varEEGNet : an Improved Neural Network for Motor Imagery based BCI

Gatien Darley
Université Grenoble Alpes
CEA, Leti
38000 Grenoble, France

Stéphane Bonnet
Université Grenoble Alpes
CEA, Leti
38000 Grenoble, France

***Abstract*—EEG-based brain-computer interface (BCI) aims at decoding brain signals into commands. Features used for classification depend on the paradigm and its neurophysiological correlates. EEGNet, a compact convolutional neural network, was proposed to deal both with event-related potentials and motor imagery paradigms. This paper demonstrates that adding a low computational cost variance layer to EEGNet is beneficial for motor imagery as it reduces feature dimension and improves classification results. A mathematical description of EEGNet is also given. Convincing results are shown on a BCI competition dataset with a mean improvement of 2.1% with the addition of the variance layer. Finally, the paper both describes intra- and inter-session results with a first-order statistics transfer learning strategy.**

***Index Terms*—Brain-Computer Interface, EEG, Motor-Imagery, Deep Learning, Transfer Learning**

I. Introduction

Processing electroencephalographic (EEG) signals is a challenging task notably due to the very nature of these signals. One major field of EEG signal processing is Brain Computer Interface (BCI) that aims at translating EEG signals into commands using machine learning tools. BCI application is two-fold: on one side it helps people with motor disorders or affected by neurodegenerative diseases and on the other side, healthy people can benefit from technologies integrating BCIs to enhance performances [1]. In this paper, we focus on Motor Imagery (MI) paradigm. It consists of the mental imagination of limb movements without any muscle activation. Such a task is sufficient to observe brain activation in the same sensorimotor cortex area as if the actual movements were executed.

The particularity of EEG signals is their space-frequency dependency. For a healthy human EEG, the range of frequencies observed is between around 7 and 35 Hz in MI. Those frequencies are subdivided into several groups : alpha (8-12 Hz) and beta (12-30 Hz) are the two main brain rhythms in MI. More precisely, in the motor cortex, the mu rhythm — also known as the sensorimotor rhythm (SMR)— characterizes the EEG signals as being an oscillation observed in the alpha band often synonym of resting or idle state. Therefore, motor behaviors result in a change in the ongoing EEG in form of an event-related desynchronization/synchronization (ERD/ERS) [2]. The spatial aspect is captured through the numerous electrodes placed on the scalp.

EEG analysis is very subject and session-dependent. To cope with subject's variability, research in transfer learning (TL) has been carried out notably through domain adaptation, deep neural network methods and subspace learning [3].

Space-frequency characterization translates usually into two stages : spatial and frequency filtering [4], [5], [6]. Such spatial filters are for instance applied on the frequency filtered EEG data in order to find a projection of the data that has maximal power or variance in one axis and minimal variance in another. Then the log-variance of the signals is often taken as feature for subsequent classification. This processing chain is for instance encountered in the Common Spatial Patterns (CSP) method [5].

Later it has been improved into the Filter Bank Common Spatial Pattern (FBCSP) [6] in order to capture narrow-band frequencies that consider subject-specific modulations. Yet it requires predefined frequency bands. To alleviate this issue Lawhern et al. [7] developed a deep neural network, EEGNet, based on convolutional layers learning temporal and spatial filters to specifically process EEG signals in an end-to-end manner. This approach shows better performances than some State-of-the-Art methods [7], [8]. One advantage is its versatility since it has been used across different EEG-based paradigms without any architecture change: P300 visual-evoked potentials, error-related negativity responses (ERN), movement-related cortical potential (MRCP) or MI. Other deep learning methods have also been proposed — for instance [9] — but we will focus on EEGNet in this study.

This paper proposes two main studies. The first one aims at improving the original EEGNet proposed in [7] in an efficient way, by using as feature for classification the log-variance of the filtered signals, inspired by the CSP method. We will show the contribution of adding a log-variance layer with the EEGNet and how this modification improves the network in both intra- and inter-session evaluation framework. Through the second one, we will perform a basic transfer learning (TL) procedure to improve the performances when testing on another session.

