IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 9, 2020, accepted October 24, 2020, date of publication November 3, 2020, date of current version November 18, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3035347

A Novel Deep Learning Scheme for Motor Imagery EEG Decoding Based on Spatial Representation Fusion

JUN YANG[1], ZHENGMIN MA[1], JIN WANG[2], AND YUNFA FU[1]
[1]Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650093, China
[2]Faculty of Information, Yunnan University, Kunming 650500, China

Corresponding author: Jun Yang (yang-jun@kust.edu.cn)

This work was supported in part by the Introduction of Talent Research Start-Up Fund Project of Kunming University of Science and Technology under Grant KKSY201903028, in part by the Postdoctoral Research Foundation of Yunnan Province 2020, and in part by the National Natural Science Foundation of China under Grant 31760281.

ABSTRACT Motor imagery electroencephalography (MI-EEG), which is an important subfield of active brain–computer interface (BCI) systems, can be applied to help disabled people to consciously and directly control prosthesis or external devices, aiding them in certain daily activities. However, the low signal-to-noise ratio and spatial resolution make MI-EEG decoding a challenging task. Recently, some deep neural approaches have shown good improvements over state-of-the-art BCI methods. In this study, an end-to-end scheme that includes a multi-layer convolution neural network is constructed for an accurate spatial representation of multi-channel grouped MI-EEG signals, which is employed to extract the useful information present in a multi-channel MI signal. Then the invariant spatial representations are captured from across-subjects training for enhancing the generalization capability through a stacked sparse autoencoder framework, which is inspired by representative deep learning models. Furthermore, a quantitative experimental analysis is conducted on our private dataset and on a public BCI competition dataset. The results show the effectiveness and significance of the proposed methodology.

INDEX TERMS Brain–computer interface, discriminative and representative deep learning, feature fusion, convolution neural network, stacked sparse autoencoder.

I. INTRODUCTION

Brain–computer interface systems (BCIs) [1]–[3] try to map human intention from brain activities, providing a new pathway between the human brain and the external environment. The best-known applications of BCIs include clinical trials [4], emotion recognition [5], prosthesis or robot control [6], [7], and game interaction [8]. For EEG-based BCI studies, motor imagery electroencephalography (MI-EEG), which is the only active BCI paradigm without the requirement of external stimuli, is a popular and key research topic in BCI applications. During imaging the movement of certain parts of the body, the subjects' intention can be detected from a specific brain signal response, which is a phenomenon named as event-related synchronization (ERS) or event-related desynchronization (ERD) [9]. Therefore, the key step in MI-BCI tasks is to decode the MI-EEG signals efficiently.

Several studies have been conducted on BCIs tasks, highlighting the significance of MI-EEG data in the non-invasive BCI domain [10], [11]. The EEG-based BCIs is regarded as a pipeline framework, which includes three main parts: 1) Signal pre-processing involving data augmentation, noise and artefact removal, and electrode channel selection; 2) Feature extraction and representation of the appropriate properties or subcomponents of the constructed signal; 3) Classification that involves outputting the discrimination result by decoding the brain intention. Conventional techniques employ machine learning approaches with handcrafted features for decoding EEG signal [12], [13]. Moreover, the parameters are adjusted and handcrafted individually, leading to inefficient learning and unsatisfactory local optima [14]. In summary, it is challenging to translate brain dynamics and classify the cognitive outcomes. The

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

conventional techniques perform poorly considering the low signal-to-noise ratio, low spatial resolution, and highly non-stationary characteristics of MI-EEG signals. Furthermore, it's rather time-consuming to calibrate BCI system which involves in a large number of labelled data training to optimize the MI-EEG decoding framework before online testing [15]. It is essential to adapt the BCI system to the objective and handle the variations from subject to subject.

To address the challenge mentioned above, we utilize an innovative multi-layer convolution neural network (CNN) framework to learn complex temporal and spatial features from different granular-grouped channels. Furthermore, a stacked sparse autoencoder (SSAE) framework is proposed to construct the overall representation and capture the invariant features from the diversity of subjects. The experimental results show that the proposed model can help capture useful sub-spatial representations and perform well in the subject-to-subject transfer leaning. The significance of this article can be summarized as follows: first, the decoding method exploits the granular-grouped spatial and temporal structures of the intervention to obtain significantly invariant useful features for the across-subjects learning. Second, this study combines the discriminative and representative deep learning (DL) models [16] for an end-to-end MI-EEG decoding, in order to extract the latent spatial representation and elevate the generalization capability of the model.

II. RELATED WORK

DL methods have emerged as a promising technology that can feasibly provide end-to-end learning in a BCI system [17], [18]. DL framework can be broadly divided into generative, discriminative and miscellaneous types [19]–[21]. A discriminative DL (DDL) framework is applied to the classification, realized by characterizing through the conditional distribution of the categories. In the other aspect, considering the distribution of joint probability on the respective categories, the generative DL (GDL) architecture investigates the correlation of the observed data. The results of recent experiments show that DL methods, such as CNNs, deep belief networks (DBN) and recursive neural network (RNN) exhibit better classification accuracy than some state-of-the-art BCI methods [22]–[24]. Specifically, studies have focused on employing DL for MI-EEG decoding [25], [26]. Amin *et al.* employed multi-layer CNNs to learn temporal and spatial features from MI-EEG signal [27]. Ma *et al.* proposed channel-correlation architecture to construct the overall representation extracting from channels for MI-EEG decoding [28]. Considering single-trial MI EEG, Zhichuan *et al.* proposed a deep CNN to take the role of feature extraction and classification [29]. Lu *et al.* proposed a novel DL framework based on the restricted Boltzmann machine-combined frequential deep belief network to generate new representations of EEG features [30]. Yang *et al.* used the RNN–LSTM to extract the temporal dependence of MI-EEG data [31]. However, few studies have focused on exploring the influence and efficiency of spatial features. To this end, we construct a multi-layer CNN framework to exploit different granular-grouped information and utilize it to improve the decoding performance.

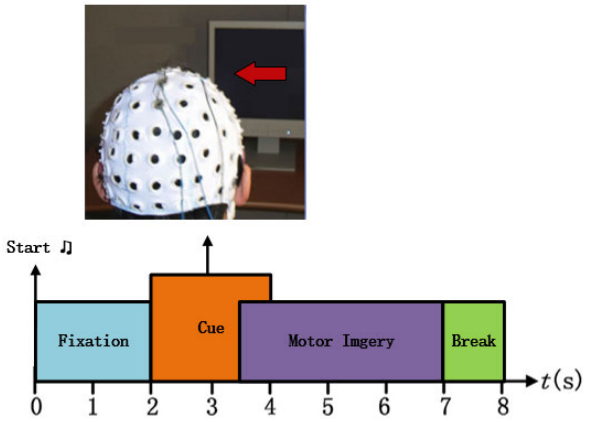

FIGURE 1. Paradigm of a motor imagery recording session.

