IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 26 May 2024, accepted 7 June 2024, date of publication 11 June 2024, date of current version 20 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3412710

RESEARCH ARTICLE

Multi-Stream 1D CNN for EEG Motor Imagery Classification of Limbs Activation

DANILO AVOLA[1], (Member, IEEE), LUIGI CINQUE[1], (Senior Member, IEEE), ANGELO DI MAMBRO[1], ROMEO LANZINO[1], (Graduate Student Member, IEEE), DANIELE PANNONE[1], (Member, IEEE), AND FRANCESCO SCARCELLO[2]

[1]Department of Computer Science, Sapienza University of Rome, 00185 Rome, Italy
[2]Department of Computer Engineering, Modeling, Electronics and Systems, University of Calabria, 87036 Arcavacata, Italy

Corresponding author: Daniele Pannone (pannone@di.uniroma1.it)

This work was supported in part by the "A Brain–Computer Interface (BCI) Based System for Transferring Human Emotions Inside Unmanned Aerial Vehicles (UAVs)" Sapienza University Research Projects under Grant RM1221816C1CF63B; in part by "EYE-FI.Artificial Intelligence (AI): Going bEYond computEr Vision Paradigm Using wi-FI Signals in AI Systems" Project of Italian Ministry of Universities and Research (MUR) within Progetti di Rilevante Interesse Nazionale (PRIN) 2022 Program under Grant 2022AL45R2 (CUP: B53D23012950001); in part by Made in Italy–Circular and Sustainable (MICS) Extended Partnership; in part by the Next-Generation European Union (EU) (Italian PNRR–M4 C2, Invest 1.3–D.D. 1551.11-10-2022, PE00000004) under Grant CUP MICS B53C22004130001; and in part by PNRR Project FAIR - Future AI Research (PE00000013), Spoke 9 - Green-aware AI, through the NRRP MUR Program funded by the NextGenerationEU.

ABSTRACT Determining the motor intentions of an individual through the analysis of electroencephalograms (EEGs) is a challenging task that concurrently holds considerable potential in aiding subjects with motor dysfunctions. Moreover, thanks to the recent advances in artificial intelligence models and EEG acquisition devices, such analyses can be carried out with ever higher accuracy. The latter aspect covers great importance, since the EEG analysis of subjects whose mental efforts are focused on moving limbs is frequently used for crucial tasks, including the control of interactive interfaces and prosthetic devices. In this paper, a novel multi-stream 1D Convolutional Neural Network (CNN) architecture is proposed. The input EEG signal is processed by four convolutional streams, which differ in the size of convolutional kernels, thus allowing the extraction of information at different time scales. The resulting 1D feature maps are then fused together and provided to a dense classifier to identify which limb the subject intended to move. Comprehensive experiments conducted on PhysioNet EEG motor movement/imagery dataset, which remains the reference collection of data in this application context, have demonstrated that the proposed model surpasses the key works in the current state-of-the-art in both cross-subject and intra-subject settings.

INDEX TERMS Motor imagery, EEG analysis, multi-stream, 1D convolutional neural networks (1D CNNs), brain–computer interfaces (BCIs).

I. INTRODUCTION

The last two decades have seen significant advances in invasive and non-invasive Brain-Computer Interfaces (BCIs). Invasive techniques for BCI applications entail the direct placement of electrodes onto brain tissue, thus enabling the acquisition, and occasionally induction, of high-definition and granular electrical signals. These interfaces find application in a wide range of tasks, including cognitive enhancement [1], epilepsy monitoring and treatment [2], restoration of sensory functions [3], and many others. Despite the outstanding outcomes achieved, invasive techniques for BCI applications are presently restricted to specific cases where alternative solutions are not feasible. The utilization of these interfaces involves several critical aspects, including surgical intervention and constant monitoring of the implants over time. Thanks to the ever-increasing quality of the information acquired in terms of spatial and temporal resolution, non-invasive techniques for BCI applications have taken on a more prominent role in recent years. As with invasive interfaces, non-invasive ones are also used in different application areas, depending on the type of information acquired and

The associate editor coordinating the review of this manuscript and approving it for publication was Md Kafiul Islam.

the nature of the devices, which can vary in terms of cumbersomeness. For example, Functional Near-Infrared Spectroscopy (fNIRS), which measures brain activity by detecting changes in the concentration of oxygenated and deoxygenated hemoglobin in brain tissues, is utilized in a broad spectrum of brain functional and cognitive research, such as mental workload assessment [4], identification of effective connectivity [5], and visuo-motor analysis [6]. On the other hand, Functional Magnetic Resonance Imaging (fMRI), which typically measures changes in blood flow associated with neuronal activity, is mainly employed for analysis and investigation purposes [7], [8], [9], as well as clinical and brain mapping applications [10], [11], [12]. Among the non-cumbersome and non-invasive techniques, the most used in terms of versatility and usability are, without a doubt, the EEG-based interfaces. Their high accuracy in acquiring signals is making them indispensable in an extensive range of challenging tasks, such as person identification [13], [14], [15], emotion recognition [16], [17], [18], neural rehabilitation [19], [20], [21], motor imagery [22], [23], [24], and many others. Regarding motor imagery (MI), it is undoubtedly one of the most fascinating and valuable fields for EEG-based interfaces. These devices enable the design of low-cost, high-performance systems capable of significantly improving the quality of life for a wide range of individuals with motor dysfunctions.

It is well-known that methodologies addressing MI focus on interpreting predominant brain areas that are activated when a subject mentally simulates movements or actions without actual physical execution. The nature of these imagined activities depends on their intended goals, which can range from simple device interactions and performing rehabilitation exercises to piloting mechanical prostheses. Unsurprisingly, recent years have seen remarkable results in complex tasks, thanks to the ever-increasing use of modern BCIs. These interfaces allow for the acquisition of EEG signals with increasingly accurate and coherent spatio-temporal content.

The current literature features a growing number of works from different research groups, each pushing the boundaries of MI forward. For example, the authors of [25] propose a novel three-dimensional capsule network EEG signal recognition model. Initially, this model employs a multi-layer 3D convolution module for feature extraction. Subsequently, it integrates a capsule network to obtain advanced spatial features, and finally, it uses dynamic routing connections and squash functions for classification. Similarly, the approach presented in [26] introduces an enhanced complex-valued common spatial pattern, which aims to analyze EEG signals by decomposing them into spatial patterns through covariance. The primary objective of their method is to maximize the variance between signals associated with distinct classes while minimizing the variance between signals within the same class. An alternative solution for analyzing MI tasks is reported in [27], where a phase locking value-based method linked with the synchronization between EEG signals from different areas of the brain is presented. The authors in [28] propose a real-time MI analysis based on examining single-channel EEG signals; in particular, they suggest a thresholding method to facilitate the differentiation between transients and low-frequency components within such signals.

Moving on to a more generalized model, the authors in [29] propose a method that uses wavelet transform for the EEG feature extraction phase and a simple Multi-Layer Perceptron (MLP) with a single hidden layer for classification. They applied it to both mental and MI tasks to demonstrate the versatility of the proposed approach. Similar to the work reported in [29], the authors in [30] present an algorithm for cursor manipulation on a screen through a BCI. In this case, a wavelet transform is used for the feature extraction stage. For the subsequent classification process, a hybrid system made up of a Support Vector Machine (SVM) and an MLP was developed. This combination of models aimed to enhance the accuracy and efficiency of cursor control using EEG signals. By imagining moving fists and feet, the subject moves the cursor in the four cardinal directions, while closing the eyes acts as a left-button click.

Analyzing the vast literature on MI, it can be seen that various research groups, in addition to solving specific tasks, are also focused on addressing intrinsic aspects of EEG signals, such as redundancy, dimensionality, and optimization. For example, an interesting paper reported in [31] describes how the authors use a shrinkage-regularized common spatial pattern to project EEG signals and select features using a minimum redundancy maximum relevancy algorithm. Subsequently, an SVM is adopted for the classification of intentions. Authors in [32] and [33], on the other hand, focus their work on the dimensionality reduction of EEG signals through an iterative multi-objective optimization for channel selection algorithm and Granger causality, respectively. The research outlined in [34] centers around a pre-processing stage designed to decompose μ and β rhythms from EEG signals. The approach employs uncorrelated transform complex common spatial patterns to enhance the independence between the identified rhythms. Subsequently, machine learning (ML) algorithms are used to extract various valuable features from both rhythms. The authors opposing the work reported in [35] primarily focus on the analysis of the δ rhythm. The methodology employed in this research involves the use of a Light Gradient Boosting Machine (LightGBM) classifier, which is based on Decision Trees (DT). The study explores the interplay between the δ rhythm and the classifier, thus providing insights into their relationship and potential implications. Finally, a recent study presented in [36] explores the use of attention mechanism-based deep learning (DL) models to construct subject-independent MI interfaces. This study shows that these mechanisms also aid in achieving remarkable results for practical case studies in this area.