II. Methods

This section gives first a mathematical description of EEG-Net to better understand the key operations underlying the network architecture. The new variance layer is then introduced.

TABLE I
EEGNET PARAMETERS FOR TWO DIFFERENT EEGNET-L_1, D ARCHITECTURES.

	K_1	L_1	N_1	K_2	N_2	N_p
EEGNET-4,2	8	4	64	8	16	1,660
EEGNET-8,2	16	8	64	16	16	3,444

Table I gives the values of the parameters of two EEGNet architectures.

A. EEGNet

The EEGNet is a convolutional neural network (CNN) tailored for EEG-based BCIs and developed by Vernon Lawhern in 2018 [7]. The input of EEGNet is a mean-centered EEG epoch $\mathbf{X}$ with size $N_C \times N_T$ where N_C is the number of channels and N_T the number of time samples.

1) The first processing block regroups both the temporal and depthwise convolutions.

$$\mathbf{Z}_1^{(i)} = \mathbf{W}_1^{(i)} \left[\mathbf{X} \star_2 \mathbf{h}_1^{(i)} \right] \tag{1}$$

Because the temporal filters $\mathbf{h}_1^{(i)}$ are 1-d, (1) is the 2D-convolution ($\star_2$) of each row of the epoch $\mathbf{X}$ with the *same* filter of size $1 \times N_1$. The dimension of $\mathbf{W}_1^{(i)}$ is $D \times N_C$. This spatio-temporal filtering is repeated for L_1 frequency bands. The output of the first block is the epoch $\mathbf{Z}_1$ with size $K_1 \times N_T$ with $K_1 = D \cdot L_1$ the number of feature maps produced by the first block.

2) The second block contains the separable convolution and implements

$$\mathbf{Z}_2 = \mathbf{W}_2 \left[\mathbf{Z}_1(\downarrow 4) \star_{2,s} \mathbf{H}_2 \right] \tag{2}$$

The operator $\star_{2,s}$ indicates the 'separable' convolution where each row of $\mathbf{Z}_1$ is filtered with a different filter. The dimension of $\mathbf{H}_2$ is $K_1 \times N_2$ while the dimension of $\mathbf{W}_2$ is $K_2 \times K_1$. The output of the second block is the epoch $\mathbf{Z}_2$ with size $K_2 \times (N_T/4)$.

3) The last step of EEGNet is a standard fully connected layer followed by a softmax that generates the probability vector

$$\mathbf{z}_3 = \text{softmax} \left[\mathbf{W}_3 \, \text{vec} \left[\mathbf{Z}_2(\downarrow 8) \right] \right] \tag{3}$$

The dimension of $\mathbf{W}_3$ is $C \times K_2(N_T/32)$ with C the number of MI tasks (here $C = 4$).

The number of learned parameters is easily computed as :

$$N_p = L_1 N_1 + K_1(N_C + N_2 + K_2) + C[1 + K_2(N_T/32)]$$

The EEGNet also contains batch normalization layers after the temporal, the depthwise and the separable convolutions which adds respectively $2(L_1 + K_1 + K_2)$ parameters to the EEGNet model.

B. varEEGNet

We propose to add a new layer that computes the logarithm of the variance of each spatially-temporally filtered signal $\mathbf{Z}_2$. The last layer generates the probability vector

$$\mathbf{z}_3 = \text{softmax} \left[\mathbf{W}_3 \log \left(\mathbb{V} \left\{ \mathbf{Z}_2 \right\} \right) \right] \tag{4}$$

where $\mathbb{V}\{\mathbf{Z}_2\}$ returns a $K_2 \times 1$ column vector that contains the variance of each row in $\mathbf{Z}_2$. The dimension of $\mathbf{W}_3$ is now $C \times K_2$ instead of $C \times K_2(N_T/32)$.
As it appears in the CSP method, taking log-variance as feature for MI classification turned out to be relevant. Indeed, the log-variance allows to capture efficiently the power variations in different frequency bands. This is ideal for detecting neuronal rhythms. Especially for EEG signal processing where the signals present many variations and highly vary between subjects and sessions. Applying the logarithm amplifies relative differences in the data.