III. TASK DEFINITION AND MATERIALS

We mainly elaborate the MI experimental task, organization and pre-processing of the data, proposed CNN and SSAE architecture, and training and testing strategies in this section.

A. MI TASK DEFINITION

For the MI task, we invited six human subjects (average age of 26.5±3.5, with three males) to perform the MI-based experiment. Each subject was seated in front of a computer screen and was instructed to perform MI tasks including executing the imaginary movement of the right or left hand. As shown in Fig. 1, each trial was designed as four steps including fixation, cue, MI, and break. A ''rest'' icon initially appeared on the screen for 2 s after the start tone. This guided the subjects to remain in a resting state, i.e., to clear their mind in preparation for the coming task. Subsequently, a left or right cue was presented for 2 s, and the subject continuously performed the specified MI task for 3.5 s, taking the reaction time into account. Finally, an interval of 1 s was provided as an additional break.

B. DATA ORGANIZATION AND DESCRIPTION

To validate the performance of the proposed approach for MI-EEG decoding, systematic and extensive experiments was conducted on public BCI competition IV dataset 1 and a private dataset collected from our laboratory based on the paradigm mentioned above. The EEG data were recorded using 64-channel caps attached to the scalp of the subjects. The dataset contains recordings of 300 trials on average for each subject. Each trial has a 3.5 s duration with a sampling frequency of 256 Hz. Table 1 lists the details of the dataset.

IV. METHOD

To exploit the end-to-end learning capability of the DL networks, the entire individual processing, including

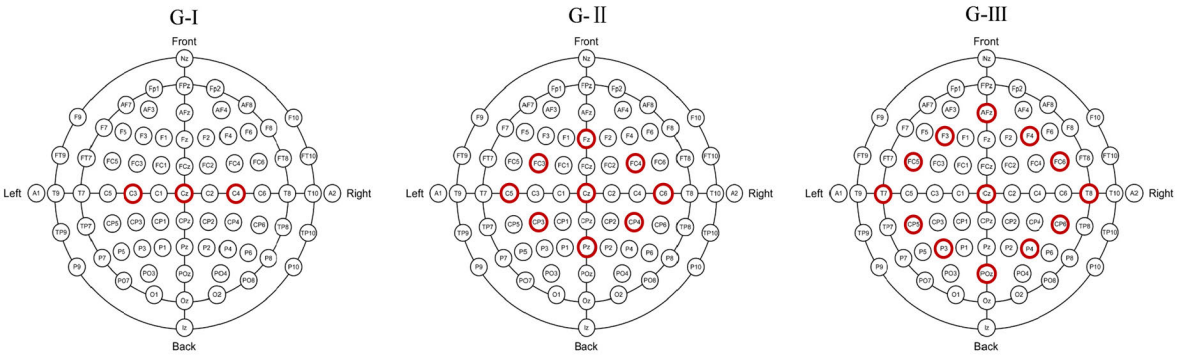

FIGURE 2. EEG channel selection of different groups.

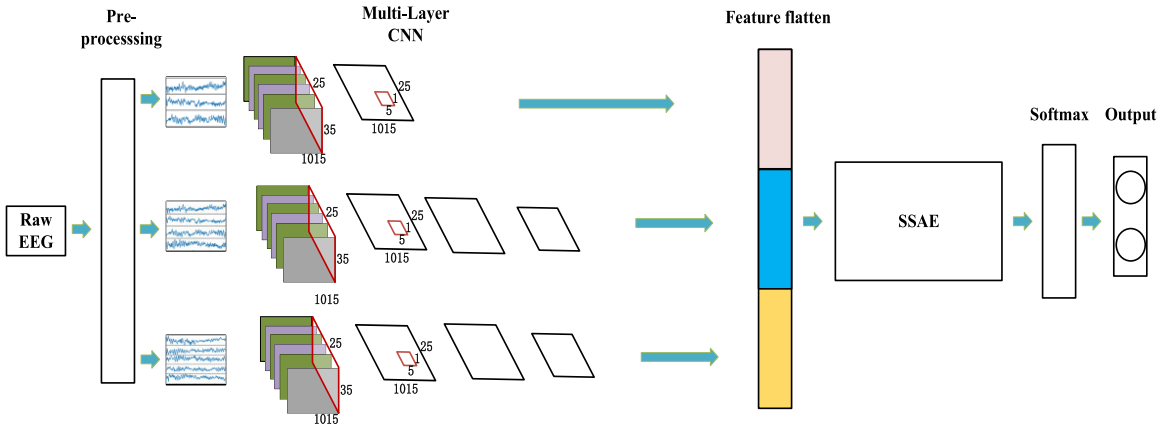

FIGURE 3. Proposed deep CNN framework based on multi-channel feature fusion.

TABLE 1. Properties of the datasets.

Datasets	Private	Public
	D_1	D_2
Subject	6	7
Channels	64	64
trials	243	270
imagery task	left, right hand (+ idle state)	left, right hand (+ idle state)
Perform periods	3.5s	3.5s

pre-processing, feature extraction and representation, and classification, should be implemented in one DL block. In this section, we introduce this decoding process in detail.

A. PREPROCESSING OF ACQUIRED DATA

In view of the noise and artifact impacts, we have to transform the available data into an informative and convenient manner for subsequent processing. The EEG data were preprocessed using the EEGLAB toolbox in MATLAB [32]. For convenience and to decreasing processing time, we first employ the common average reference (CAR) and then filter raw signals with frequencies between 7 and 35 Hz, which overlap with the main rhythmic components of ERS/ERD with regard to the MI-EEG [33]. Muscular and ocular artifacts [34] are removed through the plug-in toolbox. The sliding window approach is used to divide the MI trial data into individual segments (2 s) with an overlap of 0.5 s.

B. PROPOSED CNN ARCHITECTURE

EEG signal is diverse from the other data format because of its additional spatial resolution. Our goal is to decode the EEG from the hand MI tasks and produce invariant feature maps to translations on the input through corresponding electrode and temporal power transformation. By contrast with RNN which interested in forcing the model to process temporal and contextual correlation of EEG signals, such as LSTM [35], the standard CNN is schemed to recognize the overall shapes and is local invariant to the position of the shapes. Therefore, the CNN network is our optimum choice to exploits two important characteristics of the cortex potential: local correlation and invariance to different subjects [36,37]. Inspired by this fact, we propose a novel MI-EEG decoding method through a 2D-kernels-based CNN to capture the features of the frequency and electrode position considering the diversity of spatial granularity. Multi-layer CNN networks are applied to extract MI-related features through the 2D kernel in the

TABLE 2. Granular groups of channels.

Group	Granularity	Electrodes
G-I	Low	3
G-I+G-II	Middle	12
G-I+G-III		16
G-I+ G-II+G-III	High	25

convolutional step; the features are then subsampled to a mini-type in the pooling step. In this design, the CNN part consists of alternating convolution, pooling and fully connected layers, with the convolution depth determined by the channel granularity.

To make full use of the spatial features and exploit the global invariant features form manifold subjects, we group the most informative channel for MI-EGG into three different granularities, as listed in Fig. 2 and Table 2.

