Sinc-EEGNet for Improving Performance While Reducing Calibration of a Motor Imagery-Based BCI

Sinc-EEGNet for improving performance while reducing calibration of a motor imagery-based BCI

Pasquale Arpaia*†‡, Elisa Bertone*, Antonio Esposito*§, Angela Natalizio*¶,
Marco Parvis¶, Alessandra Laura Giulia Pedrocchi‖, Andrea Pollastro*

*Augmented Reality for Health Monitoring Laboratory (ARHeMLab),
Università degli Studi di Napoli Federico II, Naples, Italy.
†Department of Electrical Engineering and Information Technology (DIETI),
Università degli Studi di Napoli Federico II, Naples, Italy.
‡Centro Interdipartimentale di Ricerca in Management Sanitario e Innovazione in Sanità (CIRMIS),
Università degli Studi di Napoli Federico II, Naples, Italy.
§Department of Engineering for Innovation, University of Salento, Lecce, Italy.
¶Department of Electronics and Telecommunications (DET), Polytechnic of Turin, Turin, Italy.
‖NearLab, Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy.

***Abstract*—The potential of motor imagery-based brain-computer interfaces (BCIs) is hindered by long calibration times. Therefore, this study investigates a classification model that minimises BCI calibration time while maximising its accuracy by exploiting transfer learning. To this end, a modified version of the Sinc-EEGNet architecture is proposed. Analyses were carried out with data from multiple subjects. Notably, when the model was trained with data from subjects other than the test subject, Sinc-EEGNet-32 achieved a mean classification accuracy of $78 \pm 10\,\%$. This outperformed the reference EEGNet-4 architecture by $10\,\%$. Instead, when considering also data from the test subject for a fine tuning, Sinc-EEGNet-32 achieved a mean accuracy of $80 \pm 10\,\%$ by exploiting only $10\,\%$ of test subject's data and $83 \pm 10\,\%$ by exploiting $40\,\%$ of test subject's data. These correspond to a system calibration of less than $2.0\,\mathrm{min}$ and of approximately $8.0\,\mathrm{min}$, respectively. Overall, there was an increasing trend in performance for Sinc-EEGNet-32 as higher percentages of data were exploited for fine-tuning. In contrast, EEGNet-4 only achieved an accuracy of $72 \pm 13\,\%$ even with fine tuning.**

***Index Terms*—brain-computer interface, motor imagery, electroencephalography, Sinc-EEGNet, transfer learning.**

I. Introduction

A motor imagery-based brain–computer interface (BCI) relies on measuring brain signals during the mental execution of a movement. It enables a direct interaction between the human brain and external devices without peripheral (muscular) activity [1]. Such an interface demonstrated great potential as an assistive technology or in rehabilitation [2].

However, motor imagery-based BCIs still face limitations, including a low signal-to-noise ratio, strong dependency on specific tasks, and the non-stationarity. Specifically, the non-stationarity introduces intra-subject and inter-subject variability [3], [4]. Training artificial intelligence algorithms faces domain shift issues. Such an issue is associated with mismatching data distributions for training and test sets, where data from prior sessions of the subject or from other subjects would be adopted [5]. As a result, motor imagery-based BCIs often require recalibration from scratch for each new subject or session of the same subject. This poses a significant limitation for widespread usage in clinical and daily-life applications.

To address the calibration challenge, several research studies explored transfer learning techniques to cope with low data availability [6]. Many authors focused on traditional machine learning approaches, mostly relying on filter-bank common spatial pattern [7], [8]. These studies employed transfer learning to enhance the invariance of features extracted from multiple pre-recorded sessions [9], [10]. However, these approaches are constrained by the a priori extraction and selection of handcrafted features. In contrast, deep learning methods offer the advantage of processing raw data directly, with minimal or no pre-processing. However, the need to train a huge number of parameters and the low data availability typical of the context of interest could lead to ineffective learning [11]. Consequently, transfer learning emerged as a valuable resource for extending the success of deep learning methods to motor imagery-based applications.