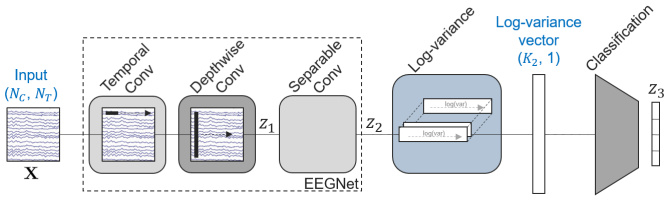


Fig. 1. varEEGnet : our proposed improvement of the EEGNet by adding a log-variance layer. The log-variance layer computes the logarithm of the variance along the temporal dimension of the feature maps resulting from the EEGNet.

One notes that performing this operation along the temporal dimension squeezes it, reducing at the same time the dimension. The EEGNet produces a feature vector of length $K_2 \cdot N_T/32$ while our method varEEGNet outputs a feature vector of length K_2.

C. Transfer learning

In order to improve inter-session evaluation results, we rely on an efficient transfer learning technique. The method involves geometrically whitening the input data from both sessions so that they present similar distributions to first order. Let $\mathbf{X}_n$ be the n-th zero-mean trial of one session of size $N_C \times N_T$. We first compute the sample covariance (SCM) matrix for each trial : $\mathbf{C}_n = \frac{1}{N_T - 1} \mathbf{X}_n \mathbf{X}_n^T$. The geometric mean $\mathfrak{G}$ of a session is given by the Fréchet mean of all the covariance matrices from the different trials in it: $\mathfrak{G}(\mathbf{C}_1, \ldots, \mathbf{C}_N) = \text{argmin}_{\mathbf{C} \in \mathcal{P}(N_c)} \sum_{n=1}^{N} \delta_R^2(\mathbf{C}, \mathbf{C}_n)$ where δ_R is the Riemannian geodesic distance and $\mathcal{P}(N_c)$ the set of $N_C \times N_C$ symmetric positive definite (SPD) matrices. It is computed in an iterative manner as in [4].
The idea behind data geometric whitening is to transform them such that the transformed geometric mean of each session becomes identity. So we are looking for a matrix $\mathbf{W}$ such that $\mathbf{W}\mathbf{G}_{session}\mathbf{W}^T = \mathbf{I}_{N_C}$ where $\mathbf{G}_{session}$ is the geometric mean of a session and $\mathbf{I}_{N_C}$ denotes the identity matrix of size N_C. A possible choice is obviously $\mathbf{W} = \mathbf{G}_{session}^{-1/2}$.

TABLE II

EEGNET-8,2 ACCURACY RESULTS (IN %) FOR EACH SUBJECT IN INTRA- AND INTER-SESSION. THE AVERAGE ACCURACY IS CALCULATED OVER ALL SUBJECTS AND WHEN REMOVING THE LESS RESPONSIVE ONES (S2, S5, S6). THIS TABLE SHOWS THE IMPACT OF THE LOG-VARIANCE LAYER.

Evaluation framework	log-var layer	S1	S2	S3	S4	S5	S6	S7	S8	S9	AVG (±SD)	AVG (±SD) best subjects
intra-session	x	**72.9**	50.0	**83.4**	**50.9**	**43.8**	43.4	**73.3**	**82.6**	63.6	**62.7 (±16.1)**	**71.1 (±12.3)**
intra-session		70.0	**50.7**	83	48.2	31.9	42.7	63.2	77.7	**78.4**	60.6 (±18.0)	70.1 (±12.8)
inter-session	x	72.7	45.5	79.9	50.4	38.5	**45.7**	58.1	68.9	65.1	58.3 (±14.2)	65.9 (±10.5)
inter-session		71.1	38.3	82.9	46.4	31.8	39.1	56.1	68.0	68.6	55.8 (±17.8)	65.5 (±12.7)

TABLE III

EEGNET-4,2 ACCURACY RESULTS (IN %) IN THE SAME FRAMEWORK AS TABLE II.

