IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received May 13, 2021, accepted June 13, 2021, date of publication June 21, 2021, date of current version July 16, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3091399

Shallow Convolutional Network Excel for Classifying Motor Imagery EEG in BCI Applications

DAILY MILANÉS HERMOSILLA[1], RAFAEL TRUJILLO CODORNIÚ[1,2], RENÉ LÓPEZ BARACALDO[3], ROBERTO SAGARÓ ZAMORA[4], DENIS DELISLE RODRIGUEZ[5], YOLANDA LLOSAS ALBUERNE[6], (Member, IEEE), AND JOSÉ RICARDO NÚÑEZ ÁLVAREZ[7], (Member, IEEE)

[1]Department of Automatic Engineering, Universidad de Oriente, Santiago de Cuba 90500, Cuba
[2]SERCONI and Department of Automatic Engineering, Universidad de Oriente, Santiago de Cuba 90500, Cuba
[3]Zimtronic, Miami, FL 33222, USA
[4]Department of Mechanical Engineering, Universidad de Oriente, Santiago de Cuba 90500, Cuba
[5]Postgraduate Program in Electrical Engineering, Federal University of Espirito Santo, Vitoria 29075-910, Brazil
[6]Department of Electricity, Universidad Técnica de Manabí, Portoviejo 130801, Ecuador
[7]Department of Energy, Universidad de la Costa, Barranquilla, Atlántico 080002, Colombia

Corresponding author: José Ricardo Núñez Álvarez (jnunez22@cuc.edu.co)

This research was conducted with the support of the Cuban Centre for Academic Supercomputing (HPC-Cuba) supported by VLIR-UOS JOINT project and the Iberoamerican Supercomputing Network (RICAP). The authors would like to thank the providers of the BCI Competition IV Datasets 2a and 2b and the BCI Competition III Dataset IVa.

ABSTRACT Many studies applying Brain-Computer Interfaces (BCIs) based on Motor Imagery (MI) tasks for rehabilitation have demonstrated the important role of detecting the Event-Related Desynchronization (ERD) to recognize the user's motor intention. Nowadays, the development of MI-based BCI approaches without or with very few calibration stages session-by-session for different days or weeks is still an open and emergent scope. In this work, a new scheme is proposed by applying Convolutional Neural Networks (CNN) for MI classification, using an end-to-end Shallow architecture that contains two convolutional layers for temporal and spatial feature extraction. We hypothesize that a BCI designed for capturing event-related desynchronization/synchronization (ERD/ERS) at the CNN input, with an adequate network design, may enhance the MI classification with fewer calibration stages. The proposed system using the same architecture was tested on three public datasets through multiple experiments, including both subject-specific and non-subject-specific training. Comparable and also superior results with respect to the state-of-the-art were obtained. On subjects whose EEG data were never used in the training process, our scheme also achieved promising results with respect to existing non-subject-specific BCIs, which shows greater progress in facilitating clinical applications.

INDEX TERMS Brain-computer interface, EEG, motor imagery, shallow convolutional neural networks.

I. INTRODUCTION

Electroencephalography (EEG) is a popular non-invasive technique, widely used for recording brain information. The raw EEG has a low signal-to-noise ratio due to its small amplitude peak-to-peak, and includes a variety of rhythms identified by their frequency range, location, and other aspects related to the brain function [1], [2]. This makes the EEG analysis extremely complex, although advantages of this technique, such as high temporal resolution, low cost, effectiveness and portability have widely motivated its application in brain-computer interfaces (BCIs). BCIs provide an alternative communication channel to patients with severe neuromotor disabilities, by capturing, processing, and translating the neural activity into control commands [3]–[5]. These systems have shown enormous potential to facilitate communication [6]–[8] and restore brain functions [9]–[11]. Various types of BCIs have been developed by researchers, using different paradigms, such as P300 [12], [13], steady-state visual evoked potentials (SSVEP) [14], [15] and motor

The associate editor coordinating the review of this manuscript and approving it for publication was Alessandra Bertoldo.

imagery (MI) [16]–[18]. Many works have been focused on MI-based BCIs due to the effectiveness of these systems for recovery of motor skills in post-stroke patients, as similar brain regions are activated over the primary sensorimotor areas when the same motor task is executed by real or imagined movements [19]. For instance, some studies [20]–[22] found that when a person executes a real or imagined unilateral movement, is attenuated or enhanced the amplitude of mu (from 8 to 12 Hz) and beta (from 13 to 30 Hz) rhythms over the primary motor cortex on both contralateral and ipsilateral hemispheres, respectively, phenomena known as event-related desynchronization/synchronization (ERD/ERS). The ERD phenomenon has been widely used to build BCIs [23], [24], [20], [25], but sometimes it may not be discernible for some subjects, producing consequently poor performance for MI classification, as described in [26]–[29]. In contrast, other studies [30], [22] detected the ERS for all subjects, which showed high specificity to discriminate movements. For instance, Pfurtscheller *et al.* [29] reported that ERD/ERS may be used for BCIs operated by MI, enhancing the hand MI discrimination when both ERD and ERS patterns are induced by executing at least one or two tasks. Furthermore, BCIs based on ERD/ERS were also proposed in [31]–[33].

Many techniques have been proposed for MI classification through EEG signals by applying hand-crafted feature extraction as a traditional method [34], [35] into the time domain, frequency domain, and spatiotemporal representation. One of the most popular and powerful methods for feature extraction is the Common Spatial Patterns (CSP) [36], [37], which has been extended in other approaches, such as Filter-Bank CSP (FBCSP) [38], achieving good accuracy on the BCI Competition IV datasets 2a and 2b. Furthermore, Linear Discriminant Analysis (LDA), Artificial Neural Networks (ANN), Support Vector Machines (SVM), and other traditional classifiers have been commonly used to discriminate hand-crafted feature vectors. Although these traditional approaches have been successfully applied to recognize MI, it is still a challenge to design accurate BCIs, leaving the door open for further improvements in this research area.

The recent success of deep learning methods has marked the state-of-the-art in many areas, such as computer vision, natural language processing, and speech recognition [39]. Deep learning has been benefited from the development of graphics processing units (GPUs) and the availability of large datasets, enabling the feature extraction and classification at a high level in contrast to conventional techniques. Deep learning does not require previous knowledge to analyze a given dataset, which represents a great advantage for processing data of high variability intra and inter subjects, such as EEG signals. For this reason, deep learning for MI-based BCIs has also increased the interest of researchers. For instance, Convolutional Neural Network (CNN) has been widely used with success for MI recognition through EEG signals [40], [41] due to its capability of learning high-level features, including the hierarchical structures in end-to-end data, and the automatic parameter optimization. As a result, some approaches employ deep learning for MI task recognition using the EEG signal as input [23]–[46], while others transform EEG channels by applying traditional methods for feature extraction, such as spectrograms and scalograms [47]–[53].

Despite the growing interest in new EEG-based BCI developments, there are very few public MI datasets available with limited amount of EEG data, and numbers of healthy subjects and patients. As a consequence, the training process to obtain best deep learning models from these small EEG datasets for successful class discrimination is still a challenging task, as these methods need to adjust many parameters (up to thousands), which may require a large training set. Evidence [23], [44] showed that the classification accuracy does not always improve when more convolutional layers are added in CNN for small sets. This undesirable effect may be associated to deeper architectures having high numbers of adjustable weights, which present high risk of overfitting. For this reason, much research applying CNN on small MI sets has proposed as an alternative shallow CNN architectures [23], [43], [44], [49], [54]. In fact, studies using shallow CNN improved the accuracy of MI classification by applying techniques based on model fusion and transfer learning [42], [43], [45], [46].

This current work proposes a neural network scheme using CNNs, inspired by the ShallowConNet (SCN) model in [23]. Our approach aims to improve the MI classification, taking into account that ERS occurs after ERD, this ERS generated over 2.5 s after beginning MI tasks, as described in [26], [27]. Our system was evaluated on three public datasets: the BCI Competition IV datasets 2a and 2b, and the BCI Competition III dataset IVa, obtaining high accuracy with respect to previous works. In particular, we tested the proposed system for non-subject-specific MI classification, achieving an average improvement of 18% on the database 2a with respect to the state-of-the-art. As a highlight, we extended our calibrated system to recognize MI tasks from new subjects of these datasets (dataset 2a and 2b) by applying inter-subject transfer learning strategy, obtaining comparable promising results with respect to the state-of-the-art, and also outperforming others studies. Our method demonstrated a great generalization when it was also applied on a small dataset (dataset IVa), outperforming various studies.