It was demonstrated in several research [38], [39] that the MI-EEG signals from the C3, C4 and Cz electrodes (G-I) can obviously demonstrate the ERS/ERD characteristics. Thus, we will discuss the classification results from various fusion grouped of electrodes including G-I. Inspired by the hypothetical fact that invariant features are contained in different granularity of channels, our CNN architecture is consequently designed to filter three grouped input data and fusion them to capture global invariant representation for inter-subject [40], [41] transfer, as shown in fig.3.

We have tried many groups of CNN hyperparameters with a different number of layers and filters and ultimately adopt this structure (illustrated in Table 3) before model training. The deep learning frameworks including CNN, SSAE and contrast networks, the model hyperparameters as the size of neurons in the hidden layer and the convolution kernel used by increasing in a certain range and the cross-validation accuracy were recorded. The outperforming hyperparameters was chosen to determine the decoding model. Most model hyperparameters as the number of convolution layers were determined according to previous experience and size of input granularity.

Noted that the input MI-EEG contains two specific details (time and electrode locations), our aim is to classify the hand MI tasks through the 2D features. Considering the effect from different electrode location (spatial features), we have designed three CNN frameworks for the corresponding channel-grouped EEG data. The input of the network is expressed by $x \in R^{T\times C}$. The first layer of the CNN contains M filters with a kernel size of A×B; this layer can be employed to extract global information from all the channels. The mapping outcomes of the CNN between input 2D signals and a kernel is given by:

$$h_{ij}^{m} = \mathrm{ReLU}(a) = \mathrm{ReLU}((W^{m} * x)_{ij} + b_m) \tag{1}$$

where x is the input 2D signals, and W^m and b_m are the weight matrix and bias value for m order filter, respectively ($m = 1, 2, \ldots, M$). The activation function is selected as the rectified linear unit (ReLU) [42] function, which is employed to incorporate the nonlinear elements. The ReLU can be expressed as:

$$\mathrm{ReLU}(a) = \max(0, a) \tag{2}$$

The output of the convolutional layer is fed into the input of the pooling layer. The pooling is carried out with a max value sampling. Consequently, the output of the convolutional layer is subsampled to small-sized datasets. Table 3 lists the parameter details of each convolutional and max-pooling layer of our CNN architecture.

The use of CNN results in the improved discriminating capability of temporal and spatial variations of motor imagery patterns of an EEG image. In addition, since the neural network consists of not only a one-dimensional kernel to extract MI patterns of the input image but also shallow layers compared to conventional models, it has the advantage of low computation complexity in training. Since the proposed method utilizes an input image with time, frequency, and electrode information, the training of the neural network is robust to variations or abnormal patterns of MI EEG signals.

C. FEATURE REPRESENTATION USING SSAE

After the CNN architecture, we need a feature fusion block to filter to the global and useful representation which dominates the classifier. As a representative deep learning method, the SSAE is an improved version of the stack autoencoder (SAE) network [43] including input, hidden and output layer; thus, it can reconstruct its own characteristic representation. In terms of the SAE, the size of output is similar to the inputs. During training, the input x is first fed into the hidden layer to generate latent feature z, which corresponds to the decoder. Subsequently, z is mapped to the output layer to reconstruct y samples with a similar dimension and distribution to the input, which is named the encoder. The two steps can be expressed as:

$$z = \sigma(W_{x\to z}x + b_{x\to z}) \tag{3}$$

$$y = \sigma(W_{z\to y}x + b_{z\to y}) \tag{4}$$

where $W_{x\to z}$ and $W_{z\to y}$ are the weight matrices from input-to-hidden and hidden-to-output layers, respectively. Accordingly, $b_{x\to z}$ and $b_{z\to y}$ are the accordingly bias values of the hidden and output layers, respectively.

$\sigma(a)$ is the activation function which is defined as:

$$\sigma(a) = \frac{1}{1 + e^{-a}} \tag{5}$$

After the training of an autoencoder, the latent representation in the hidden layer can serve for as input to the higher layer in a deep hierarchical network or classification, which is consequently named the stacked autoencoder. Some researchers have applied the SAE to capture CNN features and have demonstrated that it can extract advanced features and outperform the CNN-only framework [44]. This is why the SAE is trained such that it automatically extracts the latent and robust features by reproducing the input. To capture the

TABLE 3. modified architecture of proposed CNN.

Group	Layers	Kernel	Stride	Output	Operation
G-I	C- I -1	30×1×40	10	(97, 3, 40)	40filters Conv2D 30×1
	P- I -1	1×3×40	1	(97, 1, 1)	3×40 Max-pooling
G- II	C- II -1	20×1×20	5	(195, 9, 20)	20filters Conv2D 20×1
	P- II -1	2×1×20	1	(97, 9, 1)	2×20 Max-pooling
	C- II -2	10×1×20	2	(93, 9, 20)	20filters Conv2D 20×1
	P- II -2	1×9×40	1	(93, 1, 1)	9×40 Max-pooling
G-III	C-III-1	20×1×20	5	(195, 13, 20)	20filters Conv2D 20×1
	P-III-1	2×1×20	1	(97, 13, 1)	2×20 Max-pooling
	C-III-2	10×1×20	2	(93, 13, 20)	20filters Conv2D 20×1
	P-III-2	1×9×40	1	(93, 1, 1)	13×40 Max-pooling

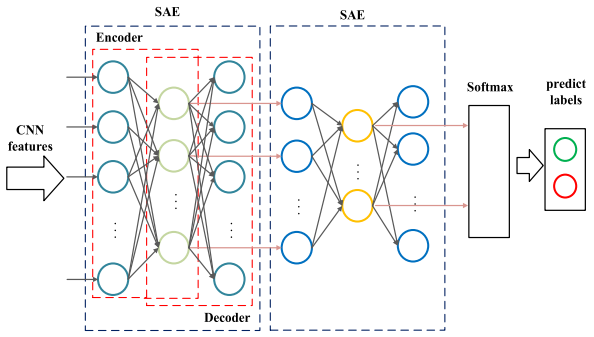

FIGURE 4. SSAE architecture.

important features from the data, a single layer of the SAE is insufficient. An SSAE can capture the features more precisely than the SAE. The SSAE can be constructed by stacking the hidden layers of the SAEs layer-by-layer. It captures the hierarchical information of the input data. The training data serve as input to the first SAE. After the first SAE training completed, the reconstruction layer is removed. Subsequently, the hidden layer output is inputted to the next SAE. In this work, we used the SAE, DBN and SSAE for fusing the features from CNN network and capturing the relevant information from the granular-grouped feature vector. Fig. 4 shows the proposed architecture of the SSAE.

