Deep Learning based Prediction of EEG Motor Imagery of Stroke Patients' for Neuro-Rehabilitation Application

Haider Raza, Anirban Chowdhury, and Saugat Bhattacharyya
School of Computer Science and Electronics Engineering, University of Essex,
Colchester, England, UK

***Abstract*—Due to the non-stationary nature of electroencephalography (EEG) signals, a Brain-computer Interfacing (BCI) system requires frequent calibration. This leads to inter-session inconsistency which is one of the main reason that impedes the widespread adoption of non-invasive BCI for real-world applications, especially in rehabilitation and medicine. Domain adaptation and deep learning-based techniques have gained relevance in designing calibration-free BCIs to solve this issue. EEGNet is one such deep net architecture that has been successful in performing inter-subject classification, albeit on data from healthy participants. This is the first paper, which tests the performance of EEGNet on data obtained from 10 hemiparetic stroke patients while performing left and right motor imagery tasks. Results obtained on implementing EEGNet have been promising and it has comparably good performance as from expensive feature engineering-based approaches for both within-subject and cross-subject classification. The less dependency on feature engineering techniques and the ability to extract generalized features for inter-subject classification makes EEGNet a promising deep-learning architecture for developing practically feasible solutions for BCI based neuro-rehabilitation applications.**

I. Introduction

Machine learning methods have attained significant progress in many knowledge mining areas including classification, regression, and clustering [1]. However, it is still challenging to bring these technologies outside the laboratory due to the divergence in the data distributions between training and testing stages (i.e. domains) [2]. A common assumption in machine learning is that the training and testing data are drawn from the same distribution or feature space [3]. However, this assumption is often violated - we often operate in non-stationary environments. The shift in the joint distribution between training and testing domains is known as a dataset shift [4]–[6].

Major victims of such dataset shift are applications based on Brain-computer Interfaces (BCI) dealing with Electroencephalography (EEG) data [7], [8]. Such applications are often hindered by the need for repeated calibration of the BCI system for each individual participant due to large inter-subject variability in the EEG signal [9]. Even when different sessions on the same participant are considered, BCI systems need re-calibration due to the non-stationary nature of the EEG signals leading to inter-session inconsistency [10]. BCIs are often used for neuro-rehabilitation and for developing control and communication systems for patients suffering from various neurological disorders [11], [12]. Often the problem is exacerbated due to the presence of varying brain lesions among the users. With regards to neuro-rehabilitation especially, the time-consuming calibration process leads to user frustration and a lack of motivation, which can hinder the recovery process. Previous attempts to solve this problem involved 1) attempting to discover globally relevant EEG features [13], 2) the use of adaptive EEG classifiers [8], and 3) the use of reinforcement learning techniques [14].

Transfer learning is often implemented by transferring stationary and/or discriminative information invariant across the subjects [15], [16]. Apart from globally relevant feature representation, other approaches to transfer learning involve ensemble learning [17], [18] and domain adaptation of classifiers [19]. A variant of the popularly used common spatial pattern (CSP) based spatial filtering, called composite CSP, proposed by Kang and colleagues, was one of the earliest efforts of inter-subject transfer learning using EEG signals [20]. Another variant of CSP called stationary subspace CSP (ssCSP) proposed by Samek and colleagues focuses on transferring stationary information from various subjects and learning a stationary subspace of the CSP matrix [21]. They showed that such an approach not only leads to better performance for inter-subject classification but also relevant to the neurophysiological changes in the brain. However, a study conducted on a large number of subjects showed that the method of using second-order baseline reduces the inter-subject variability and performs better than other popular CSP based methods for subject independent BCI without calibration [22]. As no feedback is provided during the calibration phase, the naive BCI users often find it less motivating. This degrades the quality of the recorded signal during the calibration stage making it less relevant during the feedback.