II. RELATED WORKS

The public BCI Competition IV datasets which contain EEG signals acquired from healthy subjects when performing motor tasks [55] have been widely used by BCI designers. For instance, the BCI Competition IV dataset 2a has been widely used to evaluate deep learning approaches, allowing multi-class MI classification (left hand, right hand, feet, and tongue) [56]. Schirrmeister *et al.* [23] explored through EEG three CNN architectures, such as ShallowConvNet (SCN), DeepConvNet, and also these two methods combined to classify MI, obtaining accuracy (ACC) of 73.7%

for SCN and 70.9% for DeepConvNet. Abbas and Khan [52] used CNN after applying CSP and Fast Fourier Transform Energy Map (FFTEM) for features extraction, achieving ACC of 70.75%.

Sakhavi and Guan [24] used FBCSP and the EEG envelope for feature extraction, followed by several CNN architectures, reaching highest ACC (up to 74.46%) with the C2CN model. In this approach, the kernel size used on the convolution layers was different for each subject, which add generalization for new individuals difficult. In [51] a hybrid model using CNN with Bidirectional Gated Recurrent Unit (BGRU) was built, obtaining ACC of 76.62%. The training process of this approach also included EEG data from the testing set, which may limit the generalization. In [43] three CNN models, such as multiple local spatial convolutions, global spatial convolution, and the combination of these two CNN models were proposed, attaining ACC of 71.8%, 74.6%, and 73.2%, respectively. It is worth mentioning that these aforementioned approaches have been trained and evaluated, considering only the EEG data from the same subject, therefore these systems are subject-specific.

Roy *et al.* [41] suggest two ways for training a BCI, such as using EEG data from only one subject (subject-specific BCI) or from several subjects (non-subject-specific). They remark that although best performance is obtained for a subject-specific BCI, a non-subject-specific BCI may be generalized to new subjects for whom non-training data exist, offering to specialists and patients a less time-consuming system for clinical applications. In [44] a BCI based on subject-specific training through multi-level convolutional features and fusion using convolutional networks was proposed, obtaining ACC of 74.5%. In addition, these same authors also proposed the fusion of convolutional models [45], developing two networks based on the AlexNet model also termed Multi-layer CNNs (MCNN), and CNN. These subject-specific BCIs reached ACC of 75.42% and 73.77%, respectively. Moreover, they applied for both approaches the transfer learning technique to obtain a non-subject-specific BCI, and achieved ACC of 42.09% and 55.34%, respectively.

On the other hand, the BCI Competition IV dataset 2b composed of two MI tasks from left and right hands have been widely used. Tabar *et al.* [57] proposed a CNN architecture with stacked autoencoders (CNN-SAE) using as input the EEG in time, frequency and space domain. The authors conducted first a cross-validation over all trials getting ACC of 77.4%, and after evaluated the recognition session-by-session, obtaining ACC of 75.1%. Dai *et al.* [48] combined a CNN architecture with a variational autoencoder (CNN-VAE), reaching ACC of 78.2%. Xu *et al.* [58] proposed a transfer CNN framework based on VGG-16 architecture (VGG-16 CNN), reporting ACC of 74.5%. Li *et al.* [59] employed the continuous wavelet transform (CWT) and a simplified convolutional neural network (SCNN), achieving ACC of 83.2% for a subject-specific BCI. Roy *et al.* [60] transformed the EEG epochs into the time-frequency representation by applying short-time Fourier transform (STFT), which were applied in sequence as images to CNN architectures with ACC of 77.46% for an intra-subject cross-session validation, and ACC of 70.94% when cross-subject transfer learning were obtained. Sun *et al.* [61] extracted dominant spectral EEG features by applying the spectrotemporal decomposition method (SSD-SE-CNN), achieving ACC of 79.3%. Lee *et al.* employed both CWT and 1D CNN, using three mother wavelets for evaluation, getting the best result (ACC of 83%) by applying Morlet.

The BCI Competition III dataset IVa, composed of two MI tasks linked to the right hand and foot has been widely used. This dataset has a small amount of training data, increasing the challenge for BCI reseachers. Furthermore, the EEG recording in both training and testing sets has different sizes between subjects [62]. Many studies used the training set to also validate the performance of their methods. Selim *et al.* [63] employed a subject-specific optimal time interval for CSP feature extraction, and further used hybrid bio-inspired algorithms for feature selection and to optimize the classification, obtaining ACC of 85%. Park and Chung [64] developed a channel optimal selection method to extract more discriminative CSP features for a SVM classifier, achieving ACC of 88.62%. Guo *et al.* [65] proposed a filter band component regularized common spatial patterns (FCCSP) to extract features for a LDA classifier, getting ACC of 82.01%. Singh *et al.* [66] proposed a method based on spatially regularized symmetric positive definite (SPD) matrices and minimum distance of Riemannian mean (MDRM), achieving mean ACC of 86.13%. Jin *et al.* [67] proposed a bispectrum-based channel selection (BCS) method, obtaining ACC of 86.3%. Dai *et al.* [68] developed the Transfer Kernel CSP to learn a domain-invariant kernel, achieving ACC of 81.14%. Huang *et al.* [69] employed CWT, Tensor Discriminant Analysis (TDA) and CNN, reaching ACC of 86.02%. Other studies were conducted on the full dataset to test the generalization performance. For instance, Miao *et al.* [70] converted the raw EEG into image representation by computing its energy for different frequency bands, and further used a multilayer CNN, obtaining mean ACC of 90%. Attallah *et al.* [71] proposed a hybrid system to calculate a feature set through different combinations of channels, obtaining the best performance (mean ACC of 93.46%) for 18 channels by using SVM.

III. MATERIALS AND METHODS

A. DATASET DESCRIPTION

Three public datasets were employed in this current study to evaluate the performance of our proposed system, such as the dataset 2a and dataset 2b from BCI Competition IV, and dataset IVa from BCI competition III.

1) THE BCI COMPETITION IV DATASET 2a

This dataset is composed of several trials associated to four motor imagery tasks corresponding to left hand, right hand,

both feet, and tongue [55], [56]. The EEG signals were recorded on 22 locations (Fz, FC1, FC3, FCz, FC2, FC4, C5, C3, C1, Cz, C2, C4, C6, CP3, CP1, CPz, CP2, CP4, P1, Pz, P2 and POz), using sampling rate at 250 Hz, a band-pass filter with frequency range from 0.5 and 100 Hz, and a notch filter enabled at 50 Hz to suppress the power line noise. Each subject completed two sessions on different days for a total of 576 trials (144 trials per class), each session containing 288 trials. The first session was used here for training, and the other session was employed for testing.

2) THE BCI COMPETITION IV DATASET 2b

This dataset contains two imagery tasks linked to the left hand and the right hand [55], [72]. Three bipolar EEG channels over C3, Cz and C4 locations were recorded with sampling rate at 250 Hz, using a band-pass filter with frequency range from 0.5 and 100 Hz, and a notch filter enabled at 50 Hz to suppress the power line noise. Each subject completed a total of five sessions, receiving online smiley feedback in the last three sessions. In the first two sessions, each subject performed 120 trials per session, and executed another 160 trials per session in the last three sessions, completing a total of 720 trials. The first three sessions were used here for training, and the last two sessions for testing.

3) THE BCI COMPETITION III DATASET IVa

This dataset contains EEG signals from five healthy subjects [62], which were recorded on 118 locations with sampling rate at 1000 Hz, while they performed two motor imagery tasks (right hand and right foot). Each participant completed a total of 280 trials to form the training and testing sets. In particular, these sets have different numbers of trials per subject in contrast to the aforementioned dataset 2a and 2b. A total of 168, 224, 84, 56 and 28 trials compose the training set for subjects aa, al, av, aw and ay, respectively. The remaining trials were selected to conform the testing set. We used the public EEG data available with downsampling at 100 Hz, which in our study were after upsampled at 250 Hz.

B. PROPOSED METHOD

1) PREPROCESSING AND DATA AUGMENTATION

The EEG signals are band-pass filtered in a frequency range from 4 to 38 Hz [23], [43]–[46] through a Butterworth filter [73], aiming to preserve the ERD and ERS rhythms, and also reject noise and undesirable physiological and non-physiological artifacts. Afterwards, each filtered EEG trial $\mathbf{x}$ is standardized as $(\mathbf{x}(t_i) - \mu(t_i))/\sigma(t_i)$ by applying the electrode-wise exponential moving standardization with a decay factor of 0.999, which computes both mean $\mu(t_i)$ and variance $\sigma^2(t_i)$ values taken at sample t_i [23]. The starting values of $\mu(t_i)$ and $\sigma^2(t_i)$ are calculated over periods corresponding to the rest state preceding each trial. In order to rectify outliers, the EEG amplitudes of each trial are first limited to $\mu(t_i) \pm 6\sigma(t_i)$.