In [12], an adaptive layer on the top of a filter-bank common spatial pattern was introduced into a fully connected deep network. The pre-trained model on previous sessions of the test subject was found to be effective. However, a transfer learning across subjects was not explored. In [13], the authors achieved relevant performance improvements by applying transfer learning to multiple deep learning models. When adopting an online modality, the convenience of a transfer involving different sessions of the same subject rather than multiple subjects was highlighted. More recent studies also explored the subject-to-subject transfer. In [14], a "golden" subject's data were used to train a standard deep network, followed by the design of an encoder-decoder network to align individual test subjects' data with that of the golden subject. However, the subject-specific network required extensive and time-consuming training.

In this framework, a hybrid deep neural network named Sinc-EEGNet was recently proposed for classifying motor imagery [15]–[17]. This results from the merging of two reliable and already successful architectures: the EEGNet [18] and the SincNet [19]. Different ways for combining the two architectures were proposed. In [15], [17], an attention module was added to perform automatic feature selection. In [16], modifications like the addition of extra dropout layers were introduced. However, the main goal of these works was to maximise the suitability of the models for learning from single-subject data, namely the focus was on intra-subject transfer.

Overall, while previous research tried to address the limitations of motor imagery-based BCIs through transfer learning, challenges persist in achieving efficient calibration and generalisation across subjects. The present study aims to further explore these challenges and propose a novel approach to enhance the usability and effectiveness of motor imagery-based BCIs.

To this end, a modified version of Sinc-EEGNet with a strong generalisation capability across different BCI subjects was designed. Then, an inter-subjective transfer was investigated to reduce calibration time for adaptation to a new subject, hence compliant with a real-time utilisation.

The remainder of the paper is organised as follows. Section II recalls the already existing EEGNet and SincNet architectures. Section III describes the proposed hybrid architecture, its implementation, and the data analysis approach. Sections IV presents the exploited dataset and discusses inherent results.

II. Background

This section presents two already existing architectures, namely EEGNet and SincNet, before presenting the proposed hybrid architecture.

A. EEGNet

EEGNet [18] was developed to extract various relevant features from EEG signals in a robust manner. It is a low-depth CNN that integrates two peculiar types of convolutions, namely depth-wise convolution and separable convolution.

EEGNet first adopts a standard convolutional layer to learn traditional temporal filters. It then trains frequency-specific spatial filters by means of depth-wise convolution. Each filtered version of the input EEG is individually optimised by kernels that explore the spatial correlations among the EEG channels. Following this, a separable convolution is applied. It combines a depth-wise convolution for capturing temporal summaries of each feature map and a point-wise convolution for optimal mixing of the summarised feature maps. Interestingly, EEGNet architecture resembles the main steps of the well-known filter-bank common spatial pattern, while adding to it a higher flexibility thanks to the end-to-end training procedure peculiar to deep learning models.

Currently, it is adopted as the golden standard by most studies that propose new EEG decoding approaches [20], [21]. Notably, almost half of the total amount of EEGNet trainable parameters belongs to the first standard convolutional layer. This layer is characterised by a parameter count of $N*L$, with N the number of embedded kernels and L the filters length.

B. SincNet

SincNet [19] was originally proposed for audio signals discrimination tasks. It aimed to discover more meaningful initial filters than the traditional convolutional ones. Indeed, the authors claimed that the resulting first layer kernels usually take very noisy and incongruous multi-band frequency responses. Hence, a sinc-layer was proposed as an alternative to a traditional convolutional one. Specifically, instead of the traditionally learned finite impulse response filter, a peculiar trainable kernel was adopted. It is constrained to match sinc function in the time domain or, equivalently, band-pass filter in the frequency domain.

This approach offers advantages in terms of interpretability and efficiency, as the trainable weights of a sinc-layer are only twice the number of embedded kernels.

In the context of motor imagery decoding and the present research study, the sinc-layer proves highly suitable. Since the number of its tunable parameters does not depend on the filters' length, an elevated frequency selectivity in EEG filtering can be efficiently achieved. In contrast, achieving a similar purpose with a traditional convolutional layer would require a large number of weights to balance time and frequency resolution. The compactness of sinc-filters plays a crucial role in minimising calibration requirements before BCI utilisation.