Recently, following the success of deep learning-based algorithms in image processing applications inroads have been made in the field of biomedical engineering, especially in the classification of brain signals, where reliable and stable performance is still a challenge after more than two decades of research. Lu and colleagues proposed a deep belief network method using restricted Boltzmann machine (RBM) for motor imagery (MI) classification [23]. Different architectures of deep convolutional neural network (CNNs) have also been explored for decoding EEG signals [24]. A CNN with stacked autoencoders (SAEs) has been shown to achieve better classification accuracy on BCI competition IV-2b dataset than the traditional classification approaches [25], [26]. However, none of these deep learning-based decoders addressed the issue of inter-subject transfer learning in BCI. Some recent studies [27]–[30] dealt with the problem of inter-subject transfer learning for EEG classification with limited success. Notably, the CNN architecture EEGNet proposed by Lawhern and colleagues [28] has shown the potential to be generalised across different BCI paradigms including sensori-motor rhythm (SMR). The performance of EEGNet-8,2 was found to be similar to the state-of-the-art FBCSP method [31] for within-subject classification of MI training data. For inter-subject classification, the EEGNet-8,2 slightly outperformed FBCSP, although the difference is not significant and the overall accuracy was low. However, an important advantage of using EEGNet over traditional methods is that EEGNet learns directly from the raw data, which bypasses the requirement for feature engineering. Previously, the performance of the BCI system majorly depends upon the extracted features. Thus, even if the performance of EEGNet is similar to FBCSP it is worth using if it could avoid the need for subject-specific tuning of the classifier and learn the features automatically from raw data. From our past experience of using BCI for neuro-rehabilitation [12], [32], we realised the need for calibration-free BCI as the stroke patients are the most susceptible to get frustrated during repeated calibration. We have previously used covariate shift adaptation technique [11] to adapt the EEG classifier according to the shift in data distribution and tested its feasibility on stroke patients. However, such a technique is still heavily dependent on the training data for initial parameter generation and not suitable for the calibration-free BCI system. The promising results of EEGNet on healthy subjects' data motivate us to further test its feasibility on stroke patients data for inter-subject decoding in order to realise calibration-free BCI for neuro-rehabilitation. Here, we present the performance of EEGNet on 10 hemiparetic stroke patients data for the inter-subject decoding of left-hand vs. right-hand MI [11]. To the best of authors' knowledge, this is the first implementation of any CNN based architecture on patients' EEG data for MI classification.

This paper is organized as follows. In Section II, we describe the dataset and modified EEGNet architecture implemented on this patient dataset. Parameters setting and results of EEGNet under two conditions: 1) within-subject classification and 2) cross-subject classification are shown in Section III. A discussion on the findings and its significance are presented in Section IV, followed by conclusion in Section V.

II. Materials and Methods

A. Dataset

We have used a dataset recorded from patients first reported in our previous research [11]. The experimental protocol and training/testing timing diagram are illustrated in Fig 1 and Fig 2, respectively. It involves a traditional sensory-motor rhythm

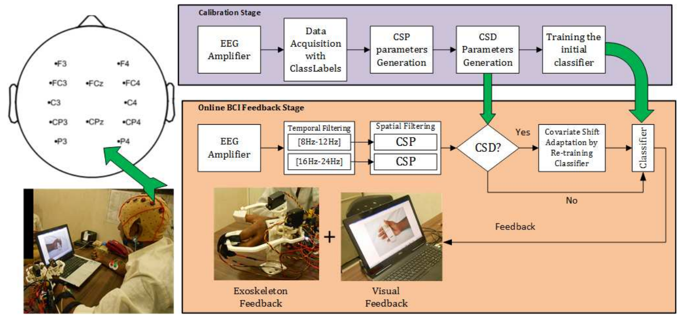


Fig. 1: Experimental protocol for the dataset recorded in our work [11]

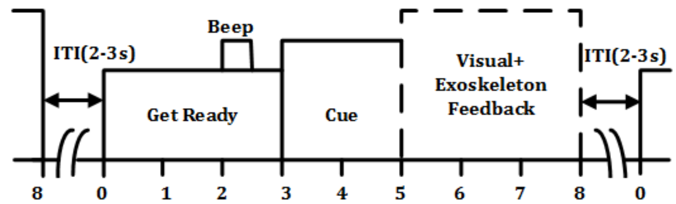


Fig. 2: Timing diagram of single-trial dataset recorded in our work [11]

(SMR) architecture consisting of two different phases. The first phase is the data acquisition without providing any feedback and data collected from this phase was used to train a classifier. The second phase is an online BCI that provides neuro-feedback on the basis of the classifier output. Data acquisition during the first phase has two runs of 40 trials and each run takes about 7 min and 30 s to complete, which is followed by one feedback run of 40 trials. In each run, the trials are equally distributed i.e., 20 trials of right-hand class and 20 trials of left-hand class. The gap between the end of phase one (i.e. training) and phase two (i.e. testing) is 16 min, which is feasible in rehabilitation settings as patients may lose attention and get tired in long runs of experiments. It is to be noted that during the original experiment the classifier was built on the common-spatial pattern (CSP) based features and re-trained during the online neuro-feedback using covariate-shift-detection (CSD) technique as shown in Fig. 1. While in this paper the data is analysed offline to evaluate the performance of EEGNet.

B. EEGNet Architecture

The EEG-based BCI system is implemented by using a compact CNN for single-trial classification (i.e. EEGNet [33]). In EEGNet, depthwise and separable convolutions were combined to construct an EEG-specific network that summarizes a few well-known EEG feature extraction methods such as optimal spatial filter with the filter bank. It also reduces the number of trainable parameters for the deep predictive model when compared with existing CNN-based EEG classification methods. For comparative evaluation, it's single-trial classification accuracy is compared with adaptive and non-adaptive methods [11].