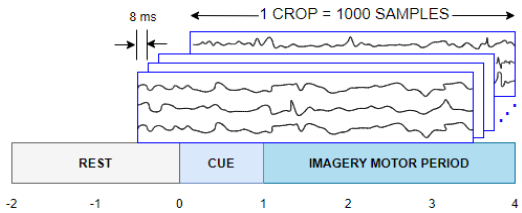


FIGURE 1. Data augmentation using trial crops strategy.

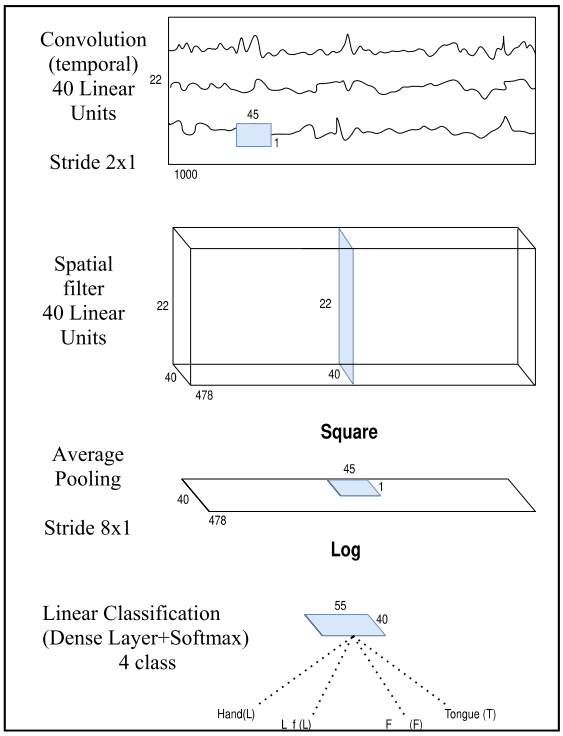


FIGURE 2. Architecture of proposed shallow convolutional network on dataset 2a.

Data augmentation has demonstrated considerably enhanced performance of deep learning approaches, reducing overfitting. Here, the trial crops strategy is adopted to increase the training set. For instance, we consider crops of 4 s each 8 ms in the interval from -0.5 to 4 s (cue onset at 0 s) over trials of the datasets 2a and 2b, while crops of 3 s each 8 ms from 0 to 3.5 s (cue onset at 0 s) were used on trials of the dataset IVa. The defined window sizes of 4 s (a total of 1000 samples) and 3 s (750 samples) aim to capture the sequence of ERD and ERS in each trial of these three datasets. Figure 1 shows for a trial the crops extraction. A similar methodology was done by Schirrmeister *et al.* [23].

2) NETWORK ARCHITECTURE

The architecture of the proposed model is based on SCN [23], which has been employed as a first stage by various researchers [43], [45], [46], [49], [54]. This SCN model, copes well with the reduced size of available datasets.

The proposed architecture for dataset 2a is shown in Figure 2. As a result, the scheme contains 2 convolutional layers and one dense classification layer. The first convolution

TABLE 1. Main differences between SCN and the proposed architecture.

Layer	SCN [23]	PROPOSED
Input	500x22x1	1000x22x1
Temporal Convolutional	40 units 25x1	40 units 45x1 Stride 2x1
Mean Pooling	75x1 Stride 15x1	45x1 Stride 8x1
Linear Classification (Dense Layer+Softmax)	30x40 to 4 class	55x40 to 4 class

has as input a $1000 \times 22 \times 1$ tensor and applies a temporal convolution with a 45×1 filter and 40 channels so that a $478 \times 22 \times 40$ tensor is obtained at the output, after performing downsampling from 250 Hz to 125 Hz, with a stride of 2. The kernel size is adjusted by applying optimization, covering periods of approximately 200 ms, which gives a greater variety of transformations. It is worth noting that our approach allows coverage of almost twice the period covered by the first convolution proposed in [23]. The second layer is a spatial convolution layer composed of 40 channels and a 1×22 filter, as done in [23]. Together, these two layers perform transformations according to [38], [23]. In sequence, other stages such as a "Square" activation function, a mean pooling layer with a downsampling, and a logarithmic activation function are added to calculate the logarithmic power average through 45×1 sliding windows with 8×1 stride.

All extracted features are then analyzed in the classification layer, composed of a dense layer with the Softmax activation function. This classification layer receives a total of 2200 features, which may add challenges when training the model, avoiding overfitting. To minimize this risk, batch normalization layers are added to normalize the outputs of previous layers with mean and variance equal to 0 and 1, respectively [74]. Moreover, a dropout layer is added, which randomly deactivates a percentage p of neurons for each training iteration [75], [76]. The "MaxNorm" regularization is also used in the convolution layers and the dense layer to limit the magnitudes of all weights to maximum values. The Adam optimizer [77], and Categorical Cross-Entropy as cost function are used here. Furthermore, the "Early Stopping" method and decay of learning rate are also used, aiming to further prevent overfitting adjusting parameters.

Table 1 shows the main differences between SCN and the proposed architecture. The architecture developed here is extended to the three datasets, but modifying the input and output layers because of the number of the EEG channels and window size for processing are different. Then, for the dataset 2b and the datatset IVa an input of $1000 \times 3 \times 1$ tensor and $750 \times 118 \times 1$ tensor is used, respectively. The output layer used for both dataset 2b and IVa have only two outputs.

C. EVALUATION AND STATISTICAL ANALYSIS

In order to validate the efficiency of our proposed recognition system, we used three public datasets, such as the datasets 2a and 2b from BCI Competition IV, and the dataset IVa from BCI Competition III. The results were compared with the state-of-the-art. Here, four different experiments were carried out, adopting subject-specific training for the first two experiments, and non-subject-specific training for the last two. As the dataset IVa is not structured by sessions, we only used it in Experiment #1. The other datasets 2a and 2b are composed of sessions, so, they were used in all four experiments.

It is worth mentioning that a balanced training set was used here, in order to build a better class recognition model.

Experiment #1. Similarly to [38] we first adjusted the parameters of our proposed architecture running a 9×10-Kfold (9 folds and 10 repetitions) stratified cross-validation over the training set corresponding to the dataset 2a. Notice that 9 folds were taken instead of 10 as in [38], to facilitate class balance. This experiment conducted individually on each subject allowed to adjust several parameters, such as the input window size of the network, the number of channels and the filter size in the convolutional layers, the characteristics of the Average Polling layer, and other parameters, such as the dropout level and the limits of weights established by MaxNorm regularization. Next, to verify the feasibility of our approach for generalization, evaluations and experiments were carried out employing these adjusted parameters.

These adjusted parameters were then extended in this current experiment for further evaluation over EEG data of the dataset 2b and dataset IVa. A 10×10-Kfold (10 folds and 10 repetitions) was only carried out on the training set of the dataset 2b, whereas a 9×10-KFold (9 folds and 10 repetitions) was exclusively implemented on the training set of the dataset IVa. Additionally, a last evaluation was conducted on the dataset IVa, considering now the training set and testing set, which consists of 280 trials. A Nested KFold cross-validation procedure was implemented on all trials, aiming to estimate an unbiased generalization performance [78]. In this case, a 10×9 (10 folds cross-validation in the outer loop, and 9 fold cross-validation in the inner loop) was realized.

Experiment #2. We applied an intra-subject session to session classification (subject-specific BCI), taking only the training set of each database for training while the testing set was used only for evaluation. Notice that this experiment provides as output 9 pretrained networks, one per subject. The first session from the dataset 2a composed of 288 trials per subject was split randomly into two new sets, one for training (240 trials representing the 5/6 part) and other for validation (48 trials representing the 1/6 part). Afterwards, the second session of this dataset, composed of another 288 trials, was finally employed by the trained system to test the generalization on new data of each subject.

A similar procedure was carried out using dataset 2b, where the first three sessions with 400 trials were divided randomly into two new sets, one for training (320 trials representing the 4/5 part) and other for validation (80 trials).

The last two sessions of this dataset (320 trials) were finally employed by the trained system to test the generalization.

To evaluate the model's generalization, a repeated holdout validation on a fixed testing set is performed, using different random seed values. The procedure is repeated 16 times to ensure that the probability of a trial to be selected at least one time in a validation set is approximately 0.95.

Experiment #3. The training process was performed, using the training set of all subjects (non-subject-specific BCI). Notice that this experiment provides as output a unique pre-trained network from all subjects. The first session from all subjects in dataset 2a (288 trials per subject) was used to form the training set with a total of 2592 trials. Then, this training set was randomly divided into new sets, one for training (2160 trials representing the 5/6 part) and the other for validation (432 trials representing the 1/6 part). Finally, the second session (288 trials per subject) in this dataset was employed to test the pretrained network and prove its generalization on new data of each subject.