The entire SSAE block is operating under supervised manner, and a softmax classifier [45] is added following this block. The SSAE model includes several alternating SAE blocks along with a softmax layer at the end. At the fine-tuning step, the output layer of each SAE is declined, and the output of latent representation will be directly fed into the next SAE. This operation transforms the multi-channel CNN features into advanced global representation, thus the learning efficiency and discriminating capacity of MI-EEG decoding networks are improved. The SSAE reconstructs the useful invariant information from different granular-grouped spatial features, and the best model is saved. When the entire network is trained for the same subject, it reconstructs a trial from another session with the same subject and same class. The softmax layer performs as a classifier for the reconstructed representation.

We can optimize the entire model by minimizing the cost function, as shown below:

$$\underset{W,b}{\operatorname{argmin}}[\theta(x, \mathrm{y})] \tag{6}$$

where $\theta(x, z)$ is the reconstruction error when the model is trained to reconstruct the global features for the final output.

V. EXPERIMENTAL VERIFICATION

In this section, an experimental demonstration of the proposed method is presented, and feature learning and certain interrelated hyper-parameters of the proposed architecture are studied. In this study, we mainly adopt inter-subject [46], [47] training type and take comparison with the intra-subject one (it will be special described if needed). Across-subjects training use one subject as a testing set and all the rest as a training set. The other is the individual training (intra-subject) that individual acts as the training and testing set. The inter-subject training methodology about MI based BCI is regarded as more challenging about subject information transfer, and more generalized and robust than the intra-subject one. Taking the intra-subject training into account, we performed 30 trials for unit validation dataset and thus 8-fold and 9-fold cross validation for D1 and D2 respectively. Fig. 5 shows the examples of inter-subject training for S_1.

A. SPATIAL FEATURES LEARNING

Table 4 lists the diverse results obtained using the different CNN layers with the SSAE fusion architecture. The combination of G- I with G-II, G-I with G-III, and G-I with G-II and G-III gave an improvement in terms of the accuracy. The accuracy obtained by fusing all the CNN networks was the best overall. Increasing the channel number in DL requires more computational resources, leading to more considerations in the construction of a BCI system.

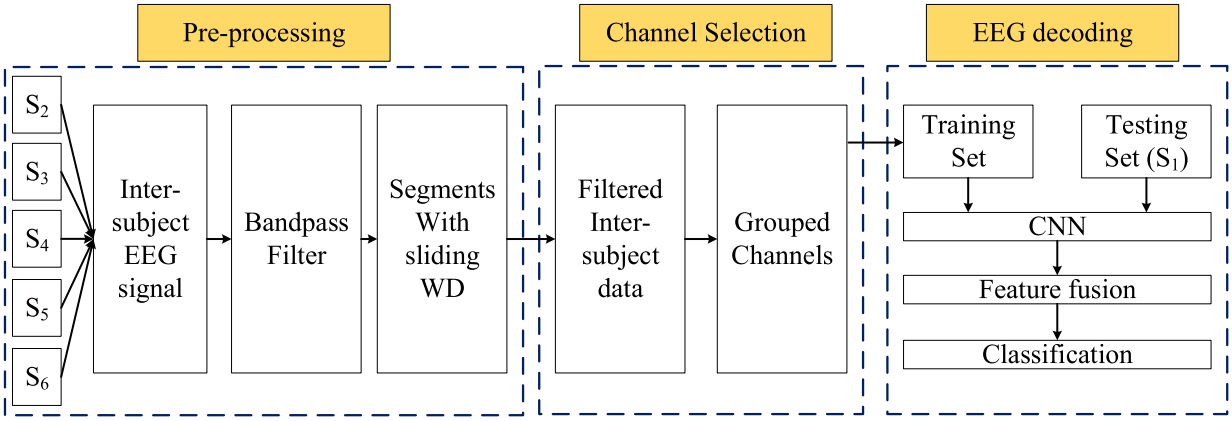

FIGURE 5. Block diagram representing the EEG trial structure for S_1 inter-subject training.

TABLE 4. recognition accuracy of various CNN combinations.

Fusion CNN scheme	Accuracy (D_1)	Accuracy (D_2)
CNN (G- I)	0.808	0.825
CNN (G-I+G-II)	0.820	0.847
CNN (G-I +G-III)	0.837	0.856
CNN (G-I+G-II+G-III)	0.847	0.864

In order to investigate the activation of spatial features, the twice mapping power topographical distribution in the D1 and D2 were computed after the proposed DL framework learning and depicted in Fig. 6, where the color encodes the average power of mapping fusion features corresponding to the position of different electrodes signals, which indicate the mapping feature energy from different electrodes inputs. We took the power of fusion features from three granularity of combination CNN processing as the indicator to explore the ERS/ERD from electrodes and the effect after spatial feature extracting. The power of the features apparently fluctuates in the high granular electrodes, indicating that our proposed method can learn more discriminant and informative features from the raw EEG data. We can obviously find extremely similar ERD for homologous MI data sets. It displays an evident contralateral dominance. This result reveals that the hand MI task activated the areas of brain motor cortex and the proposed framework can capture the useful features.

Feature fusion network are compared with the deep brief network (DBN), SAE and SSAE. The testing was performed with learning rates of 0.01 and 0.05, with the former yielding a higher accuracy. In each of the SSAE, SAE, and DBN models, the training process was initiated by increasing the number of neurons in the hidden layer in the range of 10–100, and the accuracy was recorded. The outperforming number of neurons was chosen from the set of {10, 20, ... 100}. Subsequently, the hidden layers were added consecutively, and this procedure was continued until the addition of a hidden layer no longer improved the accuracy. The parameters associated with the last stage were saved as the final parameters.

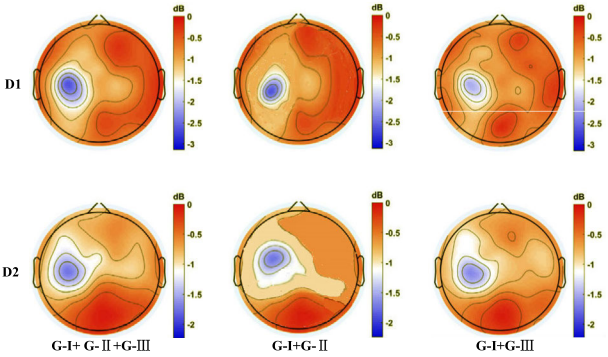

FIGURE 6. Topographical distribution of power on different data.

Tables 5 and 6 list a final comparison of the different feature fusion models employed for inter-subject training. It can be found that the SSAE obviously outperforms the other two approaches for most subjects, exhibiting average accuracies of 0.847 and 0.864 on the two datasets, respectively. Comparing the other feature fusion method, we find that the SSAE achieves a relatively lower standard deviation of all subjects, indicating that the SSAE performs more robustly on subject's diversity.

B. INSPECTION OF TRAINING PROCESS

To evaluate the proposed DL network in training, Fig. 7 shows the training and testing values for 26 epochs in cases without preprocessing, without grouped channels, and with the application of the proposed CNN–SSAE. The graphs show that when the preprocessing is absent (the case a), the peak value of accuracy for training is approximately 0.75 with a validation accuracy below 0.6. This poor result is due to the consideration of redundant information in the input, where large amounts of data are fed into the end-to-end system, and convolutional layers failure to find effective patterns that allows discriminating movement intentions. In case b, using most channel signals without grouping, the test accuracy almost reaches 0.7. However, significant differences can be

TABLE 5. Recognition accuracy of different feature fusion methods on a public dataset.