III. Materials and methods

A. The architecture

A modified merge of SincNet and EEGNet architectures was designed to enhance the advantages of both networks specifically for inter-subjective application. By incorporating a customised bandpass filter bank, the hierarchical exploration capability of EEGNet was improved. This allowed EEGNet to benefit from the advantages offered by the sinc-layer, enabling effective feature extraction from raw waveforms and maximising filter suitability for the target application.

In contrast to previously proposed Sinc-EEGNet models [15]–[17], a strong similarity with the original version of EEGNet was here voluntarily preserved to fully exploit its intrinsic generalisation capability. As reported in Table I, the proposed Sinc-EEGNet faithfully reproduced EEGNet architecture except for the first block. Specifically:

1) the first traditional convolutional layer was replaced by a sinc-layer;
2) spatial filters were individually trained for each frequency range resulting from the simultaneous optimisation of the sinc-layer;
3) the resulting feature maps were first temporally resumed on an individual basis and then mixed;
4) finally, a fully connected layer handles the classification task.

Similarly to EEGNet, Sinc-EEGNet compensates for the lack of flexibility of filter-bank common spatial pattern due

to the end-to-end training procedure. This allows for multi-resolution EEG analysis by tuning the bandwidths. Additionally, in the proposed model, the preservation of the traditional form of the bandpass filters ensures closer adherence to the gold standard of machine learning approaches. Hence, the initial filtering quality and human readability of the latter are preserved.

B. Implementation

Both Sinc-EEGNet and EEGNet were implelemted in this study. The inclusion of EEGNet was motivated by the need for a reference model for comparative analysis.

1) EEGNet implementation: Referring to [18] as starting point, firstly EEGNet was implemented. To guarantee a high architectural coherence with this study, network hyperparameters (such as temporal kernels lengths and pooling layers dimensions) were re-expressed in relative terms with respect to f_s and the same proportionality relation was kept. If not specified, the number of temporal and spatial filters, as well as dropout percentage hyperparameters settings were maintained identical.

EEG signals were band-pass pre-filtered in $4\,\mathrm{Hz}$ to $40\,\mathrm{Hz}$ to ensure a faithful reproduction. Then, each EEG trial was epoched such to include the cue and the motor imagery window of a typical synchronous motor imagery trial. Finally, a classical standardization procedure was employed.

2) Sinc-EEGNet implementation: As with EEGNet, the Sinc-EEGNet hyperparameters were expressed by maintaining proportionality with f_s. In [18] the length of the first standard layer was set equal to $\frac{f_s}{2}$. Here, the length of the sinc-filters was set to $\frac{f_s}{8}$ as this yielded better performance in preliminary experiments (Tab. I). For a similar empirical reasoning, in the third block the separable convolution kernel size was modified from $\frac{f_s}{8}$ to $\frac{f_s}{16}$. In order to maximise the potential of the proposed network to automatically learn from raw EEG signals, no prior pre-filtering was applied. As for EEGNet, each EEG trial was epoched to include the cue and the motor imagery window. Finally, a peculiar standardization strategy was adopted to mitigate the aforementioned domain shift issue. Among the effective standardization strategies proposed in [5], the best one feasible with on-line applicability was selected. Specifically, each subject belonging to the training set was individually transformed using his/her own mean and variance, while the test subject was standardized adopting the statistics computed on the whole training data.

C. Data analyses

Two main analyses were carried out in this study. In the first analysis, the models were evaluated by iteratively training them on all subjects of the selected dataset except one. The excluded subject was then used as the test subject to assess the model's performance. In the second analysis, a fine-tuning was performed on the excluded (test) subject.

1) Inter-subject training: The initial analysis involved comparing four models using the inter-subject training modality: EEGNet-4, EEGNet-32, Sinc-EEGNet-4, and Sinc-EEGNet-32. The number following the name of each architecture corresponds to the number of temporal filters in its first layer. EEGNet-4 refers to the original architecture proposed by the authors [18], while EEGNet-32 was introduced to examine the impact of increasing the number of filters. The Sinc-EEGNet architecture was then inspected in both configurations (i.e., with 4 and 32 filters).