A similar procedure was conducted with dataset 2b, using the three first sessions of all subjects to compose a training set with a total of 3600 trials. This training set, was randomly divided into two new sets, one for training (2880 trials) and the other for validation (720 trials). The trained system was finally evaluated on the last two sessions (320 trials per subject) to test the generalization.

The process of repeated holdout validation with a fixed testing set was performed 16 times for different random seeds to evaluate the model's generalization.

Experiment #4. We evaluated an inter-subject transfer learning strategy [60]. In contrast to Experiments #2 and #3, the accuracy for a specific subject in dataset 2a (or dataset 2b) was obtained here by applying the pretrained model over his/her full dataset. Notice that the pretrained model was obtained from the remaining subjects on the dataset of this specific subject (dataset 2a or dataset 2b), using the EEG data of all sessions. For example, the accuracy for Subject A01 was analyzed using the pretrained model obtained from all EEG data of Subject A02 to A09. Thus, the training set has a total of 4608 trials when using dataset 2a, and 5760 trials when using dataset 2b. A randomly selected portion of these trials (3840 or 4608 trials representing the 5/6 or 4/5 part, respectively), were then used to train our system, and the remaining trials (768 or 1152 trials, respectively) were used for validation. Finally, a repetead holdout validation (16 times selecting different random seeds) was performed to evaluate the model's generalization on the full dataset of the unknown subject with a total of 576 or 720 trials, respectively. The mean accuracy for each subject was obtained by carrying out the same procedure described for Subject A01.

To obtain the statistical significance, the significance level was set at 0.05. When the number of observations is greater than 30, a normal distribution may be considered for statistical comparison. Here, the Z test was applied to verify if there is statistical significance between our approach and other related studies. When the number of observations was less than 30, the Wilcoxon test for a sample [79] was used.

TABLE 2. K-fold cross-validation performance in terms of kappa value mean and standard deviation on dataset 2a.

Subjects	CSP [73]	FBCSP [38]	PROPOSED (9x10)
A01	0.64 ± 0.06	0.77 ± 0.07	**0.86 ± 0.06***
A02	0.42 ± 0.06	0.48 ± 0.06	**0.56 ± 0.07***
A03	0.80 ± 0.07	0.83 ± 0.07	**0.93 ± 0.05***
A04	0.37 ± 0.05	0.48 ± 0.06	**0.70 ± 0.08***
A05	0.22 ± 0.05	0.60 ± 0.06	**0.83 ± 0.06***
A06	0.28 ± 0.05	0.35 ± 0.05	**0.51 ± 0.11***
A07	0.63 ± 0.06	0.86 ± 0.07	**0.92 ± 0.05***
A08	0.77 ± 0.07	0.81 ± 0.07	**0.85 ± 0.06***
A09	0.72 ± 0.07	0.79 ± 0.07	**0.86 ± 0.06***
Average	0.54 ± 0.06	0.66 ± 0.07	**0.78 ± 0.02***

*, indicates significant difference with a 95 % confidence level

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. EXPERIMENT 1

Table 2 shows the results obtained on dataset 2a, with our approach and other related works [73], [38]. The results show that our proposed method using periods of 4 s yielded the best performance with mean Kappa of 0.78, enhancing the accuracy for all subjects with respect to previous studies applying powerful tools, such as CSP [73] and FBCSP [38]. Also notice that for all subjects the performance improved significantly by using our approach.

Table 3 shows the results obtained using dataset 2b, and comparisons with others works, such as FBCSP [38], CNN-VAE [48], CNN-SAE [57], CWT-CNN [80], SSD-SE-CNN [61] and VGG-16 CNN [58]. We observed that the proposed method outperformed some approaches, such as FBCSP, CNN-SAE, CNN-VAE and VGG-16 CNN. The CWT-CNN and SSD-SE-CNN methods achieved the best performance, notably influenced by results obtained on Subjects B02 and B03, but our system also outperformed the CWT-CNN approach for two subjects, namely B01 and B05. With respect to SSD-SE-CNN, our system outperformed it on four subjects, namely B01, B05, B07 and B08. Also notice that the performance on Subject B05 was improved significantly by our approach with respect to the other methods, as shown in Table 3.

Table 4 shows our results obtained on dataset IVa and the comparison with other related works, such as [62], CSP\AM-BA-SVM [63], TKCSP [68], SR-MDRM [66], [67], TDA+CNN [69], [64] and FCCSP [65]. Our system in comparison to TKCS, SR-MDRM and FCCSP, which that considered 118 channels, performed better on Subjects "aa", "av" and "ay". With respect to CSP\AM-BA-SVM, [67], TDA+CNN and [64], which employed channel selection, our method improved the accuracy for Subject "av". The method developed in [62] achieved mean ACC of 94.20%, which presents superior performance in comparison to our proposed approach (with ACC of 90.52%), however, other recent methods shown in Table 4 did not outperform our

TABLE 3. K-fold cross-validation performance in terms of kappa mean and standard deviation on dataset 2b.

Subjects	FBCSP [38]	CNN-SAE [57]	CNN-VAE [48]	CWT-CNN [80]	VGG-16 CNN [58]	SSD-SE-CNN [61]	PROPOSED (9x10)
B01	0.55 ± 0.02	0.52 ± 0.10	0.52 ± 0.08	0.71 ± 0.03	0.47	0.57 ± 0.03	**0.73 ± 0.10**
B02	0.21 ± 0.03	0.32 ± 0.07	0.35 ± 0.07	**0.46 ± 0.03**	0.18	0.36 ± 0.04	0.20 ± 0.11
B03	0.24 ± 0.02	0.49 ± 0.08	0.44 ± 0.06	**0.56 ± 0.04**	0.30	0.37 ± 0.04	0.17 ± 0.12
B04	0.89 ± 0.003	0.91 ± 0.02	0.91 ± 0.01	0.91 ± 0.02	0.86	**0.93 ± 0.01**	0.90 ± 0.06
B05	0.69 ± 0.005	0.66 ± 0.06	0.65 ± 0.08	0.65 ± 0.04	0.63	0.63 ± 0.04	**0.78 ± 0.08***
B06	0.53 ± 0.01	0.58 ± 0.05	0.64 ± 0.06	0.60 ± 0.03	0.42	**0.71 ± 0.03**	0.59 ± 0.10
B07	0.41 ± 0.01	0.49 ± 0.07	0.55 ± 0.07	**0.66 ± 0.03**	0.47	0.54 ± 0.04	0.65 ± 0.11
B08	0.41 ± 0.01	0.49 ± 0.11	0.51 ± 0.08	**0.70 ± 0.04**	0.56	0.59 ± 0.04	0.60 ± 0.11
B09	0.58 ± 0.01	0.46 ± 0.15	0.52 ± 0.08	**0.71 ± 0.04**	0.53	0.59 ± 0.04	0.51 ± 0.11
Average	0.50 ± 0.01	0.54 ± 0.08	0.56 ± 0.07	**0.66 ± 0.03**	0.49	0.59 ± 0.03	0.57 ± 0.03

*, indicates significant difference with a 95 % confidence level

TABLE 4. K-fold cross-validation performance in terms of accuracy mean on dataset IVa.

Subjects	Winner [62]	CSP\AM-BA-SVM [63]	TKCSP [68]	SR-MDRM [66]	TDA+CNN [67]	TDA+CNN [69]	TDA+CNN [64]	FCCSP [65]	PROPOSED (9x10)
aa	96.00	86.61	68.10	79.46	82.10	88.39	**98.21**	72.32	94.33
al	**100**	**100**	93.88	**100**	95.00	98.21	89.29	98.21	97.24
av	**81.00**	66.84	68.47	73.46	72.10	66.33	73.47	68.87	79.90
aw	**100**	90.63	90.58	89.28	90.70	94.64	92.86	78.57	93.54
ay	**98.00**	80.95	84.65	88.49	91.80	82.54	89.29	92.06	87.58
Average	**94.20**	85.01	81.14	86.13	86.30	86.02	88.62	82.01	90.52

TABLE 5. Nested K-fold cross-validation performance in terms of accuracy mean on dataset IVa.

Subjects	PROPOSED
aa	88.49
al	95.12
av	74.67
aw	94.60
ay	91.69
Average	88.91

method. Our system attained similar results on Subjects "aa", "al" and "av" with respect to [62], but the mean ACC decreased, mainly due to the results achieved on Subject "ay", who has the least amount of training data among all subjects.

Table 5 shows the results obtained on dataset IVa employing a Nested KFold cross-validation procedure with the whole dataset, similar to previous studies such as [70] and [71], but considering all EEG channels. These studies applied a KFold cross-validation using selected channels, obtaining promising results. The Nested KFold cross-validation used in our work shows that our approach achieved mean accuracy of 88.91%, confirming good generalization.