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_1 (InputLayer)         [(None, 1, 12, 1536)]     0
_________________________________________________________________
conv2d (Conv2D)              (None, 8, 12, 1536)       200
_________________________________________________________________
batch_normalization (BatchNo (None, 8, 12, 1536)       32
_________________________________________________________________
depthwise_conv2d (DepthwiseC (None, 16, 1, 1536)       192
_________________________________________________________________
batch_normalization_1 (Batch (None, 16, 1, 1536)       64
_________________________________________________________________
activation (Activation)      (None, 16, 1, 1536)       0
_________________________________________________________________
average_pooling2d (AveragePo (None, 16, 1, 384)        0
_________________________________________________________________
dropout (Dropout)            (None, 16, 1, 384)        0
_________________________________________________________________
separable_conv2d (SeparableC (None, 16, 1, 384)        512
_________________________________________________________________
batch_normalization_2 (Batch (None, 16, 1, 384)        64
_________________________________________________________________
activation_1 (Activation)    (None, 16, 1, 384)        0
_________________________________________________________________
average_pooling2d_1 (Average (None, 16, 1, 48)         0
_________________________________________________________________
dropout_1 (Dropout)          (None, 16, 1, 48)         0
_________________________________________________________________
flatten (Flatten)            (None, 768)               0
_________________________________________________________________
dense (Dense)                (None, 2)                 1538
_________________________________________________________________
softmax (Activation)         (None, 2)                 0
=================================================================
Total params: 2,602
Trainable params: 2,522
Non-trainable params: 80
```

Fig. 3: EEGNet architecture in cross-patient settings.

EEGNet is a compact CNN architecture, which can be applied in different BCI paradigms such as sensory-motor rhythm (SMR), P300 visual evoked potentials, error-related negativity responses (ERN), and movement-related cortical potentials (MRCP). One of the advantages of EEGNet is that it is trainable on a limited amount of data acquired during the calibration phase and can produce separable features. We have slightly modified the EEGNet model-based on the requirements of our dataset. For EEG trials, data were collected at 512 Hz sampling rate, having 12 channels. For each trials, we had used MI-related epochs of 3 s for analysis in this study. A model summary of EEGNet in cross-patient settings can be found in 3. The architecture of the model is defined as follows:

1) Block 1: combination of Conv2D and DepthwiseConv2D: In block 1, starting with input layer there are two convolution steps: first, 2D convolution filter followed by batch normalisation; second, depthwise convolution followed by batch normalisation. One of the advantages in using depthwise convolution is that it helped in reducing the number of trainable parameters to fit a deep predictive model. Notably, as an advantage, depth-wise convolution is not fully-connected to all previous feature maps, which makes lesser parameters to fit. Here in the case of EEG, combined Conv2D and depthwiseConv2D provided a direct way to learn spatial filters for each temporal filter. A depth parameter controlled the number of spatial filters to learn for each feature map. This combination is inspired by the filter-bank common spatial pattern (FBCSP) algorithm, where the spatio-temporal features were prepared during the learning process.