Once the dataset was selected, the condition of a new subject was simulated by training the models using only data from other subjects. To ensure robust classification performance and avoid subject bias, the leave-one-subject-out cross-validation method was adopted, as suggested by [22]. This method involves dividing the data into k folds, with k equal to the number of subjects in the dataset. In each iteration of cross-validation, a different subject served as the test set. For implementing it, *GroupKFold* function from *sklearn.model_selection* was employed. Following the approach in [23], the remaining subjects' data were further split in a 9:1 ratio. The first set was used for training the network weights, while the second set served as validation data for the early stopping procedure.

2) Fine-tuning adaptation: A second analysis was carried out to explore transfer learning application.

For this purpose, the pre-trained model with data from several subjects was fine-tuned on a new subject. The classification performance on this new subject was evaluated by varying his/her percentage of the data used for the fitting process. At each leave one subject out cross-validation iteration, the data of the relative test subject were chronologically split as follows: the second half was chosen as fixed and unique test set, hence it was never exploited for model subject-specific adaptation; from the first half, subsets of different sizes were extracted for applying the fine-tuning procedure to the model pre-trained on the other subjects' data. Specifically, separate tests were performed by progressively enlarging such subsets, thus exploring relative sizes of: $10\,\%$, $20\,\%$, $30\,\%$ and $40\,\%$. Note that these percentages were computed by considering as $100\,\%$ the total amount of data in the whole test subject's session (i.e. including both the first and the second half). Each subset was further split with stratification at ratio 9:1 to obtain, respectively, the fine-tuning training and validation data. To guarantee a greater robustness of the results obtained from the detailed procedure, 15 experiments were conducted per test subject and per individual percentage by repeating the random sampling process of the fine-tuning data from the first half of the session.

3) Experiments setup: For the sake of reproducibility, all the experiments were performed with random seed fixed to 42. Binary cross-entropy was minimised on mini-batches of 32 trials and on a maximum of 10000 epochs. For this purpose, Adam optimiser [24] was employed in conjunction with an early stopping method, respectively fixing to 0.02 the former weight decay and to 150 the latter epochs limit. A different learning rate was chosen for pre-training and fine-tuning processes, respectively 0.0001 and 0.00001 as this

TABLE I: Proposed Sinc-EEGNet architecture. C = number of channels, T = number of time samples, F_1 = number of sinc-filters, D = number of spatial filters, f_s = sampling frequency.

Block	Layer	# filters	Size	# tunable parameters	Output	Activation
	Input				(C,T)	
	Reshape				(1,C,T)	
1	Sinc-layer	F_1	(1, $\frac{fs}{8}$)	2*F_1	(F_1,C,T)	Linear
	Batch Normalization			2 * F_1	(F_1,C,T)	
2	Depthwise convolution	D*F_1	(C,1)	C*D*F_1	(D*F_1,1,T)	Linear
	Batch Normalization			2*D*F_1	(D*F_1,1,T)	
	Activation				(D*F_1,1,T)	ELU
	Average Pooling		(1,$\frac{fs}{32}$)		(D*F_1,1,$\frac{T}{\frac{fs}{32}}$)	
	Dropout				(D*F_1,1,$\frac{T}{\frac{fs}{32}}$)	
3	Separable convolution	D*F_1	(1,$\frac{fs}{16}$)	D*F_1*($\frac{fs}{16}$+D*F_1)	(D*F_1,1,$\frac{T}{\frac{fs}{32}}$)	Linear
	Batch Normalization			2*D*F_1	(D*F_1,1,$\frac{T}{\frac{fs}{32}}$)	
	Activation				(D*F_1,1,$\frac{T}{\frac{fs}{32}}$)	ELU
	Average Pooling		(1,$\frac{fs}{16}$)		(D*F_1,1,$\frac{T}{\frac{fs}{32}*\frac{fs}{16}}$)	
	Dropout				(D*F_1,1,$\frac{T}{\frac{fs}{32}*\frac{fs}{16}}$)	
	Flatten				(D*F_1*$\frac{T}{\frac{fs}{32}*\frac{fs}{16}}$)	
4	Dense	D*F_1*$\frac{T}{\frac{fs}{32}*\frac{fs}{16}}$		D*F_1*$\frac{T}{\frac{fs}{32}*\frac{fs}{16}}$ + 1	1	Softmax

consistently yielded good accuracy in preliminary experiments on the training set. Due to the perfect balance of the dataset, accuracy was adopted as the main metric for binary classification performance evaluation. For the sake of comparability, the F1 score was also reported.