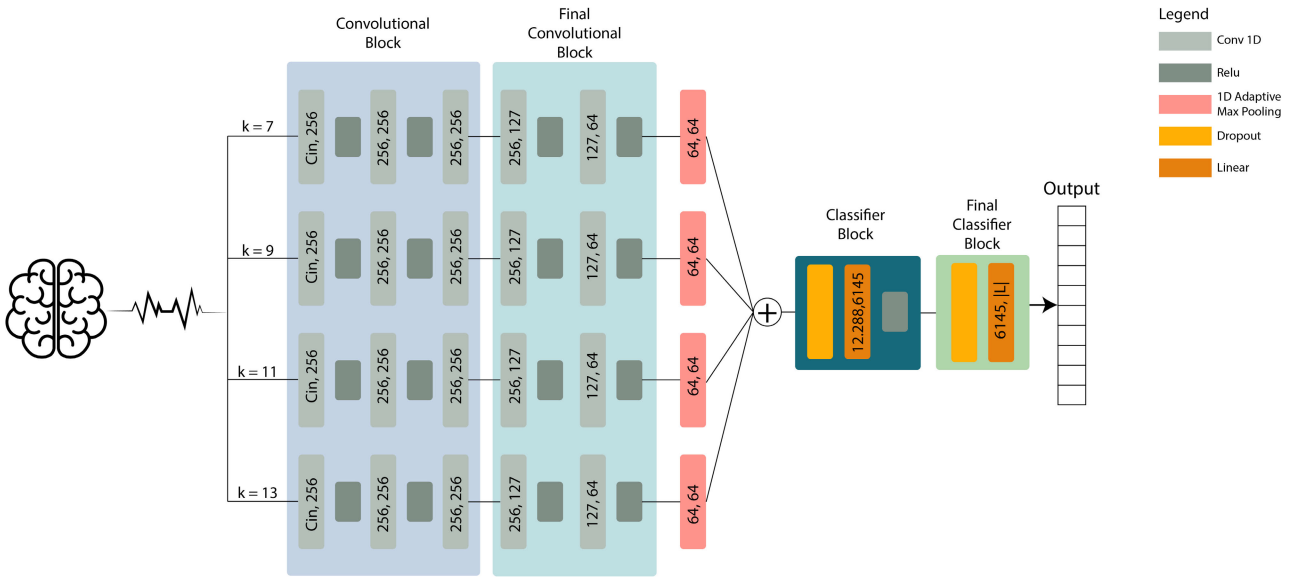


FIGURE 1. The architecture of the proposed model. The latter is composed by four CNN streams, which are defined in Table 1, and a set of fully connected layers, which are defined in Table 2.

In addition to what is reported above, and thanks to the ever-increasing computational capabilities of current Graphical Processing Units (GPUs), it should be highlighted that the last decade has seen continuous growth in the use of DL models for analyzing EEG signals, including MI-based tasks. For example, the authors in [37] employ a Convolutional Neural Network (CNN) within an end-to-end framework to simultaneously extract features and classify EEG signals. This approach eliminates the necessity for ad-hoc, non-deep feature extraction algorithms. The results show a substantial enhancement in both intra-subject and cross-subject accuracies compared to prior non-deep methodologies. Similarly, the work reported in [38] uses a Long Short-Term Memory (LSTM) neural network in conjunction with an attention mechanism, diverging from the previously mentioned approach. Alternatively, the authors in [39] introduce a virtual reality (VR) serious game designed for stroke rehabilitation. The game employs a CNN, and the outcomes align closely with those reported in [37]. The work proposed in [40], instead, utilizes the positioning of EEG electrodes to rearrange the input for a Graph-based Convolutional Recurrent Attention Model (G-CRAM), which performs feature extraction and classification by leveraging the spatiality of the data. Finally, the CNN used in [41] incorporates the local re-parameterization trick, a method that can generate a gradient estimator to reduce variance. Confirming the leap in quality due to the ever-increasing use of deep learning models to analyze EEG signals is the work described in [42] where, moreover, the absolute goodness of CNNs in terms of accuracy and versatility emerges. For all the reasons seen so far and in light of the remarkable results obtained, this paper proposes a novel architecture based on multi-stream 1D CNN where the input EEG signal is processed by four convolutional streams, which differ in the size of convolutional kernels. The resulting 1D feature maps are then fused together and provided to a dense classifier to identify which limb the subject intended to move. The key contributions of this work can be summarized as follows:

- a novel modular multi-stream 1D CNN architecture for the feature extraction stage that can retrieve meaningful information by analyzing the same data considering different time scales;
- in the conducted ablation study, it has emerged that with this specific type of data, shallow models with more streams achieve better results with respect to deeper models with less streams;
- comparative experiments on PhysioNet EEG motor movement/imagery dataset, the reference data collection for MI application field, that show how the proposed model overcomes the key works of the current literature in both cross-subject and intra-subject analysis.

As a final consideration, employing modern deep models in EEG analysis could yield interesting insights into neuroscience. For instance, CNNs could enhance feature extraction from noisy EEG signals, or alternately, by focusing on the classification of these signals, it could be possible to develop advanced prosthetic limbs. Additionally, deep models could be leveraged for diagnosing neurological disorders such as epilepsy, Alzheimer's disease, and attention-deficit/hyperactivity disorder (ADHD).

The rest of the paper is structured as follows. Section II describes in detail the proposed architecture, specifying how the different streams and the classifier are composed. Section III describes the used dataset together with the performed experiments; moreover, it shows the obtained results compared with the current literature. Section IV reports the conducted ablation study for the modeling of the final architecture. In Section V, the limitations of the

proposed method are highlighted, together with the future works. Finally, Section VI concludes the paper.

II. METHODOLOGY

The proposed deep architecture is a CNN that approximates the function $m : \mathbb{R}^{b \times s \times c_{in}} \rightarrow \mathbb{R}^{b \times |L|}$, which maps each EEG in b to a probability distribution. In detail, we have:

- $b \in \mathbb{N} \geq 1$ represents the number of EEG trials in each batch given as input to the CNN during the training and inference phases. As pointed out in [43], this hyperparameter must be accurately chosen based on the problem being treated and the available hardware since it directly affects the performance, both in means of computational costs and of classification result;
- $s \in \mathbb{N} \geq 1$ is the number of samples for each EEG trial calculated as $s = \lfloor r \cdot t \rfloor$, where $r \in \mathbb{R} > 0$ indicates the sampling rate, expressed in Hertz (Hz), and $t \in \mathbb{R} > 0$ the recording time in seconds (s). Each EEG trial used in the training and testing phases for this work has a fixed length;
- $c_{in} \in \mathbb{N} \geq 1$ is the number of EEG channels recorded during the experiments. This parameter mainly depends on the device used for recording the data. In this work, we used all the available data for each trial;
- L is the set of classes (also called labels) of each trial in the used dataset, corresponding to $\{A_l, A_r\}$ for the PhysioNet dataset, where A_l and A_r mean left or right arm activation, respectively. Each trial is labeled either A_l or A_r.

The proposed model m consists of two main sub-modules: a set of convolutional streams, m_s, described in detail in Section II-A, whose purpose is to extract different features from the input, and the classifier, m_c, covered in Section II-B, that outputs a tensor of probabilities (in the form of logits) for L.

A. CONVOLUTIONAL STREAMS

A convolutional stream $m_{s_i} \in m_s$ is composed of one or more convolutional blocks, each one composed of 1D convolutional layers. All the convolutional streams present the same structure, differing solely in the parameter denoted as k, namely, the size of the kernels of each stream. The idea behind employing kernels of varying sizes is to facilitate the extraction of information from each channel across diverse temporal scales. Let us define the kernel size of the first stream as k_1. The subsequent streams i will have a kernel size of $k_i = k_{i-1} + 2$. For example, if $k_1 = 3$, we will have that $k_2 = 5, k_3 = 7, k_4 = 9 \ldots k_i = k_{i-1} + 2$.

