

# NeuroGrasp: Real-Time EEG Classification of High-Level Motor Imagery Tasks Using a Dual-Stage Deep Learning Framework

Jeong-Hyun Cho<sup>ID</sup>, Ji-Hoon Jeong<sup>ID</sup>, Associate Member, IEEE, and Seong-Whan Lee<sup>ID</sup>, Fellow, IEEE

**Abstract**—Brain–computer interfaces (BCIs) have been widely employed to identify and estimate a user’s intention to trigger a robotic device by decoding motor imagery (MI) from an electroencephalogram (EEG). However, developing a BCI system driven by MI related to natural hand-grasp tasks is challenging due to its high complexity. Although numerous BCI studies have successfully decoded large body parts, such as the movement intention of both hands, arms, or legs, research on MI decoding of high-level behaviors such as hand grasping is essential to further expand the versatility of MI-based BCIs. In this study, we propose NeuroGrasp, a dual-stage deep learning framework that decodes multiple hand grasping from EEG signals under the MI paradigm. The proposed method effectively uses an EEG and electromyography (EMG)-based learning, such that EEG-based inference at test phase becomes possible. The EMG guidance during model training allows BCIs to predict hand grasp types from EEG signals accurately. Consequently, NeuroGrasp improved classification performance offline, and demonstrated a stable classification performance online. Across 12 subjects, we obtained an average offline classification accuracy of 0.68 ( $\pm 0.09$ ) in four-grasp-type classifications and 0.86 ( $\pm 0.04$ ) in two-grasp category classifications. In addition, we obtained an average online classification accuracy of 0.65 ( $\pm 0.09$ ) and 0.79 ( $\pm 0.09$ ) across six high-performance subjects. Because the proposed method has demonstrated a stable classification performance when evaluated either online or offline, in the future, we expect that the proposed method could contribute to different BCI applications, including robotic hands or neuroprosthetics for handling everyday objects.

**Index Terms**—Brain–computer interface (BCI), deep learning, electroencephalogram, high-level tasks, motor imagery (MI), real-time classification.

Manuscript received 24 November 2020; revised 21 March 2021, 27 June 2021, and 30 August 2021; accepted 12 October 2021. Date of publication 8 November 2021; date of current version 18 November 2022. This work was supported in part by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant, funded by the Korean Government under Grant 2017-0-00432, Grant 2017-0-00451, and Grant 2019-0-00079. This article was recommended by Associate Editor P. P. Angelov. (Jeong-Hyun Cho and Ji-Hoon Jeong contributed equally to this work.) (Corresponding author: Seong-Whan Lee.)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board at Korea University under Application No. 1040548-KU-17-172-A-21 and performed in line with the Declaration of Helsinki.

Jeong-Hyun Cho is with the Department of Brain and Cognitive Engineering, Korea University, Seoul 02841, South Korea (e-mail: jh\_cho@korea.ac.kr).

Ji-Hoon Jeong and Seong-Whan Lee are with the Department of Artificial Intelligence, Korea University, Seoul 02841, South Korea (e-mail: jh\_jeong@korea.ac.kr; sw.lee@korea.ac.kr).

Digital Object Identifier 10.1109/TCYB.2021.3122969

## I. INTRODUCTION

BRain-COMPUTER interfaces (BCIs) represent one of the novel approaches to translating human intentions into external device commands through brain activity. BCI systems have been developed to exploit different types of paradigms using electroencephalogram (EEG) signals, such as a powered exoskeleton, drone, and robotic arm [1]–[5]. Motor imagery (MI) is one of the most common paradigms using EEG, and recent studies on MI-based BCIs (MI-BCIs) have proven their potential for facilitating the reliable control of external devices, such as robotic arms and neural prostheses [6]–[9]. These advances in MI-BCIs, which have allowed users to communicate with external devices through thoughts rather than the peripheral nervous system, have demonstrated an impressive capacity for enhancing the prosthetic performance, which helps improving day to day tasks in the user’s life.