Finally, a statistical analysis was conducted to assess the differences between the performance obtained with the different architectures and with the different fine-tuning conditions. The Jarque-Bera test was used to test the assumptions of normality. Then, depending on the normality of the sample, the paired t-test or the Wilcoxon signed rank test were used to compare each models pair. To be noted, these tests were performed individually for each fine-tuning percentage. Successively, repeated measures ANOVA test was adopted to compare the performance of a single model across the multiple investigated percentages. For all statistical tests, the significance level was set to 0.05.

IV. Results

A. Dataset

A benchmark dataset was used to conduct the analyses, namely the BCI competition IV dataset 2a [25]. It contains EEG signals of nine healthy subjects recorded by means of 22 wet electrodes. The sampling frequency was equal to $250\,\mathrm{Sa/s}$. Four motor imagery tasks were contained in the dataset. However, two tasks were selected as classes for the discrimination task for this study, namely right hand and left hand. For each subject, two sessions were recorded in two different days, namely "T" and "E" sessions. As the present work inspected an inter-subjective approach, all investigations in this study considered only session "T". It contains 72 trials per motor imagery class. Each trial comprises a $2.00\,\mathrm{s}$ of fixation cross at the beginning. Then, a cue appeared and remained on the screen for $1.25\,\mathrm{s}$. The subjects were asked to perform the motor imagery task from $3.00\,\mathrm{s}$ to $6.00\,\mathrm{s}$. Finally, a short break followed. In the current study, each EEG trial was epoched from $2.00\,\mathrm{s}$ to $6.00\,\mathrm{s}$.

B. Inter-subject training results

Table II presents a comparison of the four models investigated in this study (i.e., EEGNet-4, EEGNet-32, Sinc-EEGNet-4, and Sinc-EEGNet-32). The comparison includes the number of trainable parameters and across subjects mean accuracy and F1 score, together with their associated standard deviations. The performance of the replicated EEGNet-4 model aligns with previous literature [26] when employing the same dataset, binary task, and leave-one-subject-out cross-validation. The sole statistically significant improvement was achieved by Sinc-EEGNet-32. Increasing the number of initial temporal filters from 4 to 32 in EEGNet resulted in negligible average gain, confirming the compactness emphasised by EEGNet authors and supporting EEGNet-4 as the optimal model version.

Conversely, when comparing EEGNet-4 and Sinc-EEGNet-4, an increase from $68 \pm 9\,\%$ ($65 \pm 15\,\%$) to $73 \pm 10\,\%$ ($72 \pm 11\,\%$) in the average accuracy (F1) performance was obtained. This average improvement supports the convenience of adopting a sinc-layer in place of a standard convolutional one. As a matter of fact, the former forces the network to focus on high-level tunable parameters with the greatest impact on the final temporal/frequency filtering process. Moreover,

	EEGNet-4	EEGNet-32	Sinc-EEGNet-4	Sinc-EEGNet-32
# params	1101	12385	481	7425
accuracy (%)	68 ± 9	69 ± 9	73 ± 10	78 ± 10*
F1 (%)	65 ± 15	68 ± 11	72 ± 11	75 ± 12*

TABLE II: Inter-subject training results in terms of number of trainable parameters and across subject mean accuracy and F1 with their associated standard deviation. * indicates statistical significance ($p < 0.05$) of paired t-test in comparison with the reference model EEGNet-4.