B. EXPERIMENT 2

Table 6 shows the achievements obtained by the proposed system and previous methods, such as SCN [23], FBCSP [38], GLOBAL [43], and MCNN and CCNN [45]. It is worth noting that significant improvement by applying our approach was obtained on three subjects (A03, A04, and A05) with respect to previous methods, such as MCNN that also presented good performance for other subjects. Also notice that ACC higher than 69% was obtained for almost all subjects, except for Subjects A02 and A06 (ACC around 52%). MCNN, which achieved the highest average accuracy from all subjects, slightly outperformed our method, which also reached promising results.

Figure 3 shows the confusion matrix for Subjects A02 and A03. The diagonal elements represent patterns correctly classified by our proposed system. Notice that for Subject A03 the four analyzed classes were very well discriminated, achieving the highest accuracy classifying MI tasks of left and right hands. For Subject A02, the four class discrimination was lower with respect to Subject A03. It is worth noting that the left hand MI tasks from Subject A02 were relatively well-identified by our system with respect to the right hand and the tongue MI. As a result, the best performance was obtained on this subject, decoding feet MI. This confirms the high inter-subject variability in performing MI tasks. Additionally, 2D representations were obtained by using the "T-distributed Stochastic Neighbor Embedding" (t-SNE) method [81]. Figure 4 shows the spatial feature projection before training (Figure 4a), as well as after training for Subjects A02 (Figure 4b), A03 (Figure 4c), and A09 (Figure 4d.)

The class projection by applying convolutional layers seen in Figure 4, may explain the negative impact on recognition due to overlapped classes. In particular, good class separation was obtained on Subject A03, which justifies the highest accuracy using our approach. In contrast, for Subject A02 it was hard to discriminate appropriately his/her MI tasks for both left and right hands, as shown in Figure 4b. On the other hand, the graph of Subject A09 shows good discrimination for MI tasks linked to MI from left hand and tongue, and poor

TABLE 6. Intra-subject session to session classification results (% accuracy) on dataset 2a.

Subjects	FBCSP [38]	SCN [23]	C2CM [24]	MCNN [45]	CCNN [45]	GLOBAL [43]	PROPOSED
A01	76.00	86.56	87.50	**90.21**	87.14	88.6	83.81
A02	56.28	62.29	**65.28**	63.40	63.10	55.9	51.97
A03	80.88	89.86	90.28	89.35	86.76	86.7	**91.48***
A04	61.07	65.61	66.67	71.16	68.29	71.0	**73.82***
A05	54.85	55.19	62.50	62.82	63.61	66.5	**69.82***
A06	45.48	48.47	45.49	47.66	48.32	**56.0**	53.90
A07	82.98	86.07	89.58	90.86	87.73	88.4	**91.17**
A08	81.63	78.41	83.33	**83.72**	80.17	80.9	81.87
A09	70.45	76.05	79.51	82.32	78.83	77.1	**82.39**
Average	67.70	72.05	74.46	**75.72**	73.77	74.6	75.58

*, indicates significant difference with a 95 % confidence level

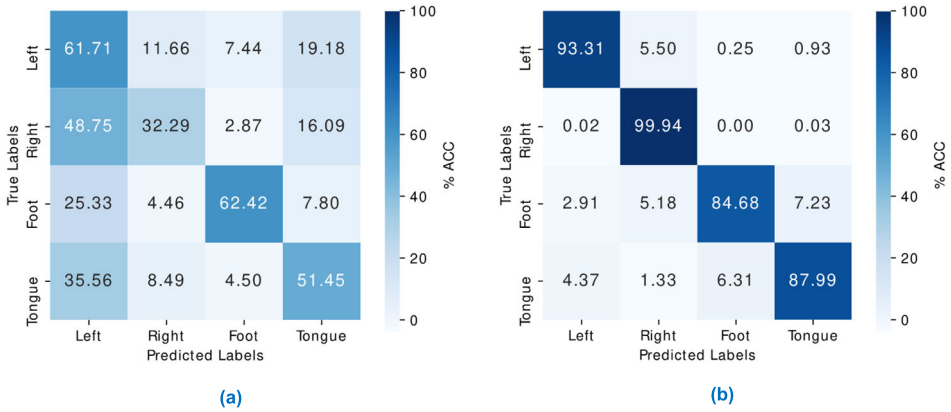


FIGURE 3. Confusion matrix of the evaluation data for subject A02 (a) and subject A03 (b).

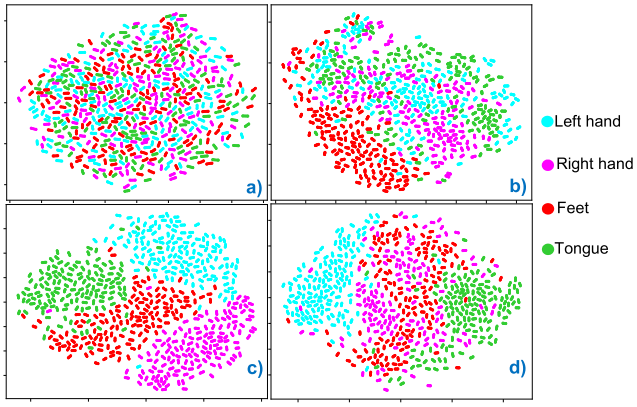


FIGURE 4. The t-SNE visualization of the proposed model-based feature embedding using the input to fully connected layer a) pre- training b) Subject A02 c) Subject A03 and d) Subject A09.

separation for the other two classes, confirming again the high inter-subject variability.

Table 7 shows the results obtained on dataset 2b. Here, we only compared our results with previous works that used all training sessions, such as CNN-SAE [57], CWT-SCNN [59] and CA_{gross}[60]. Therefore, the system proposed in [38] was excluded for this evaluation. Also notice that the best results in [60] were not considered in Table 7 for comparison, as they were obtained in other conditions.

TABLE 7. Intra-subject on session to session classification results (% accuracy) on dataset 2b.

Subjects	CNN-SAE [57]	CWT-SCNN [59]	CA_{gross} [60]	PROPOSED
B01	**78.10**	74.70	68.31	75.43
B02	63.10	**81.30**	55.10	55.36
B03	60.60	**68.30**	54.61	52.09
B04	**95.60**	96.30	91.14	94.96
B05	78.10	**92.50**	80.17	87.60
B06	73.80	**86.90**	72.23	79.71
B07	70.00	73.40	67.79	**79.77***
B08	71.30	**91.60**	88.65	87.87
B09	**85.00**	84.40	80.23	84.69
Average	75.10	**83.20**	73.13	77.50

*, indicates significant difference with a 95 % confidence level

For this reason, we show the results obtained in [60] by using CA_{gross}. We observed that our method outperformed the results obtained by applying CNN-SAE and CA_{gross}, but CWT-SCNN performed better. Notice that our method achieved mean ACC higher than 77.50% for six subjects B04, B05, B06, B07, B08 and B09. Also, notice that the results improved significantly (ACC of 79.77%) for Subject B07, with respect to the others approaches shown in Table 7.

C. EXPERIMENT 3

A similar experiment using non-subject-specific classification was carried out in [45] employing the dataset 2a, where

TABLE 8. Inter-subject on session to session classification results (% accuracy) on dataset 2a.

Subjects	SCN [23]	MCNN [45]	CCNN [45]	PROPOSED
A01	47.06	51.91	62.07	**72.29***
A02	31.22	38.06	42.44	39.26
A03	41.02	43.34	63.12	**81.59***
A04	33.19	35.81	52.09	**60.90***
A05	41.57	41.50	49.96	**54.03***
A06	34.71	31.11	37.16	**51.58***
A07	43.09	48.09	62.54	**74.70***
A08	46.01	45.01	59.32	**77.11***
A09	51.78	51.29	69.43	**76.86***
Average	41.07	42.09	55.34	**65.37***

*, indicates significant difference with a 95 % confidence level

TABLE 9. Inter-subject on session to session classification results (% accuracy) on dataset 2b.

Subjects	PROPOSED
B01	66.94
B02	55.66
B03	54.08
B04	93.78
B05	79.09
B06	80.81
B07	74.18
B08	90.05
B09	83.42
Average	75.33

the results were obtained by also using the testing set for the training process of an Autoencoder. The performance of other approaches in [45] are shown in Table 8 together with our achievement using the proposed system. As such a comparison with three methods, SCN [23], MCNN and CCNN [45] indicates that our recognition system achieved the best performance with a significant difference for almost all subjects, except for Subject A02. Furthermore, our approach reached higher mean ACC (65.37%) with a significant difference over the other methods.

This experiment was also conducted on dataset 2b, and the results are shown in Table 9. We noted that our approach decreased its performance slightly, getting mean ACC of 75.33%, which is comparable with respect to Experiment #2 (see Table 7) and also enhanced the accuracy for Subjects B03, B06 and B08, demonstrating good generalization using a non-subject-specific training stage.