2) Block 2: separable convolution: In block 2, after receiving inputs from block 1, a depth-wise convolution was followed by point-wise convolution. There are two main advantages of using separable convolutions. First, it reduces the number of parameters to fit; and second, principally separating the relationship with and across feature map by learning a kernel and summarising each feature map individually by optimally merging the output. In other words, this method separates learning on how to summarise individual feature maps in time using depth-wise convolution and learns how to optimally combine feature maps using point-wise convolution. This method represents different feature maps at different time-scales and combines the output afterwards.

3) Block 3: classification: In block 3, the features are passed to a softmax/sigmoid function. The softmax function is used here because EEGNet is a multi-class classification model. However, our data consisting of binary classes should give the same results as softmax is a generalization of sigmoid for a larger number of classes [34].

The next section presents the results for two different cases: 1) within-subject: is a type of experimental design in which a predictive model is trained and tested on data of each subject, albeit acquired at different run/session of recording. 2) cross-subject: is a type of experimental design in which the subjects are divided into two groups. The first group that consists of data from 9 subjects are used to train a predictive model while the second group is made of data from remaining subjects (i.e., $10th$ in this study) that are used to evaluate the performance of the model. In cross-subject, the procedure is repeated 10 times, so each subject get a chance to be selected for performance evaluation.

III. Results

A. Parameters setting

A description of the setting parameters is given as follows:

a) EEG dataset is stored in a 3D format (N, C, T), where N is the number of trials, C is the number of channels, and T is the time samples). b) The EEG data were band-pass filtered from 8 Hz as a lower cut-off to the variable limit as a parameter to be selected for upper cut-off comprising of [24, 30, 40] Hz. c) Model parameters: In block 1, 2D convolution filter of size (1,flt_size), where generally it is recommended to use a filter length that equals to half the sampling rate because it captures the information from 2 Hz and above. In our study, we have evaluated the results on three different kernel lengths (flt_size = [32, 64,128]) (i.e. temporal filter); depth-wise convolution of size (C,1), where C is number of channels (i.e. C = 12) to learn the spatial filter with depth parameter D controls the number of spatial filters to learn for each feature map. In block 2, separable convolution is used (i.e. depth-wise convolution) of size (1,16). d) Fitting parameters: the models was fitted using 'adam' optimizer and minimized by ‘categorical_crossentropy’ function with number of epochs (i.e. epochs = [100, 300, 500]). The codes were executed on Google Colab [35] environment, where Tesla K80 GPU is freely available. Deep learning Tensorflow [36] and Keras API [37] were used to create the learning model.

B. Within-subject classification

Table I compares the performance of EEGNet with adaptive and non-adaptive methods [11]. The EEGNet was trained for a different number of epochs during the training and interestingly, the testing accuracy gradually improved with an increase in the number of epochs. We performed grid search and obtained a set of best parameter (i.e. freq [8-24], dropout = 0.25, and flt_size (i.e. kernel length = 64 for 100 epochs of training and kernel length = 32 for 300 and 500 epochs of training)). The average test classification accuracy under different number of training epochs is given as follows: 1) 100 epochs of training: test accuracy 66.75$\pm$15.90%; 2) 300 epochs of training: test accuracy 68.50$\pm$15.86%; and 3) 500 epochs of training: test accuracy 70.25$\pm$16.56%. In [11], with non-adaptive classifier (EEG-NAC), the average classification accuracy is 70.25%, which is same as the performance of EEGNet with 500 epochs of training. In the case of adaptive classifier [11], the EEGNet