A convolutional block is formed by a sequence of two 1D convolutional layers with kernel size k, padding $\lfloor \frac{k}{2} \rfloor$ and stride 1, and a ReLU activation function after each convolutional layer. All convolutional blocks, except for the final one, present a final 1D convolutional layer with a stride equal to 2, to reduce the length of the samples. Next to the final convolutional block, there is a 1D Adaptive Max Pooling layer, that shapes the input to a fixed length preserving the number of channels [44], and a flattening layer, which purpose is to reshape the obtained feature map to the form accepted by the classifier. A convolutional stream approximates the function $m_{s_i} : \mathbb{R}^{b \times s \times c_{in}} \rightarrow \mathbb{R}^{b \times (c_{out} \cdot p)}$, where $c_{out} = 64$ is the number of output channels of the final convolution and $p = 48$ the final output size of the 1D Adaptive Max Pooling layer.

In the proposed model, we have that the optimal number of streams is 4, and the parameters k for the kernel are, respectively, 7, 9, 11, and 13 for m_{s_1}, m_{s_2}, m_{s_3}, and m_{s_4}. The rest of the parameters are reported in Table 1.

TABLE 1. Layers and parameters for each stream of the proposed model.

	Layers	Input channels	Output channels	Stride	Padding
Convolutional Block	1D Conv.	c_{in}	256	1	$\lfloor \frac{k}{2} \rfloor$
	ReLU	-	-	-	-
	1D Conv.	256	256	1	$\lfloor \frac{k}{2} \rfloor$
	ReLU	-	-	-	-
	1D Conv.	256	256	2	$\lfloor \frac{k}{2} \rfloor$
Final Convolutional Block	1D Conv.	256	127	1	$\lfloor \frac{k}{2} \rfloor$
	ReLU	-	-	-	-
	1D Conv.	127	64	1	$\lfloor \frac{k}{2} \rfloor$
	ReLU	-	-	-	-
Pooling Block	1D Adaptive Max Pooling	64	64	-	-

TABLE 2. Layers and parameters for the classifier of the proposed model, where L is the set of labels.

	Layers	Input features	Output features	Dropout probability
Classifier Block	Dropout	-	-	50%
	Linear	$64 \cdot 48 \cdot 4$	6,145	-
	ReLU	-	-	-
Final Classifier Block	Dropout	-	-	50%
	Linear	6,145	$\lvert L \rvert$	-

B. CLASSIFIER

The classifier block m_c is composed of a sequence of fully connected layers, which get as input the concatenation of the flattened features maps obtained through the convolutional blocks of each stream $m_{s_i} \in m_s$. Each fully connected layer, is preceded by a dropout layer [45], which prevents neurons from excessive co-adaptation by randomly dropping each one and its connections with a fixed probability during training. In addition, every linear layer except the last one is followed by a ReLU activation function to add non-linearity. The classifier block approximates the function $m_c : \mathbb{R}^{b \times (c_{out} \cdot p \cdot |m_s|)} \rightarrow \mathbb{R}^{b \times |L|}$. The parameters chosen for the model used in the results are reported in Table 2.

III. EXPERIMENTS AND DISCUSSION

In this section, the dataset, experimental setup, and experimental protocol are presented first. Subsequently, the results

obtained from the intra-subject and cross-subject experiments, along with a comparison to the current literature, are discussed.

A. DATASET

The PhysioNet EEG motor movement/imagery dataset [46], [47] is a collection of EEGs from 109 subjects recorded during the execution of four tasks: opening and closing left or right hand, imagining opening and closing them, opening and closing both hands or both feet, imagining opening and closing them. Such data has been recorded using a 64-channel, 160 Hz Brain Computer Interface. In this work, we focused on the recordings belonging to the second task, i.e., imagining opening and closing the right and left hand. No pre-processing steps were used on the dataset samples. This choice is due to the robustness of the CNN models in extracting features also in contexts in which the data present some noise.

B. EXPERIMENTAL SETUP

The final model architecture employed in the experiments was determined by considering the optimal configuration that emerged from the ablation study delineated in Section IV. Consequently, the model comprises four distinct streams in m_s, each containing two convolutional blocks. The kernel size for the first stream is set to $k = 7$ to analyze a sufficiently wide temporal window. In our experiments, we obtained the best results with one convolutional block, one final convolutional block, and one pooling block. The structural delineation of each of these blocks is explained in detail in Table 1. The classifier, m_c, comprises two linear layers, as depicted in Table 2. Figure 1 illustrates our final multi-stream CNN architecture.

The model has been implemented by using the Pytorch framework, while the EEG data has been handled with the MNE [48] framework. The model has been trained and tested on a machine with a 2-core Intel Xeon CPU at 2.2 GHz, 13 GiB of RAM and with an nVidia Tesla T4 GPU bundled with 16 GiB of GDDR6 RAM. The code related to our model is available on GitHub.[1]

C. EXPERIMENTAL PROTOCOL

In this work, we focused on two types of experiments: cross-subject (Section III-C1) and intra-subject (Section III-C2) classification of MI-based EEGs.

1) CROSS-SUBJECT

The following is the formal definition of the cross-subject classification task. Given a pool $E = \bigcup_{1 \leq i \leq N}(x_i, y_i)$ of N labeled EEGs belonging to all the subjects, a model is trained on a subset $E_t \subset E$ to identify patterns in EEG signals to predict the thought action, and its performance is evaluated on a subset $E_v \subset E$, where $E_t \cap E_v = \emptyset$ and $E_t \cup E_v = E$. Each sample $(x, y) \in E$ is formed by an EEG record $x \in \mathbb{R}^{s \times c_{in}}$ of a fixed duration of s samples and c_{in} channels, and a label $y \in L$ that indicates the action being thought by the subject in that trial.

[1] https://github.com/Prometheus-Laboratory/2024_MultiStream1D

TABLE 3. Comparison with the related work on the PhysioNet dataset. Bold values are the best ones in their respective columns.

Work	Year	Subjects used	Intra-subject accuracy	Cross-subject accuracy
[29]	2013	109	68.2%	-
[30]	2014	100	75%	-
[34]	2016	24	80.1%	-
[41]	2022	50	92.4%	-
[27]	2014	103	-	71.6%
[26]	2014	56	-	72.4%
[40]	2020	105	-	74.7%
[39]	2019	105	-	80.5%
[32]	2016	85	63.6%	60.1%
[37]	2018	105	86.5%	80.4%
[38]	2020	103	98.3%	83.2%
[33]	2021	105	98.3%	83.6%
Ours	2023	105	**99.4%**	**85.9%**

2) INTRA-SUBJECT

Intra-subject classification task has been structured as follows. For each subject $S_i \in S$ in a pool S of subjects, where $S_i = \bigcup_{1 \leq i \leq N}(x_i, y_i)$ is a collection of N labeled EEGs of a single subject, a model is trained on a subset $E_t \subset S_i$ to identify patterns in EEG signals and predict the thought action. Its performance is evaluated on a subset $E_v \subset S_i$ where $E_t \cap E_v = \emptyset$ and $E_t \cup E_v = S_i$. As in section III-C1, each sample $(x, y) \in E$ is formed by an EEG record $x \in \mathbb{R}^{s \times c_{in}}$ and a label $y \in L$. At the end of the process, a model is produced for each subject, and the final intra-subject performances are calculated as the mean of the performances of each model.

D. RESULTS

In this section, the results obtained in the performed intra-subject and cross-subject experiments, together with the comparison with the current literature, are presented.