D. EXPERIMENT 4

The results obtained are shown in Table 10 (for dataset 2a) and Table 11 (for dataset 2b). On both datasets, the performance by applying inter-subject transfer learning declined in comparison to other experiments, but the accuracy decreased notably on dataset 2a. For this same dataset 2a, the average accuracy of our system by applying inter-subject transfer learning was higher with respect to SCN [23] and MCNN [45] in Table 8, which is remarkable because these works used

TABLE 10. Inter-subject transfer learning classification results (% accuracy) on dataset 2a.

Subjects	PROPOSED
A01	65.62
A02	31.95
A03	63.08
A04	42.73
A05	31.57
A06	33.21
A07	41.45
A08	60.70
A09	58.38
Average	47.63

TABLE 11. Inter-subject transfer learning classification results (% accuracy) on dataset 2b.

Subjects	CA_{gross} [60]	PROPOSED
B01	68.23	**71.85***
B02	55.41	**55.46**
B03	54.20	**55.32***
B04	81.42	**88.63***
B05	65.40	**68.58***
B06	71.59	**74.11***
B07	68.53	**69.89***
B08	72.13	**75.91***
B09	73.13	**74.75***
Average	67.78	**70.50***

*, indicates significant difference with a 95 % confidence level

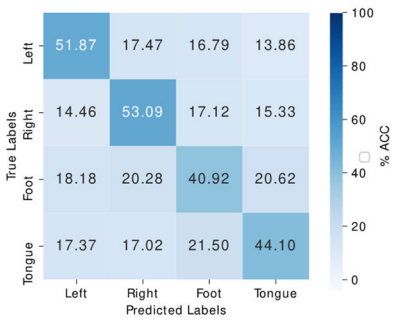


FIGURE 5. The mean confusion matrix over all subject on the dataset 2a by applying inter-subject transfer learning.

data from all subjects during the training process. Moreover, the results for Subjects A01 and A08 are encouraging.

Figure 5 shows the mean confusion matrix over all subjects. The best MI discrimination was obtained for both left and right hands. This finding is encouraging, as upper limb rehabilitation of post-stroke patients is widely used, in order to increase their independence in performing daily activities [4], [5], [82].

Our results using inter-subject transfer learning on dataset 2b are compared with CA_{gross} proposed in [60]. We observed that our system reached a higher mean accuracy of 70.50%, and also enhanced the accuracy for all subject, with a significant difference for almost all subjects, except for Subject B02.

It is worth noting that performance was consistent across all experiments for two groups of subjects: 1) A02, A06, B02 and B03; 2) A01, A03 and B04, which agrees with other published studies. Therefore, we consider that the variation of results (good and poor accuracy) may be more related to the quality of the data [60].

V. CONCLUSION

A motor imagery-based BCI using a convolutional network was developed in this work, employing SCN for multi-class classification. Several experiments to evaluate the proposed model were carried out using well-known international databases from BCI Competition IV and BCI Competition III. Our results showed improvements over existing state-of-the-art methods. In this research we explored the performance of our recognition system for subject-specific and non-subject-specific classification. We also evaluated the pre-trained model to recognize MI tasks from unknown subjects, which may have a relevant impact on real-life scenarios in clinical applications, as it can enable new individuals to use a BCI without previous calibration stage.

ACKNOWLEDGMENT

The authors would like to thank the providers of the BCI Competition IV Datasets 2a and 2b and the BCI Competition III Dataset IVa.

REFERENCES

[1] D. L. Schomer and F. L. Da Silva, *Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. 6th ed. Lippincott Williams & Wilkins, 2012.

[2] M. Clerc, L. Bougrain, and F. Lotte, *Brain-Computer Interfaces: Methods and Perspectives*. Hoboken, NJ, USA: Wiley, 2016.

[3] U. Chaudhary, N. Birbaumer, and A. Ramos, "Brain–computer interfaces for communication and rehabilitation," *Nature Rev. Neurol.*, vol. 12, pp. 513–525, Aug. 2016.

[4] K. K. Ang and C. Guan, "Brain-computer interface in stroke rehabilitation," *J. Comput. Sci. Eng.*, vol. 7, no. 2, pp. 139–146, 2013.

[5] K. K. Ang and C. Guan, "Brain–computer interface for neurorehabilitation of upper limb after stroke," in *Proc. IEEE*, vol. 103, no. 6, pp. 944–953, May 2015.

[6] N. Birbaumer, N. Ghanayim, T. Hinterberger, I. Iversen, B. Kotchoubey, A. Kübler, J. Perelmouter, E. Taub, and H. Flor, "A spelling device for the paralysed," *Nature*, vol. 398, no. 6725, pp. 297–298, Mar. 1999.

[7] D. Beukelman, S. Fager, and A. Nordness, "Communication support for people with ALS," *Neurol. Res. Int.*, vol. 2011, 2011.

[8] U. Chaudhary and N. Birbaumer, "Communication in locked-in state after brainstem stroke: A brain-computer-interface approach," *Ann. Transl. Med.*, vol. 3, no. 1, 2015.

[9] N. Mrachacz-Kersting, N. Jiang, A. J. T. Stevenson, I. K. Niazi, V. Kostic, A. Pavlovic, S. Radovanovic, M. Djuric-Jovicic, F. Agosta, K. Dremstrup, and D. Farina, "Efficient neuroplasticity induction in chronic stroke patients by an associative brain-computer interface," *J. Neurophysiol.*, vol. 115, no. 3, pp. 1410–1421, Mar. 2016.

[10] G. Pfurtscheller, G. R. Müller, J. Pfurtscheller, H. J. Gerner, and R. Rupp, "'Thought'—Control of functional electrical stimulation to restore hand grasp in a patient with tetraplegia," *Neurosci. Lett.*, vol. 351, no. 1, pp. 33–36, Nov. 2003.

[11] J. Lüdemann-Podubecká, K. Bösl, and D. A. Nowak, "Repetitive transcranial magnetic stimulation in post stroke upper limb spasticity," *Case Med. Res.*, vol. 218, pp. 281–311, Aug. 2019.

[12] A. Kübler, A. Furdea, S. Halder, E. M. Hammer, F. Nijboer, and B. Kotchoubey, "A brain-computer interface controlled auditory event-related potential (P300) spelling system for locked-in patients," *Ann. New York Acad. Sci.*, vol. 1157, no. 1, pp. 90–100, Mar. 2009.

[13] G. Pires, U. Nunes, and M. Castelo-Branco, "Statistical spatial filtering for a P300-based BCI: Tests in able-bodied, and patients with cerebral palsy and amyotrophic lateral sclerosis," *J. Neurosci. Methods*, vol. 195, no. 2, pp. 270–281, Feb. 2011.

[14] D. Zhu, J. Bieger, G. G. Molina, and R. M. Aarts, "A survey of stimulation methods used in SSVEP-based BCIs," *Comput. Intell. Neurosci.*, vol. 2010, pp. 1–12, Jan. 2010.

[15] D. Lesenfants, D. Habbal, Z. Lugo, M. Lebeau, P. Horki, E. Amico, C. Pokorny, F. Gómez, A. Soddu, G. Müller-Putz, S. Laureys, and Q. Noirhomme, "An independent SSVEP-based brain–computer interface in locked-in syndrome," *J. Neural Eng.*, vol. 11, no. 3, Jun. 2014, Art. no. 035002.

[16] K. K. Ang, C. Guan, K. S. Phua, C. Wang, L. Zhou, K. Y. Tang, G. J. E. Joseph, C. W. K. Kuah, and K. S. G. Chua, "Brain-computer interface-based robotic end effector system for wrist and hand rehabilitation: Results of a three-armed randomized controlled trial for chronic stroke," *Frontiers Neuroeng.*, vol. 7, p. 30, Jul. 2014.

[17] J. Wolpaw, N. Birbaumer, D. McFarland, G. Pfurtscheller, and T. Vaughan, "Brain-computer interfaces for communication and control," *Clin. Neurophys.*, vol. 113, no. 6, pp. 767–791, 2002.

[18] G. Pfurtscheller, C. Guger, G. Müller, G. Krausz, and C. Neuper, "Brain oscillations control hand orthosis in a tetraplegic," *Neurosci. Lett.*, vol. 292, no. 3, pp. 211–214, Oct. 2000.

[19] G. Pfurtscheller and F. H. L. da Silva, "Event-related EEG/MEG synchronization and desynchronization: Basic principles," *Clin. Neurophysiol.*, vol. 110, no. 11, pp. 1842–1857, Nov. 1999.

[20] G. Pfurtscheller and A. Aranibar, "Evaluation of event-related desynchronization (ERD) preceding and following voluntary self-paced movement," *Electroencephalogr. Clin. Neurophysiol.*, vol. 46, no. 2, pp. 138–146, Feb. 1979.