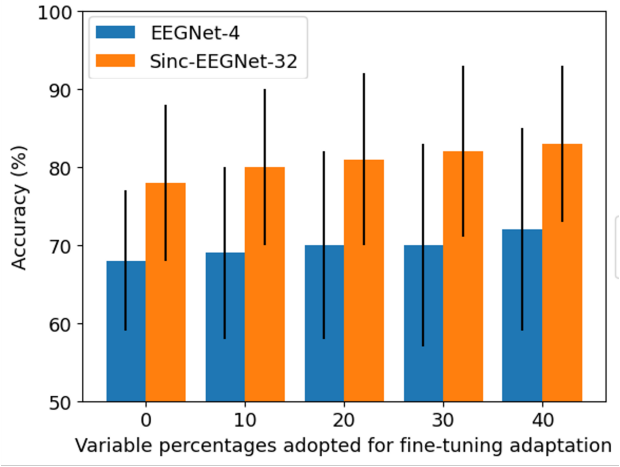

Fig. 1: Inter-subject transfer learning results when applying fine-tuning on the test subject with variable percentages. The across subject mean accuracy with their associated standard deviation are represented by bars heights and error bars, respectively.

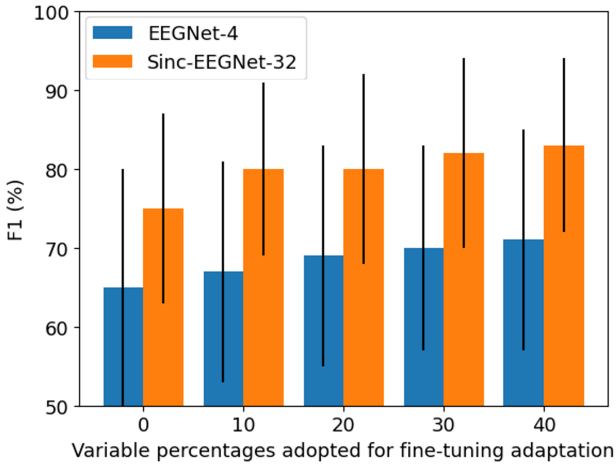

Fig. 2: Inter-subject transfer learning results when applying fine-tuning on the test subject with variable percentages. The across subject mean F1 with their associated standard deviation are represented by bars heights and error bars, respectively.

Sinc-EEGNet-4 offers a very efficient global architecture: the total amount of trainable parameters is more than halved with respect to the EEGNet-4.

Finally, $10\,\%$ average improvements in accuracy (F1) were obtained when EEGNet-4 was compared with Sinc-EEGNet-32. Specifically, the results increased from $68 \pm 9\,\%$ ($65 \pm 15\,\%$) to $78 \pm 10\,\%$ ($75 \pm 12\,\%$). In that case, the difference resulted statistically significant. Hence, in Sinc-EEGNet architecture, the employment of sinc-filters truly enables to benefit from the augmentation of the exploitable features for the final classification. Notably, this possibility is intrinsically addressable to the more efficient sinc-layer introduction, as lack of improvement was reported in EEGNet-32 enlarged version. The chance to dispose of more learnable temporal filters that directly process raw EEG signals is peculiarly relevant for motor imagery decoding. Indeed, this represents a key resource to allow an exhaustive exploration of the wide motor imagery-related frequency range of interest without necessarily renouncing an elevated frequency selectivity.

C. Fine-tuning adaptation results

Figures 1 and 2 report a quantitative characterization of the trade-off between the achievable classification performance and the amount of data employed for the fine-tuning adaptation to the test subject. Results obtained with EEGNet-4 and Sinc-EEGNet-32 were compared. As the previous analysis revealed a negligible gain when increasing the initial temporal filters in the EEGNet model, EEGNet-4 was considered as reference for the present analysis. In the figures, the percentage of data exploited for the test subject is reported on the x-axis. The across subject mean accuracy and F1 expressed in percent and their associated standard deviations are represented by bars heights and error bars, respectively. It is worth noting that the $0\,\%$ bars exactly correspond to the results reported in Table II. For the employed dataset, each trial lasts $8.0\,\mathrm{s}$ and the total number of trials in a session is 144, with a total duration of approximately $19.0\,\mathrm{min}$. Hence, the investigated percentages correspond to calibrations time of: $1.9\,\mathrm{min}$, $3.7\,\mathrm{min}$, $5.7\,\mathrm{min}$ and $7.6\,\mathrm{min}$, respectively.