E. PERFORMANCE METRICS

Given the intricate nature of EEG data, which is characterized by high dimensionality and susceptibility to noise, it is imperative to employ a variety of metrics to comprehensively assess the model performance on the validation sets. In the following paragraphs, the metrics employed for measuring the goodness of our model are described [49].

a: ACCURACY

Represents the ratio between correctly classified samples and their total amount within the dataset. Hence, the accuracy gives an overall measure of model performance.

b: PRECISION

Represents the ratio between correctly classified positive samples and the number of positive predictions. Precision quantifies the number of true positive predictions among

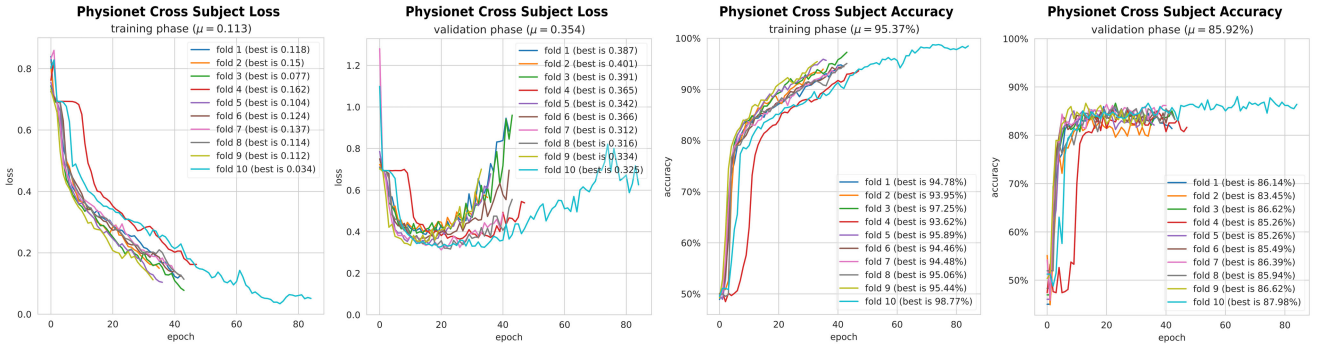


FIGURE 2. Loss values for a) training and b) testing phases on the PhysioNet dataset using 10-fold cross validation, and accuracy values for c) training and d) testing phases on the PhysioNet dataset using 10-fold cross validation. Both accuracy and loss refer to the cross-subject experiment. Each line in the plots represents the i-th fold, while the value μ represents the average value obtained among the several folds. Formally, let $M \in \mathbb{R}^{k_f \times e}$ be the accuracy matrix, where k_f is the number of folds in the cross validation and e the maximum number of epochs, $\mu = \frac{1}{k_f} \sum_{1 \leq i \leq k_f} \max(M_i)$ is equal to the mean of the maximum accuracy reached for each fold. In our experiments, $k_f = 10$ and $e = 100$.

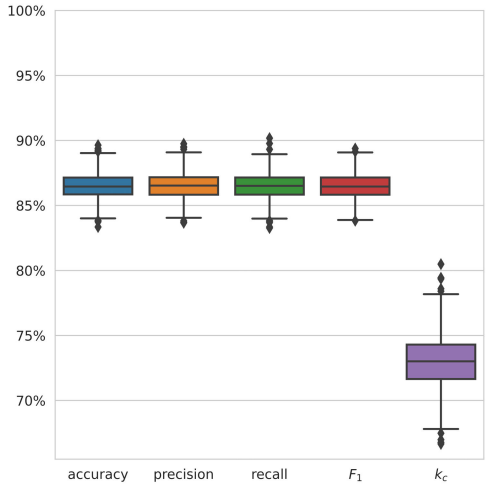


FIGURE 3. The four metrics computed for the cross-subject experiments. For each of the metrics, a box plot is provided, which allows a better overview of the metrics distribution and their variability in the 10 folds.

the total number of positive predictions made, reflecting the model reliability in predicting positive instances.

c: RECALL

Represents the ratio between correctly classified positive samples and the number of positive samples. Recall represents the proportion of actual positive instances that the model correctly identified, providing insight into the model sensitivity to positive instances.

d: F_1 SCORE

Represents the relation between precision and recall using their harmonic mean. F_1 offers a balanced measure of the model's performance, especially when the class distribution is imbalanced. Accuracy, precision, recall, and F_1 score are critical for assessing the model's ability to make correct predictions and minimize errors, providing insights into the model's reliability in real-world applications.

e: COHEN'S K

Represents the agreement between two annotators that classify the items, and also considers the agreement obtained by chance, ranging in $[-1, 1]$ [50]:

$$k_C = \frac{accuracy - p_e}{1 - p_e}, \tag{1}$$

where p_e is the expected agreement when both annotators assign labels randomly. Cohen's k offers a measure of the degree to which the classifications made by the model align with expert annotations, considering random agreement, and thus it adds an additional layer of validation to the assessment. This aspect plays a key role, especially in the medical context, where reliability is fundamental.

The meticulous application of different metrics is crucial in our cross-subject experiments due to the inherent variability and complexity arising from the diversity in subject characteristics, behaviors, and responses. Such a comprehensive approach ensures the nuanced evaluation of model generalizability, reliability, and robustness across different subjects. Conversely, intra-subject experiments typically exhibit less variability, allowing for a more streamlined evaluation with fewer metrics, yet adequately capturing the essence of model performance within individual subjects. The proposed selection of metrics provides a robust way for evaluating the model performance, enabling the revelation of insights and subtleties that may remain elusive when constrained by a narrower set of evaluative tools, often observed in prior research.

Since the dataset is not split into training and validation sets, k-fold cross-validation has been used. It consists in splitting a dataset E in k subsets E_i such that $\bigcap_{1 \leq i \leq k} E_i = \emptyset$ and $\bigcup_{1 \leq i \leq k} E_i = E$, train k models using $k - 1$ splits as the training set and a single split as the validation set such that

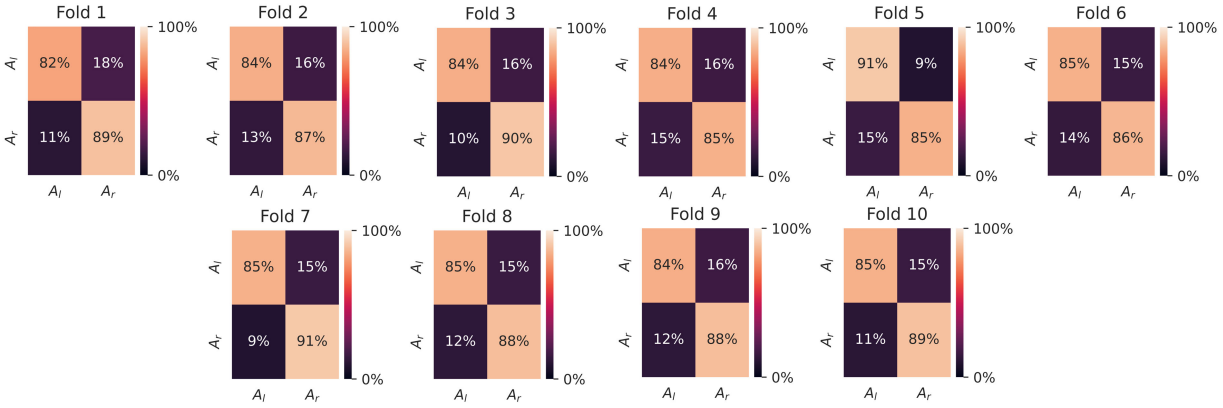


FIGURE 4. Confusion matrix for the cross-subject experiment on the PhysioNet dataset. A_l and A_r are respectively the labels for left and right arm activation.

the validation split is different in each of the k iterations, and finally averaging the performances of the k models.

F. COMPARISON WITH RELATED WORKS

In this section, a comparison of our work with the state-of-the-art is performed, emphasizing distinctions in both cross-subject and intra-subject evaluations.

It is worth noting that numerous works [29], [30], [39], predominantly showcase the accuracy metric due to its straightforwardness, providing a snapshot of model performance. However, this simplification can miss subtleties and imbalances in model evaluation. In opposition, our method, similar to [33] and [38], employs multiple, complimentary metrics, as described in Section III-E, ensuring a nuanced understanding of model performance and offering a more well-rounded insight into its reliability and efficacy. Additionally, the incorporation of both cross-subject and intra-subject evaluations marks a departure from the norm, reflecting a depth observed in only a few studies like [32], [33], [37], and [38]. This inclusive approach affords a comprehensive view of model performance and adaptability across diverse subjects and singular subject scenarios. Focusing on one kind of evaluation potentially overlooks critical insights and nuances that are integral to understanding model behavior and efficacy comprehensively.