has 5% less average accuracy. It is important to note that in adaptive classifier [11], the feedback trigger time instant (FTTI) was selected for each patient individually. However, in EEGNet, the raw data was given as input and data were in the range [8-24] Hz. The result obtained using grid search is illustrated in Fig 4 using a heatmap, where the x-axis is for flt_size (i.e. kernel length) and the y-axis is for dropout. Fig 4(a), (b) and (c) shows the average test accuracy with 100, 300 and 500 epochs of training, respectively. The training and validation performance for the subjects with 500 epochs are illustrated in Fig 5, where the parameters are as follows: freq [8-24], dropout = 0.5, and flt_size (i.e. kernel length = 32).

C. Cross-subject classification

Cross-subject classification results are shown in Table II. In cross-subject analysis, an increased number of epochs in training did not help in performance improvement. Here, we had performed grid search in a similar fashion to within-subject settings to get the best set of parameters. The average test classification accuracy under different number of training epochs are given as follows: 1) 100 epochs of training: test accuracy 69.75$\pm$11.89% with freq [8-24], dropout = 0.25, and flt_size (i.e. kernel length = 32) ; 2) 300 epochs of training: test accuracy 67.00$\pm$10.72% with freq [8-24], dropout = 0.25, and flt_size (i.e. kernel length = 128); and 3) 500 epochs of training: test accuracy 64.50$\pm$12.73% with freq [8-24], dropout = 0.5, and flt_size (i.e. kernel length = 32). The result obtained using grid search is illustrated in Fig 6 using a heatmap, where the x-axis is flt_size (i.e. kernel length) and the y-axis is a dropout. Fig 6(a), (b) and (c) reports the average test accuracy with 100, 300 and 500 epochs of training, respectively.

IV. Discussion

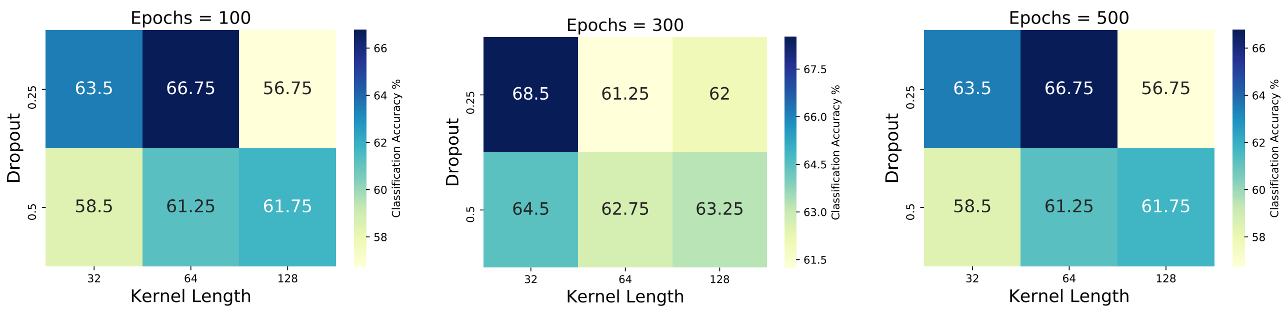
In the original EEGNet paper [28] the within subject classification accuracy for 4-class motor imagery classification on BCI Competition IV-2a data was roughly 67.25%. As our data comprises of two-classes,we had expected a much higher within-subject classification accuracy than 70.25% (with 500 epochs). However, there are two facts to be considered which may have impeded the classification accuracy. First of all, unlike the original BCI Competition IV-2a which is recorded on healthy individuals, we have used stroke patient’s data. As it is a well-known fact that brain-lesions seriously alters the dynamics of the EEG signals and hence adds more non-stationarity in the data distribution over the trials. Therefore, we may not expect a performance equivalent to a healthy subjects’ dataset. The second factor is that unlike the original EEGNet paper [28], we have reported our results on test data rather than cross-validated results on the training data. For the cross-subject analysis the classification accuracy of nearly 40% reported in [28] was not optimistic. However, in our paper the cross-subject performance of 67% was closer to the within-subject performance of 70.25%. The promises obtained from the cross-subject results has improved the chances of realising a calibration-free BCI for neuro-rehabilitation applications. In such a scenario, the data recorded by other patients and/or the data recorded by the same patient on previous days can be used as a training dataset and the patient can receive BCI based neuro-feedback right from the start of the trial. This may reduce the level of frustration and tiredness in the patients and boost their motivation towards therapeutic task.