Natural grasp tasks are associated with more dynamic brain activity than the movements of other extremities because a large area of the human brain’s motor cortex is allocated to controlling the hands [6], [10], [11]. In particular, decoding grasp actions within a single arm is more difficult and complicated than decoding the conventional MI related to the movements of other body parts [12], [13]. Therefore, we consider high-level MI for natural grasp actions. Recently, numerous studies have attempted to decode EEG signals related to actual movements or motor intentions using diverse approaches; consequently, the more significant advances in technology have inspired our present research. Recent advanced studies on the use of MI-BCIs have been aimed at human-inspired artificial intelligence (AI) robots using intuitive upper extremity movement [14], [15]; several of these have investigated the decoding of the combinations of hand, wrist, and arm movements [7]. Hence, in this article, we focus on decoding the movement intention of natural grasp types (e.g., cylindrical, spherical, lateral, and pincer grasps) and categories (e.g., power and precision grasps) from the EEG signals.

Conventional BCI studies concerning motor activity have been undertaken under different EEG paradigms [1]–[3], [16], [17]. For example, some research groups have conducted experiments to classify the different upper limb movements present within a single arm. Schwarz *et al.* [6] attempted to decode natural reach-and-grasp actions from human EEG signals. They attempted to identify three executed reach-and-grasp actions, namely, lumbrical, pincer, and palmar

grasp, using EEG neural correlates. Ofner *et al.* [7] encoded single upper limb movements into the time domain of low-frequency EEG signals. The primary goal of their experiment was to classify six different actions: 1) elbow flexion; 2) elbow extension; 3) hand grasp; 4) hand spread; 5) supination; and 6) pronation.

Similarly, decoding the different movements of the upper extremities in a single arm is important, and EEG-based grasp task decoding is one of the significant challenging issues for BCI studies. The aforementioned studies were carefully executed regarding the user's motor intention; however, a significant problem was the unsatisfactory decoding performance, which predominantly existed when tested in the MI paradigm. To overcome this limitation, other studies combined different EEG paradigms with MI to improve the performance of MI-BCIs; subsequently, this hybrid approach demonstrated improved results in several cases. Unfortunately, the hybrid methods sacrificed their intuitiveness-of-control because supplemental EEG paradigms were induced by exogenous stimuli [18]–[21]. On the contrary, researchers have designed a way to use EEG and electromyography (EMG) signals together to develop another type of hybrid BCIs. As a representative example, Özdenizci *et al.* proposed a novel hierarchical graphical model-based context-aware hybrid brain-machine interface. They used probabilistic fusion of EEG and EMG activities to classify the EEG data into correct classes according to hierarchical and stochastic methodologies. This strategy improved the classification performance of EEG-based interfaces close to the level of EMG-based interfaces, which typically show the highest performance [22]. López-Larraz *et al.* proposed a decoding method to detect movement attempts based on features extracted from EEG and EMG signals. Their main idea was to use EMG activities as complementary information of EEG to detect movement intentions accurately. As a result, this EEG-EMG hybrid BCI showed significantly higher classification accuracy than EEG-or EMG-based comparable methods [23].

To develop robust BCIs driven by pure MI, numerous contemporary studies have focused on using advanced machine learning and deep learning techniques [9], [24]–[28]. For example, Zhang *et al.* [29] proposed a temporally constrained sparse group spatial patterns (TSGSPs) method that improves the classification accuracy of MI-BCIs. TSGSP demonstrated the feasibility of solving a problem inherent in the conventional filter bank common spatial patterns (CSPs) that cannot extract enough information in the temporal domain and showed a robust feature extraction. In addition, deep learning is a noteworthy option for improving the performance of MI-BCIs. Two approaches have emerged from the combination of deep learning and BCIs. One approach focuses on designing enhanced neural network architectures [30]–[33]; others focus on determining a method to develop a novel form of input data for networks [34]–[36]. Numerous studies have attempted to increase the performance by tuning the deep learning network structure, and many have achieved notable classification performance [37]–[39]. Moreover, as part of another approach to high-performance BCI development, other researchers are focusing on developing subject-independent

decoding models based on transfer learning and adversarial inference learning from EEG signals [40]–[42]. However, convolutional neural network (CNN)-based deep learning could not achieve an effect on processing EEG data that it has with image data. Moreover, long short-term memory (LSTM) and other methods for sequence-to-sequence processing are not effective for EEG analysis.