Feature fusion	D_1						Avg.	Standard deviation
	S_1	S_2	S_3	S_4	S_5	S_6		
SAE	0.787	0.803	0.724	0.782	0.798	0.815	0.785	0.0292
DBN	0.746	0.743	0.713	0.751	0.812	0.749	0.752	0.0296
SSAE	0.853	0.841	0.887	0.797	0.844	0.857	0.847	0.0267

TABLE 6. Recognition accuracy of different feature fusion methods on a private dataset.

Feature fusion	D_2							Avg.	Standard deviation
	S_A	S_B	S_C	S_D	S_E	S_F	S_G		
SAE	0.770	0.803	0.724	0.782	0.828	0.793	0.835	0.791	0.034
DBN	0.786	0.743	0.753	0.771	0.792	0.822	0.803	0.781	0.026
SSAE	0.883	0.841	0.877	0.857	0.854	0.883	0.847	0.864	0.016

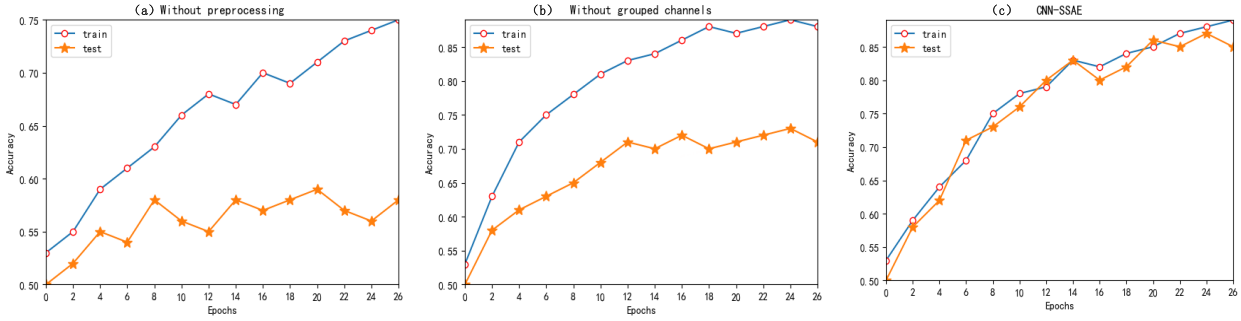

FIGURE 7. Train and test validation behavior.

observed between the training and testing values, suggesting overfitting [48]. Finally, in case c, i.e., when applying the proposed method, both the training and validation values are higher than 0.8, indicating less overfitting. This result validates the previous works, where channel-grouped CNN algorithms were found to exhibit superior performance.

Fig. 8 shows the values of training and testing losses indicated by red and blue dotted lines, respectively. The classification loss of the entire framework first decreases and then steadies at a fixed level during the training process, showing no evident signs of overfitting.

Fig. 9 shows the result of the epoch size on the kappa value performance, along with the average training period (each training set contains 280 trials) on D_1. The kappa coefficient [49] is a measure of the overall accuracy, obtained by eliminating the randomness of the classification result. It can be defined as follows:

$$K_{appa} = \frac{p_0 - p_e}{1 - \mathrm{p_e}} \tag{7}$$

where P_0 is the overall classification accuracy, and P_e is the rate of theoretical consistency which is calculated as:

$$p_e = \frac{\alpha_1\beta_1 + \alpha_2\beta_2 + \ldots + \alpha_c\beta_c}{n \times n} \tag{8}$$

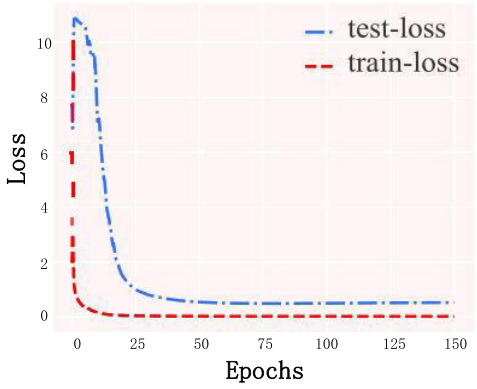

FIGURE 8. Training and testing loss.

where $\{\alpha_1, \alpha_2, \ldots, \alpha_c\}$ and $\{\beta_1, \beta_2, \ldots, \beta_c\}$ denote the numbers of real and predicted samples in the specific category, respectively, c is the category size, and n is the total sample size. The kappa value typically falls between 0 and 1, with the higher value indicating a high degree of consistency. As shown in Fig. 9, the performance corresponding to an epoch size of 400 with an appropriate computation

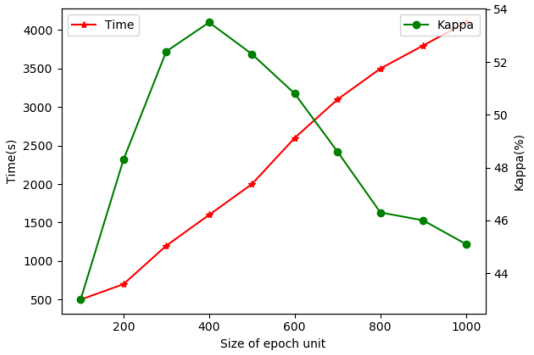

FIGURE 9. Performance of epoch on kappa value and training time.

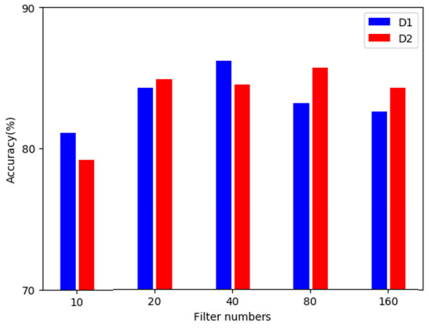

FIGURE 10. Performance of epoch on kappa value and training time.

TABLE 7. Average accuracy between two training strategy.

Framework	Intra-subject		Inter-subject	
	D1	D2	D1	D2
CNN(without grouped)	0.763	0.742	0.717	0.723
CNN(without fusion)	0.773	0.763	0.703	0.694
CNN-SSAE	0.851	0.833	0.847	0.864

time (1500 s) is the best. The number of epochs is set as 300. The time consumption increases in a largely linear manner.

The size of the filters (hyper parameters) in the convolutional network can affect the performance and efficiency of the end-to-end decoding process, as shown in Fig. 10. In this case, we mainly arrange the network with filter numbers of approximately 20 and 40.

In order to detect the effectiveness of the different DL steps, we divided the proposed framework for only grouped feature extractor (multi-layer CNN) and extracting by Multi-Layer Perceptron (MLP) with SSAE fusion which ignore the spatial correlation. We get the following results (Table 7) through two training strategies. We can obviously perceive the low average accuracy (less than eighty percent) and larger gaps between two training strategies by virtue of divided methods which indicate the two DL steps function in useful and subjects-invariant features learning.