Another important study worth comparing with is the study by Taber and colleagues [25]. They had used CNN for classification of MI EEG data provided by BCI Competition IV-2b. The average classification accuracy achieved in [25] was 72.4%, while in our case it is 70.25%. However, in [25] features were pre-constructed using time-frequency plots in the form of an image before providing it into the CNN. In contrast to that, the EEGNet architecture [28] used in this paper uses raw data as inputs to the CNN and lets CNN generate its own features. Thus one advantage of using EEGNet over time-frequency based approach is that we can skip the feature engineering part to save computational cost. This may help in designing real-time continuous prediction based neuro-feedback BCI systems using CNN, which is subjected to be validated in future works. Another important point to note is that unlike BCI Competition IV-2b, the dataset used here is a stroke patients’ dataset, which is subjected to be affected by more non-stationarity in trial to trial data distribution. Therefore, a difference of 2.15% can be attributed to lower signal quality, although no significant ($p-$value$<$0.05) difference is found in performance using a two-sample t-test. Moreover, a cross subject analysis was not provided in [25], while in this paper we have provided the cross-subject analysis along with the within-subject results.

The performance of EEGNet for cross-subject prediction also outperformed the multi-variate empirical mode decompo-

TABLE I: Within Subject Classification Accuracy with the following parameter: Frequency: [8-24] Hz, dropout: 0.25, kernel length = 64 for 100 epochs of training and kernel length = 32 for 300 and 500 epochs of training.

Subject	EEG-NAC (%)	EEG-NAC (%)	Test Acc (%)	Test Acc (%)	Test Acc (%)
Epochs			100	300	500
S01	70.00	72.50	67.50	67.50	67.50
S02	67.50	72.50	75.00	92.50	92.50
S03	75.00	82.50	50.00	50.00	52.50
S04	65.00	72.50	57.50	52.50	67.50
S05	75.00	77.50	60.00	57.50	60.00
S06	67.50	72.50	67.50	75.00	75.00
S07	67.50	75.00	97.50	82.50	92.50
S08	72.50	75.00	50.00	65.00	52.50
S09	72.50	82.50	55.00	52.50	52.50
S10	70.00	75.00	87.50	90.00	90.00
Mean	70.25	75.75	66.75	68.50	70.25
Std	3.43	3.92	15.90	15.86	16.56



(a) 100 epochs of training (b) 300 epochs of training (c) 500 epochs of training

Fig. 4: Comparison of the within subject average classification accuracy under different parameters: epochs = [100, 300, 500]; dropout = [025, 0.5], and kernel length = [32, 64, 128]

TABLE II: Cross-Subject Classification Accuracy. The parameter of are given as follows: 1) 100 epochs Frequency: [8-24] Hz, dropout: 0.25, Kernel Length = 32; 2) 300 epochs: Frequency: [8-24] Hz, dropout: 0.25, Kernel Length = 128; and 3) 500 epochs: Frequency: [8-24] Hz, dropout: 0.5, Kernel Length = 32.

Subject	Test Acc (%)	Test Acc (%)	Test Acc (%)
Epochs	100	300	500
S01	77.50	75.00	75.00
S02	85.00	80.00	80.00
S03	72.50	65.00	75.00
S04	75.00	67.50	75.00
S05	77.50	77.50	75.00
S06	55.00	45.00	47.50
S07	57.50	57.50	50.00
S08	57.50	67.50	57.50
S09	72.50	75.00	60.00
S10	67.50	60.00	50.00
Mean	69.75	67.00	64.50
Std	11.89	10.72	12.73

sition (MEMD) based filtering technique. The MEMD results for cross-subject classification reported in [38] showed the average classification accuracy of 61.75% on BCI-competition IV-2a data, while the average cross-subject classification accuracy achieved using EEGNet is 69.75%. It is important to note that the 69.75% cross-subject accuracy using EEGNet is achieved on patient data, while 61.75% accuracy using MEMD was on healthy subjects' data. This further reinforces the efficacy of EEGNet for cross-subject classification.

V. Conclusion

In this paper, we have shown the efficacy of a popular CNN based architecture, EEGNet, for classifying the stroke patients motor imagery EEG data for the first time. Results show that EEGNet can achieve satisfactory levels of accuracy for classifying stroke patients data, which is comparable to the performance on healthy subjects data for within-subject prediction. Moreover, for cross-subject prediction, it outperformed traditional MEMD based approach and the performance was even better than the healthy individuals' data. Thus, the study can have an impact on how we will be designing the BCI systems for neuro-rehabilitation in the future, paving

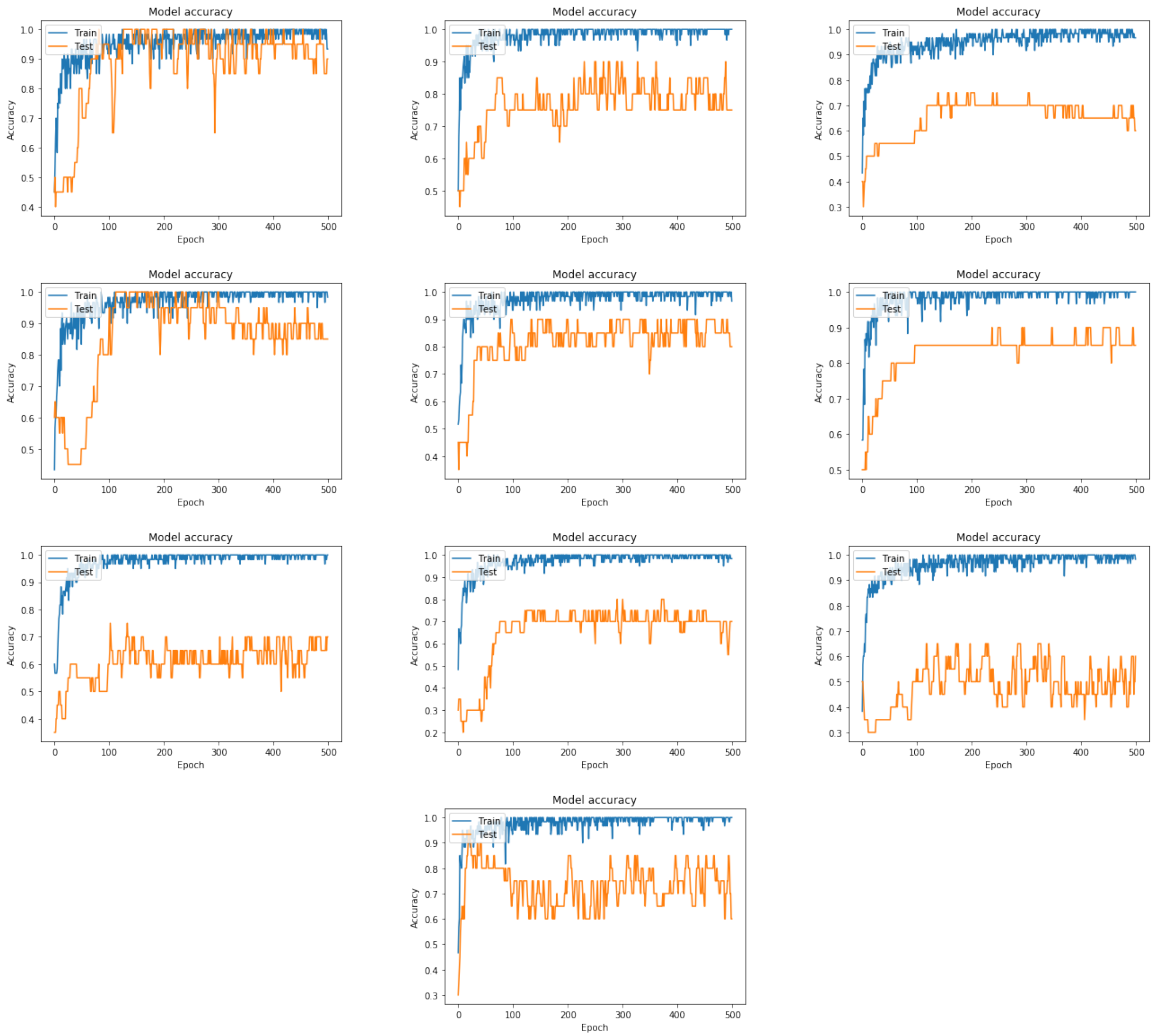


Fig. 5: EEGNet training and validation performance for all 10 subjects with 500 epochs of training, dropout = 0.25, and kernel length = 32.

