and physionet,” *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.
- [66] D. Clevert, T. Unterthiner, and S. Hochreiter, “Fast and accurate deep network learning by exponential linear units (elus),” *CoRR*, vol. abs/1511.07289, 2015. [Online]. Available: <http://arxiv.org/abs/1511.07289>
- [67] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv*, vol. abs/1412.6980, 2014. [Online]. Available: <http://arxiv.org/abs/1412.6980>
- [68] Theano Development Team, “Theano: A Python framework for fast computation of mathematical expressions,” *arXiv e-prints*, vol. abs/1605.02688, May 2016. [Online]. Available: <http://arxiv.org/abs/1605.02688>
- [69] F. Chollet, “Keras,” <https://github.com/fchollet/keras>, 2015.

- [70] B. Rivet, A. Souloumiac, V. Attina, and G. Gibert, “xDAWN algorithm to enhance evoked potentials: Application to brain-computer interface,” *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 8, pp. 2035–2043, Aug 2009.
- [71] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” *arXiv*, vol. abs/1502.03167, 2015. [Online]. Available: <http://arxiv.org/abs/1502.03167>
- [72] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A simple way to prevent neural networks from overfitting,” *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014. [Online]. Available: <http://jmlr.org/papers/v15/srivastava14a.html>
- [73] Y.-L. Boureau, J. Ponce, and Y. LeCun, “A theoretical analysis of feature pooling in visual recognition,” in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, J. Frnkranz and T. Joachims, Eds. Omnipress, 2010, pp. 111–118. [Online]. Available: <http://www.icml2010.org/papers/638.pdf>
- [74] O. Ledoit and M. Wolf, “A well-conditioned estimator for large-dimensional covariance matrices,” *Journal of Multivariate Analysis*, vol. 88, no. 2, pp. 365 – 411, 2004.
- [75] H. Cecotti, A. R. Marathe, and A. J. Ries, “Optimization of single-trial detection of event-related potentials through artificial trials,” *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 9, pp. 2170–2176, Sept 2015.
- [76] U. Hoffmann, J.-M. Vesin, T. Ebrahimi, and K. Diserens, “An efficient p300-based braincomputer interface for disabled subjects,” *Journal of Neuroscience Methods*, vol. 167, no. 1, pp. 115 – 125, 2008, brain-Computer Interfaces (BCIs).
- [77] F. Lotte and C. Guan, “Regularizing common spatial patterns to improve bci designs: Unified theory and new algorithms,” *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 2, pp. 355–362, Feb 2011.
- [78] Y. Benjamini and D. Yekutieli, “The control of the false discovery rate in multiple testing under dependency,” *The Annals of Statistics*, vol. 29, no. 4, pp. 1165–1188, 2001. [Online]. Available: <http://www.jstor.org/stable/2674075>