C. MODEL EVALUATION

Aimed to detect the across-subjects discrimination and demonstrate the effectiveness of our method, the results

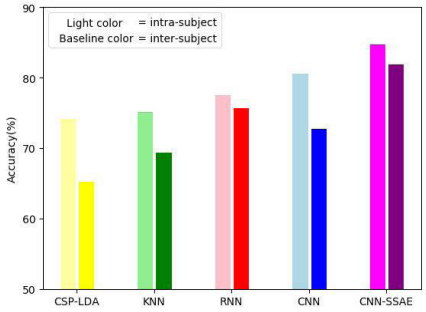

FIGURE 11. Comparison of five methods in different training.

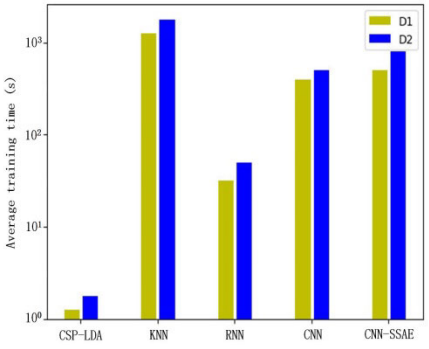

FIGURE 12. Performance of training time.

obtained the reference mentioned method CSP–LDA [50], KNN [51], RNN [52], and CNN [53] for decoding the same MI tasks are compared with different training strategy as illustrated in Fig. 11.

It is apparent from the histogram that the proposed approach achieves a higher accuracy and lower deviation between different training strategies, which shows that the proposed scheme enhances the capability of decoding complex MI tasks and outstanding generalization. It is interesting that there is also small gap in the RNN method which guides us to exploit it in further work.

Further, we experimentally analyze the computational complexity of five methods by comparing their average training time for all subjects with two datasets. From Fig. 12, we can clearly observe that the average training time of the KNN are longer than other approaches. The fact that the CSP-LDA runs faster than DL models demonstrates its efficiency, but unfortunately with extremely low accuracy revealed in Fig. 11. It is noted that RNN is unexpectedly time-consuming, although there is no channel selection about it. In addition, all methods are slow in D2 which can be explained by involving more subjects and trials.

For model convergence evaluation, we replace the feature extracting part (multilayer CNN) of the DL decoding models. The kappa value for the MI task according to the increasing epochs employed among the three models was illustrated in Fig. 13. It was observed that the kappa value climbs sharply after the 13-15 epochs. All models showed that kappa converged as the training epochs increased. The scatter line

TABLE 8. Kappa value result of CNN-SSAE methods compared with CNN and FBCSP (public data).

Method	D1						Avg.
	S_1	S_2	S_3	S_4	S_5	S_6	
FBCSP	0.244	0.347	0.247	0.431	0.535	0.486	0.381
CNN	0.325	0.417	0.317	0.387	0.426	0.541	0.402
CNN-SSAE	**0.514**	0.401	**0.434**	**0.469**	**0.448**	**0.503**	**0.462**

TABLE 9. Kappa value result of CNN-SSAE methods compared with CNN and FBCSP (private data).

Method	D2							Avg.
	S_A	S_B	S_C	S_D	S_E	S_F	S_G	
FBCSP	0.328	0.203	0.279	0.263	0.381	0.430	0.214	0.299
CNN	0.414	0.305	0.404	0.379	0.289	0.457	0.329	0.368
CNN-SSAE	**0.488**	0.289	**0.593**	**0.489**	**0.443**	0.405	**0.494**	**0.457**

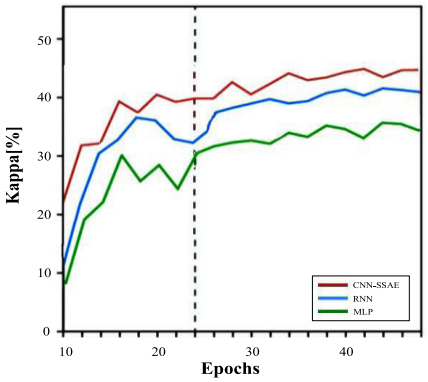

FIGURE 13. Mean classification accuracy and convergence point for different feature extracting frameworks.

TABLE 10. Accuracy of model in different initializations.

Initialization cases	FBCSP		CNN		CNN-SSAE	
	D1	D2	D1	D2	D1	D2
1	0.717	0.723	0.763	0.742	0.831	0.813
2	0.821	0.774	0.732	0.773	0.774	0.834
3	0.805	0.780	0.773	0.842	0.803	0.784
4	0.786	0.767	0.831	0.833	0.774	0.836
5	0.768	0.784	0.813	0.826	0.766	0.812
6	0.803	0.694	0.734	0.763	0.832	0.794
7	0.817	0.824	0.847	0.831	0.805	0.853
8	0.720	0.716	0.781	0.816	0.834	0.833
mean	0.780	0.758	0.784	0.803	0.802	0.820
Standard deviation	0.041	0.043	0.043	0.038	0.028	0.023

in graph shows the convergence performance of the different frameworks' training loss. It indicates that at least 23 epochs were needed to classify motor imagery tasks reliably with the converged performance. In general, the proposed feature learning framework achieve a high classification and fast convergence performance.

For evaluating the different architectures, we used all the subjects of D1 and D2 for inter-subject training under the filter bank common spatial patterns (FBCSPs), CNN, and CNN–SSAE decoding framework. To compare our approach with these methods, we highlighted the best-case kappa values on bold labels. As listed in Tables 8 and 9, the average kappa value among subjects was 0.381 for FBCSP [54] and 0.402 for the CNN. The average kappa value of our approach is higher than those of the two approaches. Our method outperforms the FBCSP and CNN methods for 5 out of the 6 subjects in D1 and 5 out of the 7 subjects in D2.

To explore the aftereffect of the random initializations of the model hyperparameter, eight initialization setting cases (including various combinations of the spatial filter size, CNN kernel size, and increasing number of neurons in the hidden layer) are used to make a comparison on different datasets through inter-subject training, as listed in Table 10. The results show that the CNN–SSAE noticeably outperforms the other two methods in most cases in terms of the average accuracy. Moreover, the CNN–SSAE achieves a much lower standard deviation value for the different initializations, indicating that the CNN–SSAE model depends less on the hyperparameter and performs more robustly on the different subjects.

VI. CONCLUSION AND FUTURE WORK

In this work, a novel neural network framework was proposed for decoding multi-channel MI-EEG signals into movement intention categories. A CNN based on different granular-grouped channels and a stacked sparse autoencoder (SSAE)-combined CNN method were applied to MI data for improving the discriminative and generalization capability of the MI-EEG decoding model. The datasets collected from our laboratory and from BCI Competition IV dataset 1 were used. Our experiments showed that the feature fusion network SSAE exhibits an accuracy of up to 86.41%. This inspires us to study the spatial feature and channel selection for MI-EEG decoding. Comparison and simulation experiments validate that the MI state of hand movement from different limbs could be classified with high accuracy using the proposed DL methods. In addition, several experiments were conducted to

verify the efficiency and computational complexity of the proposed DL models, particularly in terms of the training period. Finally, we found evidence that the generalization capability for different subjects can be improved by the CNN–SSAE.