[21] G. Pfurtscheller, "Event-related synchronization (ERS): An electrophysiological correlate of cortical areas at rest," *Electroencephalogr. Clin. Neurophysiol.*, vol. 83, no. 1, pp. 62–69, Jul. 1992.

[22] G. Pfurtscheller, K. Zalaudek, and C. Neuper, "Event-related beta synchronization after wrist, finger and thumb movement," *Electroencephalogr. Clin. Neurophysiol./Electromyograp. Motor Control*, vol. 109, no. 2, pp. 154–160, Apr. 1998.

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

[24] S. Sakhavi and C. Guan, "Convolutional neural network-based transfer learning and knowledge distillation using multi-subject data in motor imagery BCI," in *Proc. 8th Int. IEEE/EMBS Conf. Neural Eng. (NER)*, May 2017, pp. 588–591.

[25] G. Pfurtscheller, C. Neuper, A. Schlogl, and K. Lugger, "Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters," *IEEE Trans. Rehabil. Eng.*, vol. 6, no. 3, pp. 316–325, Sep. 1998.

[26] C.-I. Hung, P.-L. Lee, Y.-T. Wu, L.-F. Chen, T.-C. Yeh, and J.-C. Hsieh, "Recognition of motor imagery electroencephalography using independent component analysis and machine classifiers," *Ann. Biomed. Eng.*, vol. 33, no. 8, pp. 1053–1070, Aug. 2005.

[27] Y. Zhang, X. Zhang, H. Sun, Z. Fan, and X. Zhong, "Portable brain-computer interface based on novel convolutional neural network," *Comput. Biol. Med.*, vol. 107, pp. 248–256, Apr. 2019.

[28] S. Kelly, D. Burke, P. de Chazal, and R. Reilly, "Parametric models and spectral analysis for classification in brain-computer interfaces," in *Proc. 14th Int. Conf. Digit. Signal Process. (DSP)*, 2002, pp. 307–310.

[29] G. Pfurtscheller, C. Brunner, A. Schlögl, and F. H. L. da Silva, "Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks," *NeuroImage*, vol. 31, no. 1, pp. 153–159, May 2006.

[30] G. Pfurtscheller, A. Stancák, and C. Neuper, "Post-movement beta synchronization. A correlate of an idling motor area?" *Electroencephalogr. Clin. Neurophysiol.*, vol. 98, no. 4, pp. 281–293, Apr. 1996.

[31] Z. Tang, S. Sun, S. Zhang, Y. Chen, C. Li, and S. Chen, "A brain-machine interface based on ERD/ERS for an upper-limb exoskeleton control," *Sensors*, vol. 16, no. 12, p. 2050, Dec. 2016.

[32] G. Prasad, P. Herman, D. Coyle, S. McDonough, and J. Crosbie, "Applying a brain-computer interface to support motor imagery practice in people with stroke for upper limb recovery: A feasibility study," *J. NeuroEngineering Rehabil.*, vol. 7, no. 1, p. 60, Dec. 2010.

[33] Y. Hashimoto and J. Ushiba, "EEG-based classification of imaginary left and right foot movements using beta rebound," *Clin. Neurophysiol.*, vol. 124, no. 11, pp. 2153–2160, Nov. 2013.

[34] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 4, no. 2, pp. R1–R13, 2007.

[35] 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," *J. Neural Eng.*, vol. 15, no. 3, Jun. 2018, Art. no. 031005.

[36] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 4, pp. 441–446, Dec. 2000.

[37] K. Fukunaga, *Statistical Pattern Recognition*, 2nd ed. New York, NY, USA: Academic, 1990.

[38] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, "Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b," *Frontiers Neurosci.*, vol. 6, p. 39, Mar. 2012.

[39] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning nature," *Nature*, vol. 521, pp. 436–444, 2015.

[40] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: A review," *J. Neural Eng.*, vol. 16, no. 3, Jun. 2019, Art. no. 031001.

[41] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert, "Deep learning-based electroencephalography analysis: A systematic review," *J. Neural Eng.*, vol. 16, no. 5, Aug. 2019, Art. no. 051001.

[42] K. Roots, Y. Muhammad, and N. Muhammad, "Fusion convolutional neural network for cross-subject EEG motor imagery classification," *Computers*, vol. 9, no. 3, p. 72, Sep. 2020.

[43] J. J. Liao, J. J. Luo, T. Yang, R. Q. Y. So, and M. C. H. Chua, "Effects of local and global spatial patterns in EEG motor-imagery classification using convolutional neural network," *Brain-Comput. Interfaces*, vol. 7, nos. 3–4, pp. 47–56, 2020.

[44] S. U. Amin, M. Alsulaiman, G. Muhammad, M. A. Bencherif, and M. S. Hossain, "Multilevel weighted feature fusion using convolutional neural networks for EEG motor imagery classification," *IEEE Access*, vol. 7, pp. 18940–18950, 2019.

[45] S. U. Amin, M. Alsulaiman, G. Muhammad, M. A. Mekhtiche, and M. S. Hossain, "Deep learning for EEG motor imagery classification based on multi-layer CNNs feature fusion," *Future Gener. Comput. Syst.*, vol. 101, pp. 542–554, Dec. 2019.

[46] H. Wu, Y. Niu, F. Li, Y. Li, B. Fu, G. Shi, and M. Dong, "A parallel multiscale filter bank convolutional neural networks for motor imagery EEG classification," *Frontiers Neurosci.*, vol. 13, p. 1275, Nov. 2019.

[47] Z. Wang, L. Cao, Z. Zhang, X. Gong, Y. Sun, and H. Wang, "Short time Fourier transformation and deep neural networks for motor imagery brain computer interface recognition," *Concurrency Comput. Pract. Exper.*, vol. 30, no. 23, p. e4413, Dec. 2018.

[48] M. Dai, D. Zheng, R. Na, S. Wang, and S. Zhang, "EEG classification of motor imagery using a novel deep learning framework," *Sensors*, vol. 19, p. 551, 2019.

[49] T. Wang, E. Dong, S. Du, and C. Jia, "A shallow convolutional neural network for classifying MI-EEG," in *Proc. Chin. Autom. Congr. (CAC)*, Nov. 2019, pp. 5837–5841.

[50] Z. Tayeb, J. Fedjaev, N. Ghaboosi, C. Richter, L. Everding, X. Qu, Y. Wu, G. Cheng, and J. Conradt, "Validating deep neural networks for online decoding of motor imagery movements from EEG signals," *Sensors*, vol. 19, no. 1, p. 210, Jan. 2019.

[51] W. Qiao and X. Bi, "Deep spatial-temporal neural network for classification of EEG-based motor imagery," in *Proc. Int. Conf. Artif. Intell. Comput. Sci.*, Jul. 2019, pp. 265–272.

[52] W. Abbas and N. A. Khan, "DeepMI: Deep learning for multiclass motor imagery classification," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2018, pp. 219–222.

[53] Z. Zhang, F. Duan, J. Sole-Casals, J. Dinares-Ferran, A. Cichocki, Z. Yang, and Z. Sun, "A novel deep learning approach with data augmentation to classify motor imagery signals," *IEEE Access*, vol. 7, pp. 15945–15954, 2019.

[54] 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.

[55] M. Tangermann, K.-R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. J. Miller, G. R. Müller-Putz, G. Nolte, G. Pfurtscheller, H. Preissl, G. Schalk, A. Schlögl, C. Vidaurre, S. Waldert, and B. Blankertz, "Review of the BCI competition IV," *Frontiers Neurosci.*, vol. 6, p. 55, Jul. 2012.

[56] C. Brunner, R. Leeb, G. Müller-Putz, A. Schlögl, and G. Pfurtscheller, "BCI competition 2008–Graz data set A," *Inst. Knowl. Discovery (Lab. Brain-Comput. Interfaces), Graz Univ. Technol.*, vol. 16, pp. 1–6, 2008.

[57] Y. R. Tabar and U. Halici, "A novel deep learning approach for classification of EEG motor imagery signals," *J. Neural Eng.*, vol. 14, no. 1, Feb. 2017, Art. no. 016003.

[58] G. Xu, X. Shen, S. Chen, Y. Zong, C. Zhang, H. Yue, M. Liu, F. Chen, and W. Che, "A deep transfer convolutional neural network framework for EEG signal classification," *IEEE Access*, vol. 7, pp. 112767–112776, 2019.

[59] F. Li, F. He, F. Wang, D. Zhang, Y. Xia, and X. Li, "A novel simplified convolutional neural network classification algorithm of motor imagery EEG signals based on deep learning," *Appl. Sci.*, vol. 10, no. 5, p. 1605, Feb. 2020.