Evaluation framework	log-var layer	S1	S2	S3	S4	S5	S6	S7	S8	S9	AVG (±SD)	AVG (±SD) best subjects
intra-session	x	**74.3**	**50.0**	82.0	42.7	36.5	**44.8**	**69.5**	**75.4**	62.2	**59.7 (±16.6)**	**67.7 (±13.9)**
intra-session		72.6	41.2	**82.6**	42.3	35.1	40.3	56.2	69.4	**77.7**	57.5 (±18.4)	66.8 (±15.0)
inter-session	x	69.3	43.4	79.2	**46.9**	**38.4**	42.6	59.8	65.4	66.1	56.8 (±14.3)	64.5 (±10.7)
inter-session		68.4	41.8	80.0	42.7	31.9	39.0	53.4	66.5	67.9	54.6 (±16.7)	63.2 (±13.1)

TABLE IV

VAREEGNET-8,2 ACCURACY RESULTS (IN %) IN THE SAME FRAMEWORK AS TABLE II. THIS TABLE EMPHASIZES THE SIGNIFICANCE OF TRANSFER LEARNING WITH EEG INTER-SESSION FRAMEWORK.

Evaluation framework	Whitening data	S1	S2	S3	S4	S5	S6	S7	S8	S9	AVG (±SD)	AVG (±SD) best subjects
inter-session		**72.7**	45.5	79.9	50.4	38.5	45.7	58.1	68.9	65.1	58.3 (±14.2)	65.9 (±10.5)
inter-session	x	71.7	**46.4**	**81.1**	**51.7**	**39.7**	**45.6**	**69.0**	**75.0**	**66.1**	**60.7 (±15.0)**	**69.1 (±10.0)**

Finally the whitened transformed data is : $\widetilde{\mathbf{X}}_n = \mathbf{G}_{session}^{-\frac{1}{2}}\mathbf{X}_n$. The goal is to apply geometric whitening on both sessions in order to take care of the existing covariance shift between them. Algorithm 1 describes this procedure applied for both sessions.

Algorithm 1 Data geometric whitening for inter-session transfer learning

Input: $\mathbf{X}_n^{tr}$, $\mathbf{X}_n^{te}$ the n-th trials of the training, testing sessions.
Output: $\widetilde{\mathbf{X}}_n^{tr}$ the whitened data of the training sessions.
Output: $\widetilde{\mathbf{X}}_n^{te}$ the whitened data of the testing session.
Training :
Compute $\mathbf{C}_n^{tr}$ the SCM of $\mathbf{X}_n^{tr}$.
Compute $\mathbf{G}_{tr} = \mathfrak{G}(\mathbf{C}_1^{tr}, \ldots, \mathbf{C}_N^{tr})$
return $\widetilde{\mathbf{X}}_n^{tr} = \mathbf{G}_{tr}^{-1/2}\mathbf{X}_n^{tr}$ [Transformed training data]
Train EEGNet model with the whitened data.
Testing :
Compute $\mathbf{C}_n^{te}$ the SCM of $\mathbf{X}_n^{te}$.
Compute $\mathbf{G}_{te} = \mathfrak{G}(\mathbf{C}_1^{te}, \ldots, \mathbf{C}_N^{te})$
return $\widetilde{\mathbf{X}}_n^{te} = \mathbf{G}_{te}^{-1/2}\mathbf{X}_n^{te}$ [Transformed testing data]
Test EEGNet model in inference with the whitened data

III. RESULTS

A. Dataset

The EEG data we will be using for this study come from the public dataset BCI Competition IV 2A that can be found on MOABB (https://moabb.neurotechx.com/docs/api.html). The signals were recorded using 22 electrodes, sampled at 250 Hz. This dataset comprises 2 sessions where the electrodes remain fixed on the subject's scalp. Each session corresponds to a series of 6 runs separated by short breaks. 9 subjects were asked to perform motor imagery tasks which consists of 4 classes of movements : left hand, right hand, feet and tongue; following a specific protocol [8]. One refers to an epoch or a trial as a segment within a run triggered by an external stimulus (the cue) corresponding to one mental MI task. In total there are 72 trials per class (12 per run), yielding a total of 288 trials per subject. We will both evaluate intra- and inter-session performances. Intra-session evaluation is performed on the first session for each subject by dividing the trials into a training set and a test set using five stratified K-fold splits. The final classification score is obtained by averaging over the five splits. For inter-session evaluation, for each subject, the model is trained on the first session and tested on the second session. We used 5-fold cross-validation. As mentioned in the literature [8], this reduces the risk of overfitting.