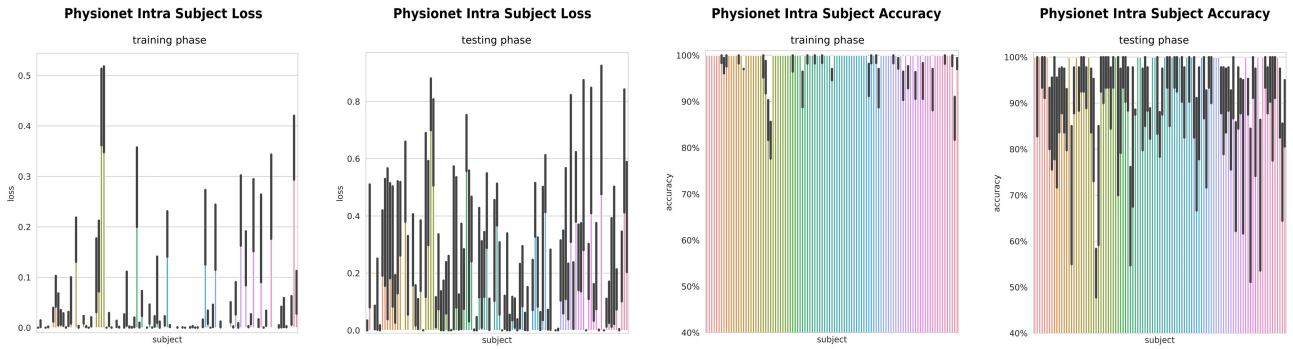
1) CROSS-SUBJECT

In cross-subject experiments, the proposed model achieves state-of-the-art results when compared with the related work. As depicted in Figure 2, the validation loss reaches its minimum at about the twentieth epoch; after that, the model starts overfitting, reaching a mean accuracy score $\mu = 85.9$. Such an accuracy score is obtained by averaging the maximum accuracy of each fold, which consistently outperforms the previous state-of-the-art score achieved in [33] by $\sim 2.3\%$. The latter, as for the work reported in [32], uses an SVM classifier, thus meaning that a data pre-processing step is mandatory to remove noise and artifacts. This step can strongly influence the result of classification since there is the risk of removing relevant information needed to accomplish the task. Instead, when using CNN-based approaches, as the one proposed in this paper and those presented in [37] and [39], it is possible to provide raw EEG data directly to the CNN-based model thanks to its ability, on the one hand, to automatically extract relevant features from input data, on the other hand, to effectively capture local patterns and filter out irrelevant information, thus allowing the model to be robust in real application contexts. In Figure 4, the normalized confusion matrix for the experiments is presented, highlighting that the left arm samples are classified with higher accuracy. The boxplot in Figure 3 allows a more precise examination and comparison of the metrics described in Section III-E, regarding the cross-subject experiments conducted. It is observed that accuracy, precision, recall, and F_1 scores are predominantly situated within an 84% to 90% range, indicating the model's robust performance and efficacy in correctly identifying and classifying instances across subjects. On the contrary, Cohen's k value displays a range from 68% to a maximum of 81%. The relative reduction in Cohen's k values compared to the aforementioned metrics underscores potential disparities in the classification of instances, hinting at a possible presence of false positives/negatives that may be affecting the inter-annotator reliability. This variance in metrics accentuates the intricate dynamics of model performance and emphasizes the criticality of having several metrics to ensure the reliability and validity of the model's predictive capabilities in diverse experimental conditions.

It is worth noting that the feature extraction part of the pipeline has been entirely handled by the streams through the use of the learned convolutions that automatically filter and extract meaningful information from the data, thus allowing the human not to necessarily define a hand-crafted algorithm for noise reduction or channel selection but to focus on the architecture of the neural network.

2) INTRA-SUBJECT

The results of the intra-subject experiments are very promising with a μ value of 0.996, where μ is computed as



(a) (b) (c) (d)

FIGURE 5. Bar plots of a,b) loss and c,d) accuracy per subject on the PhysioNet dataset using 10-fold cross-validation for intra-subject experiments. Each bar represents a subject of the dataset, and the gray part of the bar represents the variability among the several folds of the considered measure.

the mean of the maximum accuracy scores per subject for each fold, with almost all subjects correctly classified as can be seen in Figure 5. In contrast to Figure 2, which utilizes a line plot to represent cross-subject variations, reflecting the results of a single training and validation process, Figure 5 employs a bar plot to illustrate intra-subject variability due to multiple training sessions proportional to the number of subjects. In this case, a bar plot effectively represents and contrasts the discrete, categorical data points, allowing for clear comparisons between individual subjects. The use of different plot types in Figures 2 and 5 is integral to accurately conveying the distinct aspects of the neural network's performance and variability in cross-subject and intra-subject contexts. The results are comparable to the ones obtained by [35], although in the latter work just the δ rhythm has been used: those are waves from 0 to 4 Hz [51] observed in the frontal and occipital region of the cortex and are associated with sleep and tasks in which the subject is required to carefully pay attention [52].

The previous state-of-the-art accuracy of 0.983, achieved by [33] and [38], is approximately 1.3% less accurate than the proposed method. Despite these two works achieving equal intra-subject accuracy, it is worth noting that [33] uses 105 subjects for the experiments, while [38] uses 103 subjects. Also, in the intra-subject experiments, the works that use deep model-based approaches are the ones that achieve higher results, except for the work presented in [33], in which an SVM is used. The reason behind their good results can be imputable to the automatic selection of the channels based on their significance during the MI actions.

IV. ABLATION STUDY

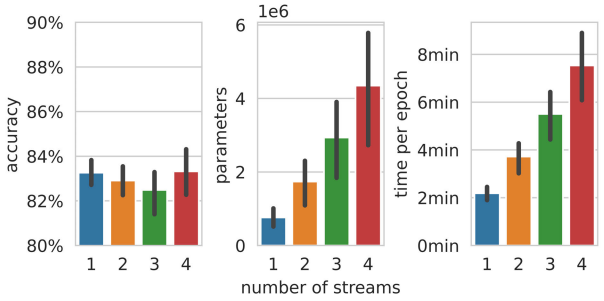
In this section, the choices made during the selection of the architecture of the final model are discussed. The several experiments have been made using a split A_t of the PhysioNet dataset D as the training set, with $|A_t| \approx 0.7 \cdot |D|$, and a test set A_v such that $A_t \cup A_v = D$ and $A_t \cap A_v = \emptyset$. This ablation study consisted of several experiments that focused on different parts of the network: the number of streams and their depth, the structure of the convolutional block, and the depth of the final classifier. In Figure 6, a quantitative representation of the performed experiments is depicted. Within each experimental configuration, three distinct bar plots are systematically illustrated, representing, in sequence, the accuracy, the number of parameters, and the average time required for each training epoch on our benchmark machine described in Section III-B. The y axis delineates the scale of the applied metric, with respective representations for percentage, quantity, and minutes. The x axis portrays the varying values of the parameter under examination. Given that more than one run has been computed for every unique configuration, the bar plots further incorporate a representation of the fluctuations observed among them. These fluctuations are manifested through a black line superimposed on each respective box. This allows for a more comprehensive understanding of the variability inherent in each configuration, thereby providing a nuanced insight into the relative stability and reliability of the tested parameter values.

1) NUMBER OF STREAMS

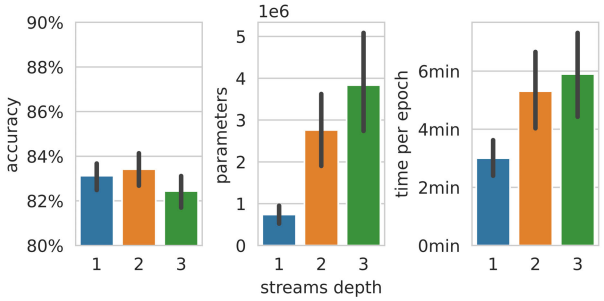
This experimental phase consisted of changing the number of streams of the model, and each stream follows the structure described in Section II-A. According to the results shown in Figure 6a, the number of parameters and the time required for the training proportionally increase as the number of streams increases. On the contrary, the accuracy value does not follow this trend. In fact, while with 1, 2, and 3 streams the maximum value is comparable, with 4 streams the accuracy increases significantly. This behavior is due to the bigger kernel size in the latest streams, thus allowing the extraction of information from a bigger temporal interval.

2) DEPTH OF STREAMS

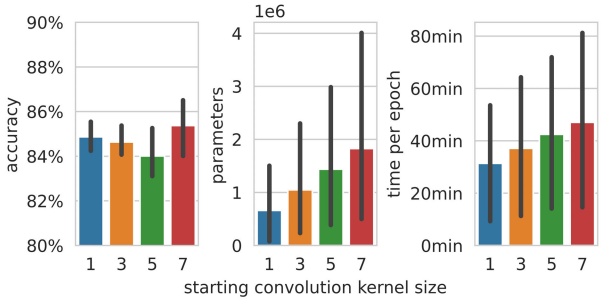
In this experimental phase, several convolutional blocks per stream have been tested. The number of blocks ranged from one to three. The structure of convolutional streams and blocks follows the description provided in Section II-A. According to the results shown in Figure 6b, the highest accuracy value is achieved by using two convolutional blocks.