[60] S. Roy, A. Chowdhury, K. McCreadie, and G. Prasad, "Deep learning based inter-subject continuous decoding of motor imagery for practical brain-computer interfaces," *Frontiers Neurosci.*, vol. 14, Sep. 2020.

[61] 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.

[62] B. Blankertz, K.-R. Muller, D. J. Krusienski, G. Schalk, J. R. Wolpaw, A. Schlogl, G. Pfurtscheller, J. R. Millan, M. Schroder, and N. Birbaumer, "The BCI competition III: Validating alternative approaches to actual BCI problems," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 153–159, Jun. 2006.

[63] S. Selim, M. M. Tantawi, H. A. Shedeed, and A. Badr, "A CSP/AM-BA-SVM approach for motor imagery BCI system," *IEEE Access*, vol. 6, pp. 49192–49208, 2018.

[64] Y. Park and W. Chung, "Optimal channel selection using correlation coefficient for CSP based EEG classification," *IEEE Access*, vol. 8, pp. 111514–111521, 2020.

[65] Y. Guo, Y. Zhang, Z. Chen, Y. Liu, and W. Chen, "EEG classification by filter band component regularized common spatial pattern for motor imagery," *Biomed. Signal Process. Control*, vol. 59, May 2020, Art. no. 101917.

[66] A. Singh, S. Lal, and H. Guesgen, "Reduce calibration time in motor imagery using spatially regularized symmetric positives-definite matrices based classification," *Sensors*, vol. 19, no. 2, p. 379, Jan. 2019.

[67] J. Jin, Y. Miao, I. Daly, C. Zuo, D. Hu, and A. Cichocki, "Correlation-based channel selection and regularized feature optimization for MI-based BCI," *Neural Netw.*, vol. 118, pp. 262–270, Oct. 2019.

[68] M. Dai, D. Zheng, S. Liu, and P. Zhang, "Transfer kernel common spatial patterns for motor imagery brain-computer interface classification," *Comput. Math. Methods Med.*, vol. 2018, pp. 1–9, Mar. 2018.

[69] S. Huang, H. Peng, Y. Chen, K. Sun, F. Shen, T. Wang, and T. Ma, "Tensor discriminant analysis for MI-EEG signal classification using convolutional neural network," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 5971–5974.

[70] M. Miao, W. Hu, H. Yin, and K. Zhang, "Spatial-frequency feature learning and classification of motor imagery EEG based on deep convolution neural network," *Comput. Math. Methods Med.*, vol. 2020, pp. 1–13, Jul. 2020.

[71] O. Attallah, J. Abougharbia, M. Tamazin, and A. A. Nasser, "A BCI system based on motor imagery for assisting people with motor deficiencies in the limbs," *Brain Sci.*, vol. 10, no. 11, p. 864, Nov. 2020.

[72] R. Leeb, C. Brunner, G. Müller-Putz, A. Schlögl, and G. Pfurtscheller, "BCI competition 2008–Graz data set B," Graz Univ. Technol., Graz, Austria, Tech. Rep., 2008, pp. 1–6.

[73] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *Proc. IEEE Int. Joint Conf. Neural Netw. (IEEE World Congr. Comput. Intell.)*, Jun. 2008, pp. 2390–2397.

[74] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

[75] I. Nusrat and S.-B. Jang, "A comparison of regularization techniques in deep neural networks," *Symmetry*, vol. 10, no. 11, p. 648, Nov. 2018.

[76] H. Alemu, W. Wu, and J. Zhao, "Feedforward neural networks with a hidden layer regularization method," *Symmetry*, vol. 10, no. 10, p. 525, Oct. 2018.

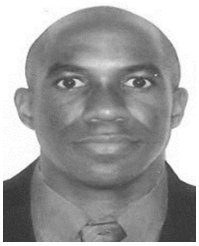
[77] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, 2015.
[78] S. Raschka, "Model evaluation, model selection, and algorithm selection in machine learning," 2018, *arXiv:1811.12808*. [Online]. Available: http://arxiv.org/abs/1811.12808
[79] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics Bull.*, vol. 1, no. 6, pp. 80–83, 1945.
[80] H. K. Lee and Y.-S. Choi, "Application of continuous wavelet transform and convolutional neural network in decoding motor imagery brain-computer interface," *Entropy*, vol. 21, no. 12, p. 1199, Dec. 2019.
[81] P. E. Rauber, S. G. Fadel, A. X. Falcao, and A. C. Telea, "Visualizing the hidden activity of artificial neural networks," *IEEE Trans. Vis. Comput. Graphics*, vol. 23, no. 1, pp. 101–110, Jan. 2017.
[82] A. A. Frolov, O. Mokienko, R. Lyukmanov, E. Biryukova, S. Kotov, L. Turbina, G. Nadareyshvily, and Y. Bushkova, "Post-stroke rehabilitation training with a motor-imagery-based brain-computer interface (BCI)-controlled hand exoskeleton: A randomized controlled multicenter trial," *Frontiers Neurosci.*, vol. 11, p. 400, Jul. 2017.
[83] A. A. S. Leon and J. R. N. Alvarez, "1D convolutional neural network for detecting ventricular heartbeats," *IEEE Latin Amer. Trans.*, vol. 17, no. 12, pp. 1970–1977, Dec. 2019, doi: 10.1109/TLA.2019.9011541.

DAILY MILANÉS HERMOSILLA was born in Santiago de Cuba, Cuba, in 1983. She received the bachelor's and M.S. degrees in automatic engineering from the Universidad de Oriente, in 2005 and 2011, respectively. Since September 2005, she has been a full-time Professor with the Department of Automatic Engineering, Universidad de Oriente, Cuba. She is currently enrolled in automatic doctoral program at the Universidad de Oriente. Her current research interests include robotic therapy, exoskeletons and rehabilitation, BCI systems, and deep learning.

RAFAEL TRUJILLO CODORNIÚ was born in Santiago de Cuba, Cuba, in 1956. He received the B.S. degree in mathematics from Odessa University, Ukraine, in 1979, and the Ph.D. degree in mathematics sciences from Rostov on Don University, in 1986. He is currently an Associate Professor with the Universidad de Oriente, Santiago de Cuba. He also works at SERCONI.

RENÉ LÓPEZ BARACALDO graduated from the Instituto Superior Politecnico Jose Antonio Echeverria, Havana, Cuba, in 2004. He currently works at Zimtronic LLC., applying deep learning in image processing and time series.

ROBERTO SAGARÓ ZAMORA was born in Santiago de Cuba, Cuba, in 1962. He received the Ph.D. degree in tribology, in 1995. He studies a postdoctoral research with Brazilian universities of Uberlandia and Piracicaba, in 2011. Since September 1985, he has been a full-time Professor with the Department of Mechanical Engineering, Universidad of Oriente, Cuba. He is the author of two book chapters and more than 120 scientific articles. His current research interests include biomechanics, robotic therapy, exoskeletons and rehabilitation, prostheses of lower and upper limb, and tribology. He received the Sciences Academic Annual Award of Cuban Republic, in 1990 and 2015.



DENIS DELISLE RODRIGUEZ received the B.E. degree in telecommunication and electronic engineering and the M.Sc. degree in biomedical engineering from the University of Oriente, Santiago de Cuba, Cuba, in 2005 and 2013, respectively, and the Ph.D. degree in electrical engineering from the Federal University of Espirito Santo (UFES), Brazil, in 2017. He is currently a Ph.D. Fellow with UFES, focused on biomedical signal processing, pattern recognition, brain–computer interfaces, and robotics for rehabilitation and assistance.

YOLANDA LLOSAS ALBUERNE (Member, IEEE) graduated in electrical engineering in automatic control from the Universidad de Oriente, Cuba, in 1979. She received the Ph.D. degree in technical sciences, in 1992. She is currently a full-time Principal Professor with the Department of Electricity, Technical University of Manabí, Ecuador. She works in the area of intelligent process control and its applications to energy efficiency. She has 40 years of experience as a university teacher and a researcher. She has supervised numerous master's theses and six doctorates, and three of them defended. She has also a coordinator of research projects and international collaboration networks for R + D + i projects. She has numerous publications in indexed journals.

JOSÉ RICARDO NÚÑEZ ÁLVAREZ (Member, IEEE) was born in Holguin, Cuba, in 1971. He received the M.Sc. degree in electrical engineering, in 2003, and the M.Sc. degree in automation, in 2014. From 1994 to 2018, he worked as a Professor with the Universidad de Oriente, Santiago de Cuba, Cuba. Since 2018, he has been working as a full-time Professor with the Department of Energy, Universidad de la Costa, Barranquilla, Colombia. He obtained the title of Electrical Engineer, in 1994. His research interests include industrial automation systems, electric power generation, transmission and distribution, and demotic and real estate systems.