the way for feature agnostic and calibration-free BCI systems, which is an important challenge in practically usable BCI design.

References

[1] C. M. Bishop, *Pattern recognition and machine learning*. springer, 2006.

[2] J. Quionero-Candela, M. Sugiyama, A. Schwaighofer, and N. D. Lawrence, *Dataset shift in machine learning*. The MIT Press, 2009.

[3] H. Raza, G. Prasad, and Y. Li, "Ewma model based shift-detection methods for detecting covariate shifts in non-stationary environments," *Pattern Recognition*, vol. 48, no. 3, pp. 659–669, Aug. 2015.

[4] J. G. Moreno-Torres, T. Raeder, R. Alaiz-RodríGuez, N. V. Chawla, and F. Herrera, "A unifying view on dataset shift in classification," *Pattern Recognition*, vol. 45, no. 1, pp. 521–530, 2012.

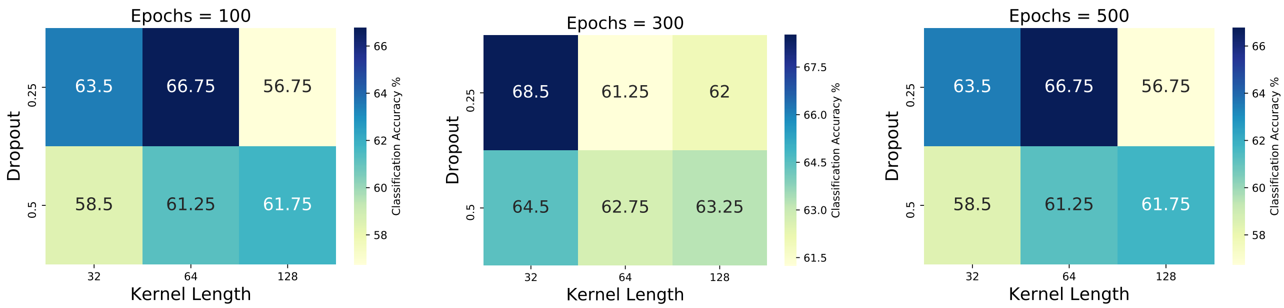
[5] H. Raza, G. Prasad, and Y. Li, "Adaptive learning with covariate shift - detection for non-stationary environments," in *Proc. 14th UK Workshop on Computational Intelligence (UKCI)*, 2014, pp. 1–8.

[6] H. Raza, H. Cecotti, and G. Prasad, "Optimising frequency band selection with forward-addition and backward-elimination algorithms in EEG-based brain-computer interfaces," in *Proc. of the Int. Joint Conf. on Neural Net.*, 2015, pp. 1–7.

[7] H. Raza, H. Cecotti, Y. Li, and G. Prasad, "Adaptive learning with covariate shift-detection for motor imagery-based brain–computer interface," *Soft Computing*, vol. 20, no. 8, pp. 3085–3096, 2016.

[8] H. Raza, D. Rathee, S.-M. Zhou, H. Cecotti, and G. Prasad, "Covariate shift estimation based adaptive ensemble learning for handling non-stationarity in motor imagery related eeg-based brain-computer interface," *Neurocomputing*, vol. 343, pp. 154–166, 2019.

[9] M. Arvaneh, I. Robertson, and T. E. Ward, "Subject-to-subject adaptation to reduce calibration time in motor imagery-based brain-computer



(a) 100 epochs of training (b) 300 epochs of training (c) 500 epochs of training

Fig. 6: Comparison of the cross subject average classification accuracy under different parameters: epochs = [100, 300, 500]; dropout = [025, 0.5], and kernel length = [32, 64, 128]

interface," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2014, pp. 6501–6504.

[10] H. Raza, H. Cecotti, Y. Li, and G. Prasad, "Learning with covariate shift-detection and adaptation in non-stationary environments: Application to brain-computer interface," in *2015 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2015, pp. 1–8.

[11] A. Chowdhury, H. Raza, Y. K. Meena, A. Dutta, and G. Prasad, "Online covariate shift detection-based adaptive brain–computer interface to trigger hand exoskeleton feedback for neuro-rehabilitation," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 10, no. 4, pp. 1070–1080, 2017.

[12] A. Chowdhury, Y. K. Meena, H. Raza, B. Bhushan, A. K. Uttam, N. Pandey, A. A. Hashmi, A. Bajpai, A. Dutta, and G. Prasad, "Active physical practice followed by mental practice using bci-driven hand exoskeleton: a pilot trial for clinical effectiveness and usability," *IEEE journal of biomedical and health informatics*, vol. 22, no. 6, pp. 1786–1795, 2018.

[13] J. d. R. Millán, M. Franzé, J. Mouriño, F. Cincotti, and F. Babiloni, "Relevant eeg features for the classification of spontaneous motor-related tasks," *Biological cybernetics*, vol. 86, no. 2, pp. 89–95, 2002.

[14] I. Iturrate, L. Montesano, and J. Minguez, "Robot reinforcement learning using eeg-based reward signals," in *2010 IEEE International Conference on Robotics and Automation*. IEEE, 2010, pp. 4822–4829.

[15] P. Wang, J. Lu, B. Zhang, and Z. Tang, "A review on transfer learning for brain-computer interface classification," in *2015 5th International Conference on Information Science and Technology (ICIST)*, April 2015, pp. 315–322.

[16] A. M. Azab, L. Mihaylova, K. K. Ang, and M. Arvaneh, "Weighted transfer learning for improving motor imagery-based brain–computer interface," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 7, pp. 1352–1359, 2019.