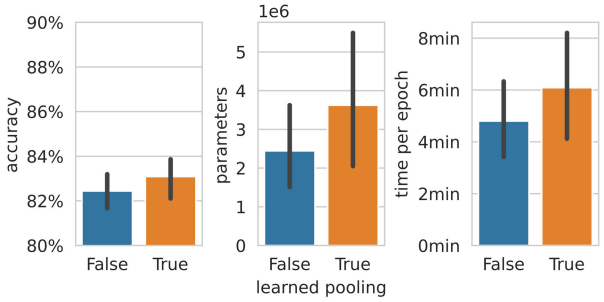
(a) Results about the number of streams in m_s.



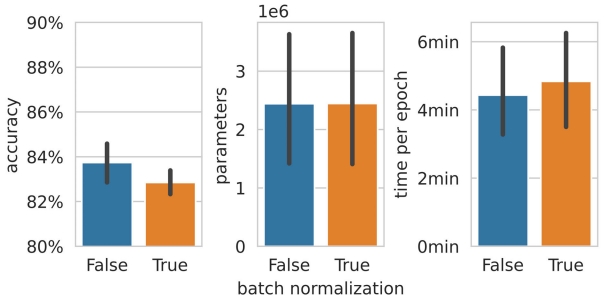
(b) Results about the number of convolution blocks in each stream.



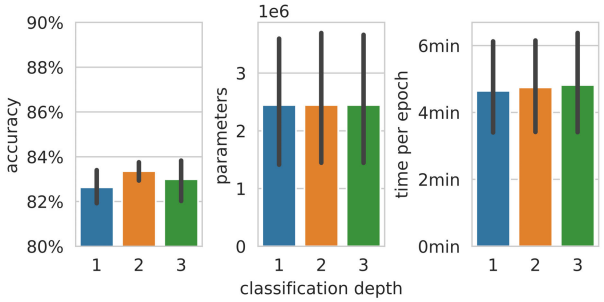
(c) Results about the starting kernel size (*i.e.* the value of k in m_{s_1}, which is progressively increased by 2 in the following streams).



(d) Results about the replacement of Max Pooling layers with 1D Convolutions at the end of the convolution blocks.



(e) Results about the presence of Batch Normalization between convolutions in convolutional blocks.



(f) Results about the number of linear layers in m_c.

FIGURE 6. Results of ablation studies on the structure of the proposed architecture. The features plotted are the accuracy on the PhysioNet dataset using 10-fold cross-validation, the number of parameters of the model, and the average minutes per training epoch.

This is amenable to the fact that a single block is not able to extract fine-grained features, while using three blocks, the signal results to be too much processed, thus leading to losing part of the information from the original signal. It is worth noting that the number of parameters and average training time are not proportional, as seen in Figure 6a. This is due to the fact that when a single convolutional block is used, the final convolution of the block having stride 2 is removed, where instead b inner convolutional blocks, i.e., all the convolutional blocks except the last one, have $b \cdot 3$ convolutions.

3) KERNEL SIZE

In this experimental phase, different kernel sizes for several streams have been tested. In detail, the kernel size of the first stream has been set, respectively, to a size of [3, 5, 7]. Then, the size of subsequent streams has been computed by following the process defined in Section II-A. As depicted in Figure 6c, this parameter does not have a significant impact on the final accuracy.

4) LEARNED POOLING

In this experimental phase, tests were conducted to determine which layer, between convolution and max pooling, must be placed at the end of the convolutional block to achieve the best results. As shown in Figure 6d, the learned convolution allows us to get better accuracy results at a slightly higher computational cost compared to the max pooling alternative, which is not a learnable layer. These findings are in line with the current state-of-the-art, as the trend in CNNs is to completely replace pooling operations [53].

5) BATCH NORMALIZATION

In this experimental phase, the effectiveness of batch normalization has been tested. In detail, we tested the batch normalization layer after the first two convolutional layers of each convolutional block. Batch normalization is a technique adopted to reduce the internal covariate shift of batches, thus speeding up the convergence and solving various initialization problems [54]. From the results shown in Figure 6e, normalizing batches reduces accuracy and slightly increases the training time. The decrease in accuracy is due to the smoothing effect of batch normalization on signals within the same batch, which results in the loss of several peaks in the signals and makes the samples too similar. Regarding the increasing training time, instead, it is simply due to the additional operation performed.

6) DEPTH OF CLASSIFICATION LAYERS

In this last experimental phase, the depth of the linear layer has been tested. Such depth consisted of a number of layers ranging from 1 to 3. As shown in Figure 6f, the highest accuracy value was achieved using two classification blocks, although the average training time and the number of parameters remained similar across all configurations.

V. LIMITATIONS AND FUTURE WORKS

The biggest limitation of the proposed method is the optimization. By having different streams, it is more difficult to optimize the proposed architecture for different reasons. Firstly, we have a higher number of hyperparameters that must be correctly tuned. In addition, the definition of the architecture itself can be non-trivial, e.g., number of streams, depth of the streams, streams of the same length or not, and more. In future research, we plan to further evaluate the performance of the proposed method by testing it with additional state-of-the-art datasets. While our current results demonstrate the effectiveness of our approach, expanding the evaluation to include other datasets will provide a more comprehensive understanding of its capabilities and robustness. By conducting experiments on additional datasets, we aim to assess the generalizability of our method across different domains and applications of EEG signals. Additionally, we intend to explore potential modifications and enhancements to our method, e.g., by considering attention mechanisms, to verify if it is possible to reduce the number of streams or the number of convolutional layers.

VI. CONCLUSION

This paper proposes a novel multi-stream 1D CNN model for classifying EEG signals related to the MI application field. The main novelty of the model is the use of the multi-stream strategy, where parallel kernels with different sizes on each stream are able to retrieve meaningful information from the same data at different time scales. Comparative experiments on the reference dataset, i.e., PhysioNet EEG motor movement/imagery data collection, for this type of application context, i.e., MI, show that the proposed method outperforms the current state-of-the-art in cross-subject and intra-subject experiments. Furthermore, our approach demonstrates a marked advancement over related works, primarily attributed to the utilization of a more extensive and diverse array of metrics. This comprehensive evaluation methodology enables a more nuanced and thorough understanding of model performance, thus ensuring the validity and reliability of our findings across varied contexts and applications. Finally, a consistent ablation study has been conducted to assess the performances of multiple feature extraction streams and their structure, which led to the best model configurations to surpass the previous results in the literature.

REFERENCES

[1] C. Loriette, C. Ziane, and S. Ben Hamed, "Neurofeedback for cognitive enhancement and intervention and brain plasticity," *Revue Neurologique*, vol. 177, no. 9, pp. 1133–1144, Nov. 2021.

[2] K. Tan, A. Daitch, P. Pinheiro-Chagas, K. Fox, J. Parvizi, and M. Lieberman, "Electrocorticographic evidence of a common neurocognitive sequence for mentalizing about the self and others," *Nature Commun.*, vol. 2022, no. 13, p. 1919, 2022.

[3] R. Wahnoun, M. Benson, S. Helms-Tillery, and P. D. Adelson, "Delineation of somatosensory finger areas using vibrotactile stimulation, an ECoG study," *Brain Behav.*, vol. 5, no. 10, pp. 1–10, Oct. 2015.

[4] L.G. Lim, W. C. Ung, Y. L. Chan, C.-K. Lu, S. Sutoko, T. Funane, M. Kiguchi, and T. B. Tang, "A unified analytical framework with multiple fNIRS features for mental workload assessment in the prefrontal cortex," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 11, pp. 2367–2376, Nov. 2020.

[5] M. Kiani, J. Andreu-Perez, H. Hagras, E. I. Papageorgiou, M. Prasad, and C.-T. Lin, "Effective brain connectivity for fNIRS with fuzzy cognitive maps in neuroergonomics," *IEEE Trans. Cognit. Develop. Syst.*, vol. 14, no. 1, pp. 50–63, Mar. 2022.

[6] S. B. Moro, M. Carrieri, D. Avola, S. Brigadoi, S. Lancia, A. Petracca, M. Spezialetti, M. Ferrari, G. Placidi, and V. Quaresima, "A novel semi-immersive virtual reality visuo-motor task activates ventrolateral prefrontal cortex: A functional near-infrared spectroscopy study," *J. Neural Eng.*, vol. 13, no. 3, Jun. 2016, Art. no. 036002.

[7] N. Ramsey, M. van de Heuvel, K. Kho, and F. Leijten, "Towards human BCI applications based on cognitive brain systems: An investigation of neural signals recorded from the dorsolateral prefrontal cortex," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 214–217, Jun. 2006.