B. Preprocessing

The EEG signals are bandpass-filtered between 8 and 32 Hz, and each trial has a length of four seconds starting from the cue (1,000 samples). The preprocessing steps and the model parameters are chosen in order to stick to the framework of MOABB's study which goal is to offer a proper and efficient mean of comparison between existing algorithms [8]. The data are centered and have zero mean.

C. Performances

Table II and III regroup the performances in terms of accuracy in different evaluation frameworks. It emphasizes the impact of adding the log-variance layer for two EEGNet architectures : EEGNet-8,2 and EEGNet-4,2. First, we find consistent results with the literature [8], that is around 60%

average accuracy in intra-session for the EEGNet-8,2 model. Then it was found in [7] that on BCI Competition IV 2A dataset, EEGNet-8,2 performs better than EEGNet-4,2, that is also a result we observe through the tables obtained with the mentioned preprocessing.

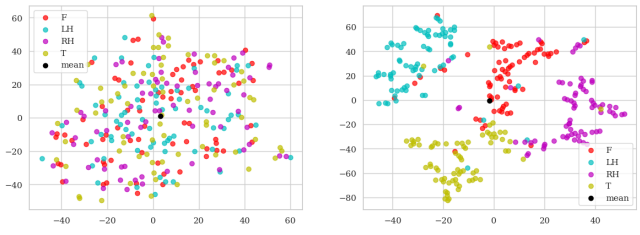
The main results are about the varEEGNet model. Adding the log-variance layer slightly improves the performances with a 2.1% (resp. 2.2%) gain in intra-session and 2.5% (resp. 2.2%) in inter-session for EEGNet-8,2 (resp. for EEGNet-4,2). We note that these dynamics also preserved when averaging on the most responsive subjects. Also, as expected, the standard deviation of the subjects is reduced when computing the log-variance. Averaging the subject scores allows to observe that gain, however, because of subject variability, this is not the case for some "*bad*" subjects. For subject S9 computing the log-variance deteriorates the performances. Another benefit of the proposed method is the dimension reduction of the features. Table V presents the number of parameters learned by the different models. We verify that the log-variance layer reduces that number by 57.8% using a EEGNet-4,2 architecture and by 55.7% based on EEGNet-8,2.

TABLE V
NUMBER OF PARAMETERS LEARNED FOR THE DIFFERENT EEGNET ARCHITECTURES. VAREEGNET REFERS TO THE EEGNET MODEL WITH THE ADDITION OF THE LOG-VARIANCE LAYER.

Model	# learned parameters
EEGnet-4,2	1,660
varEGnet-4,2	**700**
EEGnet-8,2	3,444
varEEGnet-8,2	**1,524**

We decided to project the features of both the proposed model and the original EEGNet on a 2-dimensional space using t-distributed stochastic neighbor embedding (t-SNE) algorithm [10] in order to show the benefit of using the log-variance as feature for classification. The algorithm was applied to the input feature vector of the fully connected layer. Fig.2 illustrates this for subject S8. Indeed, Fig.2(b) — displaying the features that we propose after the log-variance layer — presents well separated classes which is propitious for classification, while Fig.2(a) represents superposed classes. So at first sight, this explains why varEEGNet performs better. However, t-SNE representation may not be adapted for representing too high dimensional features [10], that might explain why classes distributions on Fig.2(a) look overlaid. Still, for a same t-SNE algorithm, varEEGNet has better features distribution, most likely due to its low feature dimension.