[17] W. Tu and S. Sun, "A subject transfer framework for eeg classification," *Neurocomputing*, vol. 82, pp. 109 – 116, 2012.

[18] S. Fazli, F. Popescu, M. Danóczy, B. Blankertz, K.-R. Müller, and C. Grozea, "Subject-independent mental state classification in single trials," *Neural Networks*, vol. 22, no. 9, pp. 1305 – 1312, 2009.

[19] C. Vidaurre, M. Kawanabe, P. von Bünau, B. Blankertz, and K. R. Müller, "Toward unsupervised adaptation of lda for brain–computer interfaces," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 3, pp. 587–597, March 2011.

[20] H. Kang, Y. Nam, and S. Choi, "Composite common spatial pattern for subject-to-subject transfer," *IEEE Signal Processing Letters*, vol. 16, no. 8, pp. 683–686, Aug 2009.

[21] W. Samek, F. C. Meinecke, and K. Müller, "Transferring subspaces between subjects in brain–computer interfacing," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 8, pp. 2289–2298, Aug 2013.

[22] B. Reuderink, J. Farquhar, M. Poel, and A. Nijholt, "A subject-independent brain-computer interface based on smoothed, second-order baselining," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug 2011, pp. 4600–4604.

[23] N. Lu, T. Li, X. Ren, and H. Miao, "A deep learning scheme for motor imagery classification based on restricted boltzmann machines," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 6, pp. 566–576, June 2017.

[24] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for eeg decoding and visualization," *Human Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.

[25] Y. R. Tabar and U. Halici, "A novel deep learning approach for classification of EEG motor imagery signals," *Journal of Neural Engineering*, vol. 14, no. 1, p. 016003, nov 2016.

[26] I. Zubarev, R. Zetter, H. Halme, and L. Parkkonen, "Robust and highly adaptable brain-computer interface with convolutional net architecture based on a generative model of neuromagnetic measurements," *CoRR*, vol. abs/1805.10981, 2018.

[27] F. Fahimi, Z. Zhang, W. B. Goh, T.-S. Lee, K. K. Ang, and C. Guan, "Inter-subject transfer learning with an end-to-end deep convolutional neural network for EEG-based BCI," *Journal of Neural Engineering*, vol. 16, no. 2, p. 026007, jan 2019.

[28] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces," *Journal of Neural Engineering*, vol. 15, no. 5, p. 056013, jul 2018.

[29] O. Kwon, M. Lee, C. Guan, and S. Lee, "Subject-independent brain-computer interfaces based on deep convolutional neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–14, 2019.

[30] H. Raza and S. Samothrakis, "Bagging adversarial neural networks for domain adaptation in non-stationary eeg," in *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2019, pp. 1–7.

[31] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, "Filter bank common spatial pattern algorithm on bci competition iv datasets 2a and 2b," *Frontiers in Neuroscience*, vol. 6, p. 39, 2012.

[32] D. Rathee, A. Chowdhury, Y. K. Meena, A. Dutta, S. McDonough, and G. Prasad, "Brain–machine interface-driven post-stroke upper-limb functional recovery correlates with beta-band mediated cortical networks," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 5, pp. 1020–1031, May 2019.

[33] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces," *Journal of neural engineering*, vol. 15, no. 5, p. 056013, 2018.

[34] G. Bouchard, "Efficient bounds for the softmax function and applications to approximate inference in hybrid models," in *NIPS 2007 workshop for approximate Bayesian inference in continuous/hybrid systems*, 2007.

[35] A. Bonner, "Getting started with google colab," 2019.

[36] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, "Tensorflow: A system for large-scale machine learning," in *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, 2016, pp. 265–283.

[37] F. Chollet *et al.*, "keras. github (2015)," 2015.

[38] P. Gaur, R. B. Pachori, H. Wang, and G. Prasad, "A multivariate empirical mode decomposition based filtering for subject independent

bci," in *2016 27th Irish Signals and Systems Conference (ISSC)*, June 2016, pp. 1–7.