In this study, we introduce a novel approach based on muscle synergies, which represent the contribution of each muscle during a specific movement to construct the novel input image generation. This approach can be defined as using a kind of EMG guidance; therefore, we present a method that effectively uses the guidance during model training so that EEG can be used to predict hand grasp types at test time. As a result, EEG information is implicitly mapped onto EMG-based muscle synergy features. Under this background, the proposed method uses an EEG and EMG-based learning, such that EEG-based inference at the test phase becomes possible. Hence, we present a dual-stage deep learning framework called NeuroGrasp to improve the classification performance and reliability of online BCIs related to decoding high-level MI tasks.

The main contributions of this study are as follows.

- 1) We develop a novel deep learning-based framework that can classify high-level MI for natural grasp motions. This framework demonstrates a higher classification performance than other comparable methods. Moreover, the proposed approach can achieve stable results in real-time EEG classification.
- 2) We present an approach that can validate if the EEG signals evoked in the MI experiment are actually produced by the relevant muscle movement imagery. By including the process of creating a kind of muscle kinematic pattern by estimating muscle synergies in the operation of the dual-stage framework, we can visually determine whether the decoding EEG can successfully describe and reconstruct the movement of the related muscles.
- 3) We perform experiments for decoding high-level MI using real-time EEG classification. The proposed framework demonstrates a comparatively high decoding performance and is sufficiently light to complete the training and test phases in real-time decoding. Consequently, this could be a preliminary process for the further study of BCI applications such as the control of a robotic hand.

The remainder of this article is organized as follows. In Section II, we introduce our BCI experiments and the structure of the proposed method more in detail. In Section III, we present the experimental results of both offline and online sessions across subjects. In Section IV, we suggest the directions for advanced MI-BCIs using NeuroGrasp, and compare the performance among several state-of-the-art methods. Finally, we discuss the conclusion of this article in Section V.

## II. MATERIALS AND METHODS

### A. Participants

Twelve voluntary subjects (Sub01–Sub12; aged 20–34; all right-handed) who were naïve BCI users participated in the

experiments. The subjects were all healthy volunteers without any known neurophysiological anomalies or musculoskeletal disorders. Before the experiment, the volunteers were informed of the experimental protocols, paradigms, and purpose. Once they confirmed that they understood the experiment, each subject provided written consent according to the Declaration of Helsinki. The experimental protocols and environments were reviewed and approved by the Institutional Review Board at Korea University [1040548-KU-17-172-A-21].

### B. Experimental Setup

During a session of the experimental protocol, the subjects were seated in a comfortable chair in front of a 24-inch LCD monitor display. The display was installed horizontally on the table to ensure that the subjects could see the objects and visual cues without moving their heads [6]. Four designated objects (a cup, ball, card, and plastic bean) were placed above the display and a visual cue (a flashing arrow pointing to the targeted object) indicated what action the subject should perform. The subjects were asked to perform MI after the visual cue (e.g., cylindrical, spherical, lateral, and pincer grasps). The locations of the objects were randomly changed to avoid the influence of artifacts. An operator sitting behind the subject minimized the subjects' unnecessary movements by cautioning or penalizing them; therefore, the operator could prevent the subjects from creating unwanted noise that reduces data quality.

### C. Experimental Protocol

Before the experiment, when the operators inserted the conductive gel on the scalp, the subjects were asked to perform rehearsal of the MI tasks. The subjects practiced the tasks 5~10 times for each class. Subsequently, in the experiment, each subject performed 50 trials per grasp. We asked the subjects to imagine a specific grasping action only once during a MI period of 4 s, as indicated in Fig. 1(b). Therefore, 200 trials (4 classes  $\times$  50 trials) were obtained from one subject. Through this experimental protocol, a balanced dataset was prepared for each grasp class. A single trial comprised three continuous stages, which posed a designated task to the subjects. These stages included rest, preparation, and MI. The duration of a single trial was 10 s and contained three sub-stages: 3, 3, and 4 s in duration. The visual cue was already presented in the 3 s preparation stage before MI. Therefore, the subjects could start MI immediately without delay, as shown in Fig. 1(b).