However, we mostly focused on the results regarding invariance from spatial features and the across-subjects transfer learning capability of the models which ignore the feature from other domains for our future research directions. Currently, few studies have applied hybrid DL framework to BCI applications. This work paves the way for using hybrid DL methods (discriminative and representative DL) in practical MI-BCI systems and also can be applied to the other BCI paradigm such as P300 and SSVEP. Moreover, the proposed method can be further transferred to learning across-sessions representations to shorten user-target BCI system calibration periods.

REFERENCES

[1] A. Kumar, L. Gao, E. Pirogova, and Q. Fang, "A review of error-related potential-based brain–computer interfaces for motor impaired people," *IEEE Access*, vol. 7, pp. 142451–142466, 2019.

[2] A. R. Sereshkeh, R. Trott, A. Bricout, and T. Chau, "Online EEG classification of covert speech for brain–computer interfacing," *Int. J. Neural Syst.*, vol. 27, no. 8, Dec. 2017, Art. no. 1750033, doi: 10.1142/S0129065717500332.

[3] 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, doi: 10.1016/s1388-2457(02)00057-3.

[4] D. Jacobs, Y. H. Liu, T. Hilton, M. Del Campo, P. L. Carlen, and B. L. Bardakjian, "Classification of scalp EEG states prior to clinical seizure onset," *IEEE J. Transl. Eng. Health Med.*, vol. 7, pp. 1–3, Aug. 2019, Art. no. 2000203, doi: 10.1109/JTEHM.2019.2926257.

[5] T. Xu, R. Yin, L. Shu, and X. Xu, "Emotion recognition using frontal EEG in VR affective scenes," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Nanjing, China, May 2019, pp. 1–4, doi: 10.1109/IMBIOC.2019.8777843.

[6] O. P. Idowu, P. Fang, X. Li, Z. Xia, J. Xiong, and G. Li, "Towards control of EEG-based robotic arm using deep learning via stacked sparse autoencoder," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Kuala Lumpur, Malaysia, Dec. 2018, pp. 1053–1057, doi: 10.1109/ROBIO.2018.8665089.

[7] C. Demirel, H. Kandemir, and H. Kose, "Controlling a robot with extraocular muscles using EEG device," in *Proc. 26th Signal Process. Commun. Appl. Conf. (SIU)*, Izmir, Turkey, May 2018, pp. 1–4, doi: 10.1109/SIU.2018.8404157.

[8] S. Z. Diya, R. A. Prorna, I. I. Rahman, A. B. Islam, and M. N. Islam, "Applying brain-computer interface technology for evaluation of user experience in playing games," in *Proc. Int. Conf. Electr., Comput. Commun. Eng. (ECCE)*, Cox'sBazar, Bangladesh, Feb. 2019, pp. 1–6, doi: 10.1109/ECACE.2019.8679203.

[9] M.-A. Li, J.-F. Han, and L.-J. Duan, "A novel MI-EEG imaging with the location information of electrodes," *IEEE Access*, vol. 8, pp. 3197–3211, 2020, doi: 10.1109/ACCESS.2019.2962740.

[10] 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. 15, Jan. 2007, Art. no. 031005.

[11] X. Gu *et al.*, "EEG-based brain-computer interfaces (BCIs): A survey of recent studies on signal sensing technologies and computational intelligence approaches and their applications," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, pp. 1–22, 2020.

[12] M. Rawashdeh, M. G. AL Zamil, M. S. Hossain, S. Samarah, S. U. Amin, and G. Muhammad, "Reliable service delivery in tele-health care systems," *J. Netw. Comput. Appl.*, vol. 115, pp. 86–93, Aug. 2018, doi: 10.1016/j.jnca.2018.04.015.

[13] 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.

[14] M. T. F. Talukdar, S. K. Sakib, N. S. Pathan, and S. A. Fattah, "Motor imagery EEG signal classification scheme based on autoregressive reflection coefficients," in *Proc. Int. Conf. Informat., Electron. Vis. (ICIEV)*, Dhaka, Bangladesh, May 2014, pp. 1–4, doi: 10.1109/ICIEV.2014.6850812.

[15] 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.

[16] X. Zhang, "A survey on deep learning based brain computer interface: Recent advances and new frontiers," in *Proc. Annu. Conf. Innov. Technol. Comput. Sci. Educ. (ITiCSE)*, vol. 1, no. 1, 2020, p. 1291.

[17] 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.

[18] W. Ko, J. Yoon, E. Kang, E. Jun, J.-S. Choi, and H.-I. Suk, "Deep recurrent spatio-temporal neural network for motor imagery based BCI," in *Proc. 6th Int. Conf. Brain-Comput. Interface (BCI)*, GangWon, South Korea, Jan. 2018, pp. 1–3, doi: 10.1109/IWW-BCI.2018.8311535.

[19] R. Salakhutdinov, "Learning deep generative models," *Annu. Rev. Statist. Appl.*, vol. 2, no. 1, pp. 361–385, Apr. 2015.

[20] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.

[21] G. Dai, J. Zhou, J. Huang, and N. Wang, "HS-CNN: A CNN with hybrid convolution scale for EEG motor imagery classification," *J. Neural Eng.*, vol. 17, no. 1, Jan. 2020, Art. no. 016025.

[22] W. Xu and M. Zhang, "Theory of generative deep learning II: Probe landscape of empirical error via norm based capacity control," in *Proc. 5th IEEE Int. Conf. Cloud Comput. Intell. Syst. (CCIS)*, Nanjing, China, Nov. 2018, pp. 470–474, doi: 10.1109/CCIS.2018.8691394.

[23] X. Ma, S. Qiu, C. Du, J. Xing, and H. He, "Improving EEG-based motor imagery classification via spatial and temporal recurrent neural networks," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Honolulu, HI, USA, Jul. 2018, pp. 1903–1906, doi: 10.1109/EMBC.2018.8512590.

[24] Z. Huang, Y. Cao, and T. Wang, "Optimization of DBN network structure based on information entropy," *J. Phys., Conf. Ser.*, vol. 1176, Mar. 2019, Art. no. 032046.

[25] M. Längkvist, L. Karlsson, and A. Loutfi, "A review of unsupervised feature learning and deep learning for time-series modeling," *Pattern Recognit. Lett.*, vol. 42, pp. 11–24, Jun. 2014.

[26] Y. Xue and Y. Li, "A fast detection method via region-based fully convolutional neural networks for shield tunnel lining defects," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 33, no. 8, pp. 638–654, Aug. 2018.