[8] B. Ng, R. Abugharbieh, X. Huang, and M. J. McKeown, "Spatial characterization of fMRI activation maps using invariant 3-D moment descriptors," *IEEE Trans. Med. Imag.*, vol. 28, no. 2, pp. 261–268, Aug. 2009.

[9] K.-R. Müller and S. M. Hofmann, "Interpreting deep learning models for multi-modal neuroimaging," in *Proc. 11th Int. Winter Conf. Brain-Comput. Interface (BCI)*, Feb. 2023, pp. 1–4.

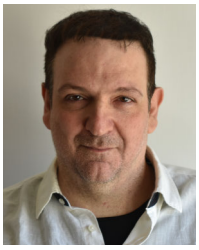
[10] S. H. Jin, S. H. Lee, G. H. Jang, Y. J. Lee, J. An, and Y.-S. Lee, "Cortical activation by fNIRS and fMRI during grasping in patient with traumatic brain injury: A case study," in *Proc. Int. Winter Workshop Brain-Comput. Interface (BCI)*, Feb. 2014, pp. 1–4.

[11] S. Leinders, M. J. Vansteensel, G. Piantoni, M. P. Branco, Z. V. Freudenburg, T. A. Gebbink, E. G. M. Pels, M. A. H. Raemaekers, A. Schippers, E. J. Aarnoutse, and N. F. Ramsey, "Using fMRI to localize target regions for implanted brain–computer interfaces in locked-in syndrome," *Clin. Neurophysiol.*, vol. 155, pp. 1–15, Nov. 2023.

[12] J. Zhang, Q. Wang, X. Wang, L. Qiao, and M. Liu, "Preserving specificity in federated graph learning for fMRI-based neurological disorder identification," *Neural Netw.*, vol. 169, pp. 584–596, Jan. 2024.

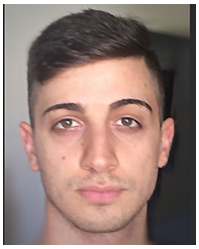
[13] P. Kumari and A. Vaish, "Information-theoretic measures on intrinsic mode function for the individual identification using EEG sensors," *IEEE Sensors J.*, vol. 15, no. 9, pp. 4950–4960, Sep. 2015.

[14] W. Chang, H. Wang, G. Yan, and C. Liu, "An EEG based familiar and unfamiliar person identification and classification system using feature extraction and directed functional brain network," *Expert Syst. Appl.*, vol. 158, Nov. 2020, Art. no. 113448.
[15] W. Li, Y. Yi, M. Wang, B. Peng, J. Zhu, and A. Song, "A novel tensorial scheme for EEG-based person identification," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–17, 2023.
[16] Y. Li, L. Wang, W. Zheng, Y. Zong, L. Qi, Z. Cui, T. Zhang, and T. Song, "A novel bi-hemispheric discrepancy model for EEG emotion recognition," *IEEE Trans. Cogn. Develop. Syst.*, vol. 13, no. 2, pp. 354–367, 2021.
[17] P. Samal and M. F. Hashmi, "Ensemble median empirical mode decomposition for emotion recognition using EEG signal," *IEEE Sensors Lett.*, vol. 7, no. 5, pp. 1–4, May 2023.
[18] Y. Yan, X. Wu, C. Li, Y. He, Z. Zhang, H. Li, A. Li, and L. Wang, "Topological EEG nonlinear dynamics analysis for emotion recognition," *IEEE Trans. Cognit. Develop. Syst.*, vol. 15, no. 2, pp. 625–638, May 2023.
[19] K. P. Thomas, A. P. Vinod, and C. Guan, "Enhancement of attention and cognitive skills using EEG based neurofeedback game," in *Proc. 6th Int. IEEE/EMBS Conf. Neural Eng. (NER)*, Nov. 2013, pp. 21–24.
[20] G. Placidi, D. Avola, A. Petracca, F. Sgallari, and M. Spezialetti, "Basis for the implementation of an EEG-based single-trial binary brain computer interface through the disgust produced by remembering unpleasant odors," *Neurocomputing*, vol. 160, pp. 308–318, Jul. 2015.
[21] A. Kamble, P. H. Ghare, V. Kumar, A. Kothari, and A. G. Keskar, "Spectral analysis of EEG signals for automatic imagined speech recognition," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–9, 2023.
[22] B. Sun, X. Zhao, H. Zhang, R. Bai, and T. Li, "EEG motor imagery classification with sparse spectrotemporal decomposition and deep learning," *IEEE Trans. Autom. Sci. Eng.*, vol. 18, no. 2, pp. 541–551, Apr. 2021.
[23] W. Zhang and D. Wu, "Lightweight source-free transfer for privacy-preserving motor imagery classification," *IEEE Trans. Cogn. Develop. Syst.*, vol. 15, no. 2, pp. 938–949, Jul. 2023.
[24] H. Zhi, Z. Yu, T. Yu, Z. Gu, and J. Yang, "A multi-domain convolutional neural network for EEG-based motor imagery decoding," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 3988–3998, 2023.
[25] X. Du, M. Kong, S. Qiu, J. Guo, and Y. Lv, "Recognition of motor imagery EEG signals based on capsule network," *IEEE Access*, vol. 11, pp. 31262–31271, 2023.
[26] C. Park, C. C. Took, and D. P. Mandic, "Augmented complex common spatial patterns for classification of noncircular EEG from motor imagery tasks," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 1, pp. 1–10, Jan. 2014.
[27] A. Loboda, A. Margineanu, G. Rotariu, and A. Mihaela, "Discrimination of EEG-based motor imagery tasks by means of a simple phase information method," *Int. J. Adv. Res. Artif. Intell.*, vol. 3, no. 10, pp. 1–5, 2014.
[28] M. Mahmoodi, B. Makkiabadi, M. Mahmoudi, and S. Sanei, "A new method for accurate detection of movement intention from single channel EEG for online BCI," *Comput. Methods Programs Biomed. Update*, vol. 1, Jan. 2021, Art. no. 100027.
[29] M. Tolić and F. Jovic, "Classification of wavelet transformed EEG signals with neural network for imagined mental and motor tasks," *Kinesiology*, vol. 45, no. 1, pp. 130–138, 2013.
[30] M. H. Alomari, A. Abubaker, A. Turani, A. M. Baniyounes, and A. Manasreh, "EEG mouse: A machine learning-based brain computer interface," *Int. J. Adv. Comput. Sci. Appl.*, vol. 5, no. 4, pp. 193–198, 2014.
[31] H. V. Shenoy, A. P. Vinod, and C. Guan, "Shrinkage estimator based regularization for EEG motor imagery classification," in *Proc. 10th Int. Conf. Inf., Commun. Signal Process. (ICICS)*, Dec. 2015, pp. 1–5.
[32] V.S. Handiru and V. A. Prasad, "Optimized bi-objective EEG channel selection and cross-subject generalization with brain–computer interfaces," *IEEE Trans. Human-Mach. Syst.*, vol. 46, no. 6, pp. 777–786, Dec. 2016.
[33] H. Varsehi and S. M. P. Firoozabadi, "An EEG channel selection method for motor imagery based brain–computer interface and neurofeedback using Granger causality," *Neural Netw.*, vol. 133, pp. 193–206, Jan. 2021.
[34] Y. Kim, J. Ryu, K. K. Kim, C. C. Took, D. P. Mandic, and C. Park, "Motor imagery classification using mu and beta rhythms of EEG with strong uncorrelating transform based complex common spatial patterns," *Comput. Intell. Neurosci.*, vol. 2016, no. 1, 2016, Art. no. 1489692.
[35] S. Abenna, M. Nahid, and A. Bajit, "Motor imagery based brain–computer interface: Improving the EEG classification using delta rhythm and LightGBM algorithm," *Biomed. Signal Process. Control*, vol. 71, Jan. 2022, Art. no. 103102.
[36] A. Keutayeva and B. Abibullaev, "Exploring the potential of attention mechanism-based deep learning for robust subject-independent motor-imagery based BCIs," *IEEE Access*, vol. 11, pp. 107562–107580, 2023.
[37] H. Dose, J. S. Møller, H. K. Iversen, and S. Puthusserypady, "An end-to-end deep learning approach to MI-EEG signal classification for BCIs," *Expert Syst. Appl.*, vol. 114, pp. 532–542, Dec. 2018.
[38] G. Zhang, V. Davoodnia, A. Sepas-Moghaddam, Y. Zhang, and A. Etemad, "Classification of hand movements from EEG using a deep attention-based LSTM network," *IEEE Sensors J.*, vol. 20, no. 6, pp. 3113–3122, Mar. 2020.
[39] T. Karácsony, J. P. Hansen, H. K. Iversen, and S. Puthusserypady, "Brain computer interface for neuro-rehabilitation with deep learning classification and virtual reality feedback," in *Proc. 10th Augmented Hum. Int. Conf. (AH)*, 2019, pp. 1–8.
[40] D. Zhang, K. Chen, D. Jian, and L. Yao, "Motor imagery classification via temporal attention cues of graph embedded EEG signals," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 9, pp. 2570–2579, Sep. 2020.
[41] W. Huang, W. Chang, G. Yan, Z. Yang, H. Luo, and H. Pei, "EEG-based motor imagery classification using convolutional neural networks with local reparameterization trick," *Expert Syst. Appl.*, vol. 187, Jan. 2022, Art. no. 115968.
[42] D. Avola, M. Cascio, L. Cinque, A. Fagioli, G. L. Foresti, M. R. Marini, and D. Pannone, "Analyzing EEG data with machine and deep learning: A benchmark," in *Proc. Int. Conf. Image Anal. Process. (ICIAP)*, 2022, pp. 335–345.
[43] T. Takase, "Dynamic batch size tuning based on stopping criterion for neural network training," *Neurocomputing*, vol. 429, pp. 1–11, Mar. 2021.
[44] B. McFee, J. Salamon, and J. P. Bello, "Adaptive pooling operators for weakly labeled sound event detection," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 26, no. 11, pp. 2180–2193, Nov. 2018.
[45] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
[46] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. 215–220, Jun. 2000.
[47] G. Schalk, D. McFarland, T. Hinterberger, N. Birbaumer, and J. Wolpaw, "BCI2000: A general-purpose brain–computer interface (BCI) system," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1034–1043, May 2004.
[48] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen, and M. S. Hämäläinen, "MEG and EEG data analysis with MNE-Python," *Frontiers Neurosci.*, vol. 7, no. 267, p. 70133, 2013.
[49] H. Dalianis, *Evaluation Metrics and Evaluation*. Cham, Switzerland: Springer, 2018, pp. 45–53.
[50] M. L. McHugh, "Interrater reliability: The Kappa statistic," *Biochemia Medica*, vol. 22, no. 3, pp. 276–282, 2012.
[51] E. P. Torres, E. A. Torres, M. Hernández-Álvarez, and S. G. Yoo, "EEG-based BCI emotion recognition: A survey," *Sensors*, vol. 20, no. 18, p. 5083, Sep. 2020.
[52] T. Harmony, T. Fernández, J. Silva, J. Bernal, L.Díaz-Comas, A. Reyes, E. Marosi, M. Rodríguez, and M. Rodríguez, "EEG delta activity: An indicator of attention to internal processing during performance of mental tasks," *Int. J. Psychophysiol.*, vol. 24, nos. 1–2, pp. 161–171, Nov. 1996.
[53] M. Sun, Z. Song, X. Jiang, J. Pan, and Y. Pang, "Learning pooling for convolutional neural network," *Neurocomputing*, vol. 224, pp. 96–104, Feb. 2017.
[54] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn. (ICML)*, vol. 37, 2015, pp. 448–456.