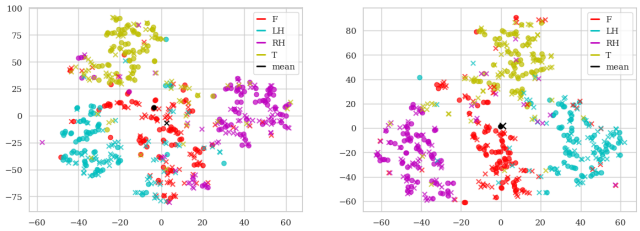
We clearly see from Tables II and III that the results in inter-session are lower than in intra-session. So we conducted further studies by setting up geometric whitening on the dataset in order to improve the inter-session results for our proposed network. The effects of the whitening transformation presented in section II are visible on Fig.3. We observe that the whitened data of the two sessions represented by discs and crosses in Fig.3(b) have now the same overall mean and present similar



(a) EEGnet-8,2 features (b) varEEGNet-8,2 features eq.(4)

Fig. 2. 2-D representation of the models features for subject S8 on training session. F, LH, RH and T are the abbreviations for each class.

distributions. As mentioned, the results in inter-session without whitening are weaker than in intra-session because without any adaptation strategy, the data of the two sessions recorded for one subject present different distributions, especially different means Fig.3(a). With our whitening procedure described in algorithm1 we partially solved covariance shift issues [11]. This translates in an improvement of the performances, observable in Table IV. For all subjects except S1, accuracy scores are better when whitening the data in inter-session framework: that is $+2.4\%$ in average and $+3.2\%$ when averaging on the most responsive subjects. In this inter-session whitening framework we were also able to verify a slight improvement brought by the addition of a log-variance layer.



(a) Original data (b) Whitened data

Fig. 3. 2-D representation of the varEEGNet-8,2 features for subject S8 on both sessions with or without whitening. F, LH, RH and T are the abbreviations for each class. The discs represent the data from the training session, the crosses the data from the testing session.

D. Interpretability

In order to verify that the proposed method is not driven by noise or artifacts in the data, we extract the learned parameters and analyse them. We use two approaches to characterize the features computed by our model : (1) visualising the temporal kernel, (2) extracting spatial patterns. Temporal kernels acts as frequency filters so plotting their Fourier transformation gives us an idea of which frequencies are filtered. Fig.4 represents a frequency filter for subject S8 learned by the temporal convolution. Spatial patterns are directly extracted from spatial filters learned by the depthwise convolution according to Haufe's method [12]. Fig.5 presents two typical patterns.

The selected temporal filter extracts frequencies located rather in the alpha band represented in purple on Fig.4. This information is consistent when practicing MI [2]. The kernel

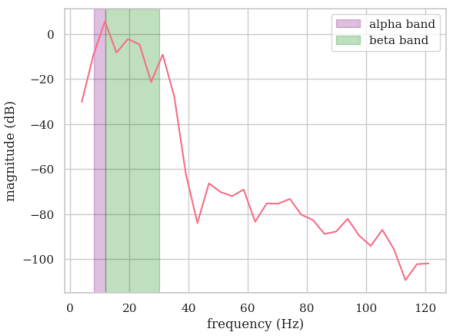


Fig. 4. Frequency filter extracted from the temporal convolution of varEEGNet-8,2 trained on subject S8, with random initialization.

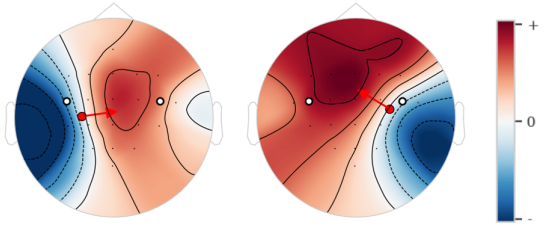


Fig. 5. Two spatial patterns extracted from varEEGNet-8,2 trained on S8. The virtual dipole approximation locates the brain activity near the electrodes for left-right hand MI. They do not correspond to physiological brain dipoles.

size conditions the number of points for the Fourier transform. Spatial patterns refer to the topographical distribution of the signal across the sensor array, which can be revealed by the spatial filters. Each of them is a component of the brain activity representation. There are 2 spatial patterns per frequency band that is 16 for varEEGNet-8,2, we displayed two of them highlighting interesting appearance. Indeed, the patterns representation Fig.5 is similar to the electrical activity of a dipole which would have been placed at the surface of the brain at the motor cortex area which is active when performing left and right hand MI. So this is characteristic of the left-right hand motor imagery [5]. It means that these two patterns better characterizes left hand (right pattern) and right hand (left pattern) movements among the four MI tasks after training.