Interestingly, Sinc-EEGNet-32 model reached an average of $80 \pm 10\,\%$ in both accuracy and F1 metrics by exploiting just the $10\,\%$ of a session data. It corresponds to a calibration lasting less than $2.0\,\mathrm{min}$. Additionally, an increasing trend was found for Sinc-EEGNet-32 when adopting larger percentages of data for fine-tuning operation. It went up to an average result of $83 \pm 10\,\%$ in both accuracy and F1 metrics when $40\,\%$ of session trials were employed. A similar trend was observed for EEGNet-4 as well, but the highest obtained average accuracy (F1) value was $72 \pm 13\,\%$ ($71 \pm 14\,\%$).

As evidenced in the first row of Table II, Sinc-EEGNet-32 globally offers reduced efficiency with respect to EEGNet-4. Notably, despite that, it was found that Sinc-EEGNet-32 was not penalized concerning the benefits it could actually retrieve from a short calibration. As a proof, the relative average improvements computed as accuracy (F1) difference between the $40\,\%$ and $0\,\%$ conditions of EEGNet-4 and Sinc-EEGNet-32 respectively were: $4\,\%$ ($6\,\%$) and $5\,\%$ ($8\,\%$). A further confirmation was provided by the statistical analysis conducted to compare the two models at different fine-tuning percentages. Statistical improvements were observed for all the investigated percentages, except for the $20\,\%$. However, repeated measures ANOVA test revealed that Sinc-EEGNet-32 performance in the no extra-calibration scenario did not statistically differ from those resulting from any percentage of fine-tuning.

V. Conclusion

The aim of this study is to enhance the usability of motor imagery-based BCIs by developing a decoding solution that can be applied to multiple subjects, offering improved performance on an individual basis and reduced calibration need. The proposed Sinc-EEGNet-32 overcomes the requirement of subject-specific data from previous sessions, and at the same time, minimal calibration times resulted sufficient for a satisfying adaptation to the present BCI session. Remarkably, referring to the popular gold deep learning architecture EEGNet-4, statistically significant performance improvements were achieved in inter-subject modality by the proposed Sinc-EEGNet-32 model due to the combination of two main modifications to the EEGNet reference model: the insertion of a sinc-layer instead of the traditional convolutional layer and the enlargement of the architecture width. Interestingly, sinc-layer was found to have a determinant role in allowing an augmentation of the features exploitable for the classification. Finally, the Sinc-EEGNet-32 high suitability for a successful transfer learning application was demonstrated by reaching an average accuracy of $80\,\%$ with less than $2.0\,\mathrm{min}$ calibration. The repeated use of the BCI system by a user can be investigated in the future. In addition, a comprehensive comparison with other existing approaches should be made to provide a more complete evaluation of the performance of the proposed model.

References

[1] O. Mokienko, L. Chernikova, A. Frolov, and P. Bobrov, "Motor imagery and its practical application," *Neuroscience and Behavioral Physiology*, vol. 44, no. 5, pp. 483–489, 2014.

[2] U. Chaudhary, N. Birbaumer, and A. Ramos-Murguialday, "Brain–computer interfaces for communication and rehabilitation," *Nature Reviews Neurology*, vol. 12, pp. 513–525, 2016.

[3] S. Saha, K. I. U. Ahmed, R. Mostafa, L. Hadjileontiadis, and A. Khandoker, "Evidence of variabilities in EEG dynamics during motor imagery-based multiclass brain–computer interface," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 2, pp. 371–382, 2017.

[4] S. C. Wriessnegger, G. R. Müller-Putz, C. Brunner, and A. I. Sburlea, "Inter-and intra-individual variability in brain oscillations during sports motor imagery," *Frontiers in human neuroscience*, vol. 14, p. 576241, 2020.

[5] A. Apicella, F. Isgrò, A. Pollastro, and R. Prevete, "On the effects of data normalization for domain adaptation on EEG data," *Engineering Applications of Artificial Intelligence*, vol. 123, p. 106205, 2023.

[6] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger, "A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update," *Journal of neural engineering*, vol. 15, no. 3, p. 031005, 2018.