### D. Data Acquisition and Preprocessing

EEG data were collected at 1000 Hz using 20 Ag/AgCl electrodes (FC1–6, C1–6, Cz, CP1–6, and CPz) in the 10/20 international system via BrainAmp (BrainProduct GmbH, Germany) [43], [44]. A 60-Hz notch filter was used to remove the power frequency interference. The FCz and FPz were used as the reference and ground electrodes, respectively. The impedance was maintained below 10 k $\Omega$ . The 20 channels were located only on the motor cortex to ensure that the recorded EEG signals corresponded to the motor-related potentials [3]. The EEG data were bandpass filtered in the

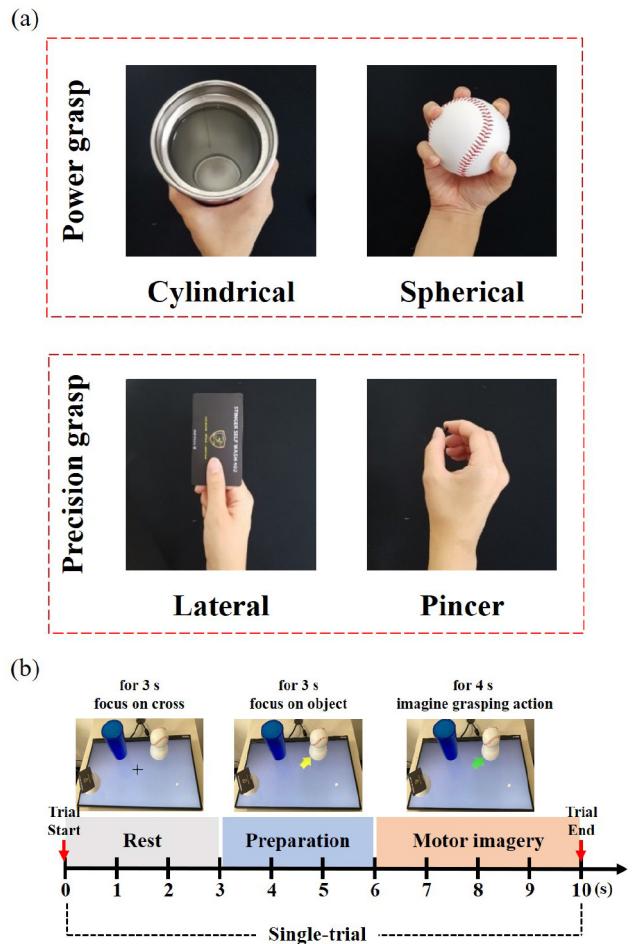


Fig. 1. Experimental tasks and protocol. (a) Tasks consisted of four hand-grasp types for different objects, such as a cup, ball, card, and bolt; cylindrical, spherical, lateral, and pincer grasps. The grasp types were classified into two grasp categories: 1) power and 2) precision grasp. (b) Experimental protocol of a single trial. The duration of a single trial was 10 s, wherein the subjects performed MI during the designated stage for 4 s after a visual cue was provided.

range [4–40] Hz [45], [46] using Hamming-windowed zero-phase finite impulse response (FIR) filters with an optimized order ( $N = 50$ ) [47]. We performed [4–40] Hz bandpass filtering and discarded all low-frequency relevant information because when bandpass filtering was performed with wide bandwidths of [4–40] Hz, the learning effect of deep learning was more pronounced than that of low-frequency bands. Because the proposed model was learned by matching EEG signals and the EMG-guidance in architecture, the learning effect did not appear when using simple brain signal patterns in low-frequency bands such as movement-related cortical potentials. A deep learning architecture was designed to use the data from only 20 EEG channels that were related to body movement located on the motor cortex, as its input. The EEG data were downsampled from 1000 to 100 Hz.

For kinematic data, we recorded surface EMG signals to detect muscle activation. The EMG data were obtained using six Ag/AgCl electrodes and a digital amplifier, which was the same equipment used to record the EEG signals. We acquired the EMG signals for each grasping action independently before the MI experiment. The actual movement data in the right