IV. Discussion

The two main areas of focus addressed in this paper show great results and are easy to handle and to implement. In this section we will discuss some results.

Firstly, as mentioned, with EEG signals there is an important variability between subjects and sessions. Some subjects are naturally less responsive than others when performing the protocol. This is why sometimes some accuracy values could seem low. That is also one of the reasons why we computed the mean removing less responsive subjects in Table II and III, because their data can be difficult to interpret.

Moreover, something to keep in mind when working on AI-based method is reproducibility issues. The fact that each learned model is different will produce slightly different evaluation results. Yet the trends and orders of magnitude should stay the same.

It is also of note that for the study of interpretability not all filters learn relevant information. Indeed among the frequency filters extracted some of them do not present the characteristic peak in the alpha band. So a further study would be to analyse which convolution kernels contribute the most to get great accuracy results. We can draw a similar reasoning on spatial filters.

Finally, while the intra-session score of 62.7% from Table II is not reached with whitening in the framework of Table IV it might be because the transformations applied on the two sessions are basic. Theoretically we could work more on the two sessions to match their distributions better, for example by performing scaling or rotation operations [13].

V. Conclusion

This article presents a low computational cost method in order to improve EEGNet performances for MI. This is done by adding a log-variance layer downstream of the EEGNet model to extract proper features for classification. This does not only improve performances but also reduces the features dimension and at the same time the number of parameters learned by the network. What has been addressed here, could potentially be applied as well on other convolutional neural networks in EEG MI, for instance [9]. Furthermore, to cope with inter-session evaluation framework, we used transfer learning and implemented a whitening operation on data which increased classification accuracy in this scenario.

References

[1] J. van Erp *et al.*, "Brain-computer interfaces: Beyond medical applications," *Computer*, vol. 45, no. 4, pp. 26–34, 2012.

[2] G. Pfurtscheller, K. Pichler-Zalaudek, and C. Neuper, *ERD and ERS in voluntary movement of different limbs*, pp. 245–268. Netherlands: Elsevier B.V., revised edition vol. 6 ed., 1999.

[3] Z. Wan *et al.*, "A review on transfer learning in EEG signal analysis," *Neurocomputing*, vol. 421, pp. 1–14, 2021.

[4] A. Barachant *et al.*, "Multiclass braincomputer interface classification by Riemannian geometry.," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 4, pp. 920–928, 2012.

[5] Z. J. Koles *et al.*, "Spatial patterns underlying population, differences in the background EEG," *Brain Topography*, vol. 2, 1990.

[6] K. K. Ang *et al.*, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pp. 2390–2397, 2008.

[7] V. J. Lawhern *et al.*, "EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces," *Journal of Neural Engineering*, vol. 15, no. 5, p. 056013, 2018.

[8] S. Chevallier *et al.*, "The largest EEG-based BCI reproducibility study for open science: the MOABB benchmark." working paper or preprint, Apr. 2024.

[9] R. Schirrmeister *et al.*, "Deep learning with convolutional neural networks for decoding and visualization of EEG pathology," in *2017 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, pp. 1–7, 2017.

[10] L. van der Maaten and G. Hinton, "Visualizing high-dimensional data using t-SNE," *Journal of Machine Learning Research*, vol. 9, no. nov, pp. 2579–2605, 2008. Pagination: 27.

[11] H. Shimodaira, "Improving predictive inference under covariate shift by weighting the log-likelihood function," *Journal of Statistical Planning and Inference*, vol. 90, pp. 227–244, 10 2000.

[12] S. Haufe *et al.*, "On the interpretation of weight vectors of linear models in multivariate neuroimaging," *NeuroImage*, vol. 87, pp. 96–110, 2013.

[13] P. L. C. Rodrigues, C. Jutten, and M. Congedo, "Riemannian procrustes analysis: Transfer learning for brain–computer interfaces," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 8, pp. 2390–2401, 2019.