[7] K. Ang, Z. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (fbcsp) in brain-computer interface," *Proceedings of the International Joint Conference on Neural Networks*, pp. 2390 – 2397, 07 2008.

[8] P. Arpaia, A. Esposito, A. Natalizio, and M. Parvis, "How to successfully classify EEG in motor imagery BCI: a metrological analysis of the state of the art," *Journal of Neural Engineering*, 2022.

[9] 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, 2009.

[10] Y. Jiao, Y. Zhang, X. Chen, E. Yin, J. Jin, X. Wang, and A. Cichocki, "Sparse group representation model for motor imagery eeg classification," *IEEE journal of biomedical and health informatics*, vol. 23, no. 2, pp. 631–641, 2018.

[11] "A review of critical challenges in mi-bci: From conventional to deep learning methods," *Journal of Neuroscience Methods*, vol. 383, p. 109736, 2023.

[12] M. Zheng and B. Yang, "A deep neural network with subdomain adaptation for motor imagery brain-computer interface," *Medical Engineering & Physics*, vol. 96, pp. 29–40, 2021.

[13] O. George, S. Dabas, A. Sikder, R. Smith, P. Madiraju, N. Yahyasoltani, and S. I. Ahamed, "Enhancing motor imagery decoding via transfer learning," *Smart Health*, vol. 26, pp. 328–339, 2022.

[14] B. Sun, Z. Wu, Y. Hu, and T. Li, "Golden subject is everyone: A subject transfer neural network for motor imagery-based brain computer interfaces," *Neural Networks*, vol. 151, pp. 111–120, 2022.

[15] H. Shimizu and R. Srinivasan, "Improving classification and reconstruction of imagined images from eeg signals," *PLOS ONE*, vol. 17, pp. 1–16, 09 2022.

[16] A. Bria, C. Marrocco, and F. Tortorella, "Sinc-based convolutional neural networks for eeg-bci-based motor imagery classification," p. 526–535, 2021.

[17] J. Chen, D. Wang, W. Yi, M. Xu, and X. Tan, "Filter bank sinc-convolutional network with channel self-attention for high performance motor imagery decoding," *Journal of Neural Engineering*, vol. 20, p. 026001, mar 2023.

[18] V. Lawhern, A. Solon, N. Waytowich, S. Gordon, C. Hung, and B. Lance, "Eegnet: A compact convolutional network for eeg-based brain-computer interfaces," *Journal of Neural Engineering*, vol. 15, 2016.

[19] M. Ravanelli and Y. Bengio, "Speaker recognition from raw waveform with sincnet," pp. 1021–1028, 2018.

[20] O. Ozdenizci, Y. Wang, T. Koike-Akino, and D. Erdogmus, "Learning invariant representations from eeg via adversarial inference," *IEEE Access*, vol. PP, pp. 1–1, 02 2020.

[21] P. Autthasan, R. Chaisaen, T. Sudhawiyangkul, P. Rangpong, S. Kiatthaveephong, N. Dilokthanakul, G. Bhakdisongkhram, H. Phan, C. Guan, and T. Wilaiprasitporn, "Min2net: End-to-end multi-task learning for subject-independent motor imagery eeg classification," *IEEE Transactions on Biomedical Engineering*, vol. 69, pp. 2105–2118, 2021.

[22] S. Kunjan, T. Grummett, K. Pope, D. Powers, S. Fitzgibbon, T. Bastiampillai, M. Battersby, and T. Lewis, *The Necessity of Leave One Subject Out (LOSO) Cross Validation for EEG Disease Diagnosis*, pp. 558–567. 09 2021.

[23] D. Han, S. Musellim, and D. Kim, "Confidence-aware subject-to-subject transfer learning for brain-computer interface," *CoRR*, 2021.

[24] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2014.

[25] C. Brunner, R. Leeb, G. Müller-Putz, A. Schlögl, and G. Pfurtscheller, "BCI competition 2008–Graz data set A," *Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology*, vol. 16, pp. 1–6, 2008.

[26] "A mutli-scale spatial-temporal convolutional neural network with contrastive learning for motor imagery eeg classification," *Medicine in Novel Technology and Devices*, vol. 17, p. 100215, 2023.