DANILO AVOLA (Member, IEEE) received the M.Sc. degree in computer science from the Sapienza University of Rome, Italy, in 2002, and the Ph.D. degree in molecular and ultrastructural imaging from the University of L'Aquila, Italy, in 2014. Since 2024, he has been an Associate Professor with the Department of Computer Science, Sapienza University of Rome. He co-founded and currently leads the Prometheus Laboratory and is the Co-Founder of 4AI, a university startup focused on pioneering new methodologies in artificial intelligence. Previously, as an Assistant Professor, he was the Research and Development Scientific Director of the Computer Vision Laboratory (VisionLab), Sapienza University of Rome. As the Principal Investigator, he currently directs several strategic research initiatives at the Sapienza University of Rome, including Wi-Fi sensing for person re-identification and human synthesis, emotion transference in humanoids via EEG, UAV navigation by view, and LieToMe systems. His research interests include artificial intelligence (including machine learning and deep learning), computer vision, Wi-Fi sensing, EEG signal analysis, human–computer interaction, human-behavior recognition, human-action recognition, biometric analysis, bioinformatics, optimized neural architectures, deception detection, VR/AR systems, drones, and robotics. He is an Active Member of several professional organizations, including IAPR, CVPL, IEEE, ACM, AIxIA, and EurAI.

LUIGI CINQUE (Senior Member, IEEE) received the M.Sc. degree in physics from the University of Napoli, Italy. From 1984 to 1990, he was with the Laboratory of Artificial Intelligence (Alenia S.p.A), involved on the development of expert systems and knowledge-based vision systems. He is currently a Full Professor of computer science with the Sapienza University of Rome, Rome, Italy. He is the author of more than 200 papers in national and international journals and conference proceedings. His current research interests include distributed systems for analysis and interpretation of video sequences, target tracking, multisensor data, and information fusion. Some of the techniques he has proposed have found applications in the field of video-based surveillance systems, autonomous vehicle, road traffic control, human behavior understanding, and visual inspection. He is a member of ACM, IAPR, and CVPL. He also serves as a reviewer for many international journals. He served on scientific committees of international conferences (e.g., CVPR, ICME, and ICPR) and symposia. He serves as a Reviewer for European Union in different research program.



ANGELO DI MAMBRO received the B.Sc. and M.Sc. (cum laude) degrees in computer science from the Sapienza University of Rome, Italy, in 2018 and 2021, respectively, where he is currently pursuing the Ph.D. degree in computer science with the Prometheus Laboratory, Department of Computer Science. He is a member of the Prometheus Laboratory, Department of Computer Science, Sapienza University of Rome. His research interests include deep learning, EEG biometrics, EEG emotion recognition, EEG synthesis, signal processing, human–computer interaction, computer vision, AR/VR, bioinformatics, optimized neural architectures, and robotics.

ROMEO LANZINO (Graduate Student Member, IEEE) received the degree (cum laude) in computer science from the Sapienza University of Rome, in 2021, where he is currently pursuing the Ph.D. degree in artificial intelligence. He is a part of the VisionLab, Computer Science Department. His research interests include deep learning, bioinformatics, computer vision, digital signal processing, and human–computer interaction.

DANIELE PANNONE (Member, IEEE) received the M.Sc. and Ph.D. degrees in computer science from the Sapienza University of Rome, in 2015 and 2018, respectively. He was a member of the Computer Vision Laboratory (VisionLab), Department of Computer Science, from 2015 to 2024. As of 2020, he holds the position of an Assistant Professor with the VisionLab, Department of Computer Science, and in 2024, he has Co-Founded the Prometheus Laboratory. In the same year, he has co-founded 4AI, a Sapienza University of Rome startup focused on pioneering new methodologies in artificial intelligence. His research interests include machine and deep learning, event recognition, object detection, person re-identification, signal analysis and processing, bioinformatics, human–computer interaction, AR/VR, and robotics. He is also a member of several professional organizations, including IAPR, CVPL, IEEE, ACM, AIxIA, and EurAI.

FRANCESCO SCARCELLO received the Ph.D. degree in computer science from the University of Calabria, in 1997. He is currently a Full Professor of computer science (SSD ING-INF/05) with the University of Calabria. His research interests include computational complexity, game theory, (hyper) graph theory, constraint satisfaction, logic programming, knowledge representation, nonmonotonic reasoning, and database theory. He has extensively published in all these areas in leading conferences and journals. He is a fellow of European Association for Artificial Intelligence. His article ''Pure Nash Equilibria: Hard and Easy Games'' received the 2008 IJCAI-JAIR Best Paper Prize, awarded to an outstanding paper published in the *Journal of Artificial Intelligence Research* in the preceding five years. His article ''Hypertree Decompositions and Tractable Queries'' received the 2009 ACM PODS Alberto O. Mendelzon Test-of-Time Award, awarded every year to an article published in the proceedings of the ACM Symposium on Principles of Database Systems (PODS) ten years prior that had the most impact in terms of research, methodology, or transfer to practice over the intervening decade. In 2016, his article ''Hypertree Decompositions: Questions and Answers'' was invited for the special gems of PODS session, at the 35th ACM Symposium on Principles of Database Systems.

Open Access funding provided by 'Università degli Studi di Roma ''La Sapienza'' 2' within the CRUI CARE Agreement