[27] 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, doi: 10.1109/ACCESS.2019.2895688.

[28] X. Ma, S. Qiu, W. Wei, S. Wang, and H. He, "Deep channel-correlation network for motor imagery decoding from the same limb," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 1, pp. 297–306, Jan. 2020, doi: 10.1109/TNSRE.2019.2953121.

[29] Z. Tang, C. Li, and S. Sun, "Single-trial EEG classification of motor imagery using deep convolutional neural networks," *Optik*, vol. 130, pp. 11–18, Feb. 2017.

[30] N. Lu, T. Li, X. Ren, and H. Miao, "A deep learning scheme for motor imagery classification based on restricted Boltzmann machines," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 6, pp. 566–576, Jun. 2017, doi: 10.1109/TNSRE.2016.2601240.

[31] B. Yang, C. Fan, C. Guan, X. Gu, and M. Zheng, "A framework on optimization strategy for EEG motor imagery recognition," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 774–777, doi: 10.1109/EMBC.2019.8857672.

[32] A. Delorme and S. Makeig, "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9–21, Mar. 2004.

[33] G. Gomez-Herrero, W. De Clercq, H. Anwar, O. Kara, K. Egiazarian, S. Van Huffel, and W. Van Paesschen, "Automatic removal of ocular artifacts in the EEG without an EOG reference channel," in *Proc. 7th Nordic Signal Process. Symp. (NORSIG)*, Jun. 2006, pp. 130–133.

[34] R. K. Srinanthini, P. Srinivasan, and S. Arun, "Spectral analysis of EEG data for ocular artifact removal using wavelet transform technique," in *Proc. Int. Conf. Recent Adv. Energy-Efficient Comput. Commun. (ICRAECC)*, Nagercoil, India, Mar. 2019, pp. 1–4, doi: 10.1109/ICRAECC43874.2019.8995021.

[35] J. Yang, S. Yao, and J. Wang, "Deep fusion feature learning network for MI-EEG classification," *IEEE Access*, vol. 6, pp. 79050–79059, 2018.
[36] S. Min, B. Lee, and S. Yoon, "Deep learning in bioinformatics," *Brief Bioinform.*, vol. 18, no. 5, pp. 851–869, 2016.
[37] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.
[38] A. Reust, J. Desai, and L. Gomez, "Extracting motor imagery features to control two robotic hands," in *Proc. IEEE Int. Symp. Signal Process. Inf. Technol. (ISSPIT)*, Dec. 2018, pp. 118–122.
[39] Z. Tang, H. Yu, C. Lu, P. Liu, and X. Jin, "Single-trial classification of different movements on one arm based on ERD/ERS and corticomuscular coherence," *IEEE Access*, vol. 7, pp. 128185–128197, 2019.
[40] F. Fahimi, Z. Zhang, W. B. Goh, T.-S. Lee, K. K. Ang, and C. Guan, "Inter-subject transfer learning with an end-to-end deep convolutional neural network for EEG-based BCI," *J. Neural Eng.*, vol. 16, no. 2, Apr. 2019, Art. no. 026007, doi: 10.1088/1741-2552/aaf3f6.
[41] S. Saha, K. I. Ahmed, R. Mostafa, A. H. Khandoker, and L. Hadjileontiadis, "Enhanced inter-subject brain computer interface with associative sensorimotor oscillations," *Healthcare Technol. Lett.*, vol. 4, no. 1, pp. 39–43, Feb. 2017, doi: 10.1049/htl.2016.0073.
[42] Z. Hu, Y. Li, and Z. Yang, "Improving convolutional neural network using pseudo derivative ReLU," in *Proc. 5th Int. Conf. Syst. Informat. (ICSAI)*, Nanjing, China, Nov. 2018, pp. 283–287, doi: 10.1109/ICSAI.2018.8599372.
[43] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P. A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, pp. 3371–3408, Dec. 2010.
[44] A. Hassanpour, M. Moradikia, H. Adeli, S. R. Khayami, and P. Shamsinejadbabaki, "A novel end-to-end deep learning scheme for classifying multi-class motor imagery electroencephalography signals," *Expert Syst.*, vol. 36, no. 6, p. e12494, Dec. 2019.
[45] X. Qi, T. Wang, and J. Liu, "Comparison of support vector machine and softmax classifiers in computer vision," in *Proc. 2nd Int. Conf. Mech., Control Comput. Eng. (ICMCCE)*, Harbin, China, Dec. 2017, pp. 151–155, doi: 10.1109/ICMCCE.2017.49.
[46] S. Saha and M. Baumert, "Intra- and inter-subject variability in EEG-based sensorimotor brain computer interface: A review," *Frontiers Comput. Neurosci.*, vol. 13, p. 87, Jan. 2020, doi: 10.3389/fncom.2019.00087.
[47] Y. Zou, X. Zhao, Y. Chu, Y. Zhao, W. Xu, and J. Han, "An inter-subject model to reduce the calibration time for motion imagination-based brain-computer interface," *Med. Biol. Eng. Comput.*, vol. 57, no. 4, pp. 939–952, Apr. 2019, doi: 10.1007/s11517-018-1917-x.
[48] H. Li, J. Li, X. Guan, B. Liang, Y. Lai, and X. Luo, "Research on overfitting of deep learning," in *Proc. 15th Int. Conf. Comput. Intell. Secur. (CIS)*, Macao, Dec. 2019, pp. 78–81, doi: 10.1109/CIS.2019.00025.
[49] B. S. Sasikala, V. G. Biju, and C. M. Prashanth, "Kappa and accuracy evaluations of machine learning classifiers," in *Proc. 2nd IEEE Int. Conf. Recent Trends Electron., Inf. Commun. Technol. (RTEICT)*, Bengaluru, India, May 2017, pp. 20–23, doi: 10.1109/RTEICT.2017.8256551.
[50] S. M. Christensen, N. S. Holm, and S. Puthusserypady, "An improved five class MI based BCI scheme for drone control using filter bank CSP," in *Proc. 7th Int. Winter Conf. Brain-Comput. Interface (BCI)*, Gangwon, South Korea, Feb. 2019, pp. 1–6.
[51] P. N. Paranjape, M. M. Dhabu, P. S. Deshpande, and A. M. Kekre, "Cross-correlation aided ensemble of classifiers for BCI oriented EEG study," *IEEE Access*, vol. 7, pp. 11985–11996, 2019.
[52] X. Ma, S. Qiu, C. Du, J. Xing, and H. He, "Improving EEG-based motor imagery classification via spatial and temporal recurrent neural networks," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Honolulu, HI, USA, Jul. 2018, pp. 1903–1906, doi: 10.1109/EMBC.2018.8512590.
[53] 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.
[54] S. M. Christensen, N. S. Holm, and S. Puthusserypady, "An improved five class MI based BCI scheme for drone control using filter bank CSP," in *Proc. 7th Int. Winter Conf. Brain-Comput. Interface (BCI)*, Gangwon, South Korea, Feb. 2019, pp. 1–6.