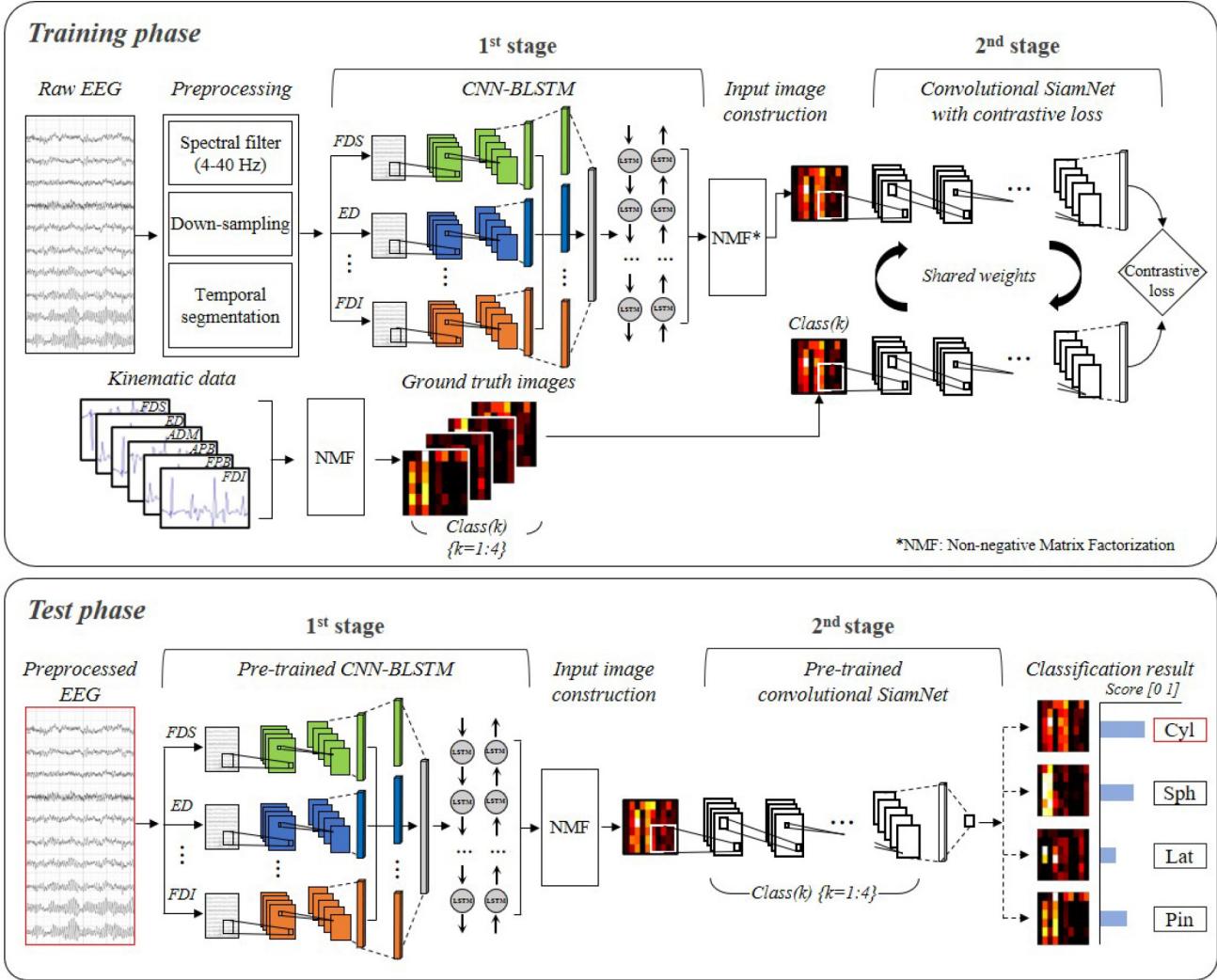


Fig. 2. Overall flowchart of proposed framework configured with dual-stage deep learning. The first stage creates SAC-based muscle patterns using CNN-BLSTM; the second stage uses this information in the form of an image to classify into one of the grasp classes.

arm were recorded from six muscles. The details of the corresponding muscles are as follows: flexor digitorum superficialis (FDS), extensor digitorum (ED), abductor digiti minimi (ADM), abductor pollicis brevis (APB), flexor pollicis brevis (FPB), and first dorsal interosseous (FDI) [48], [49]. The signals were filtered using a [60–500] Hz bandpass filter [48].

#### E. Proposed Method

As a decoding method, we employed a dual-stage deep learning framework called NeuroGrasp. This framework consisted of two independent deep learning architectures. CNN-bidirectional LSTM (CNN-BLSTM) [28] configured the first stage and convolutional *SiamNet* [50], [51] configured the second stage. Specifications of the neural networks are described in Table I. In addition, the proposed framework consisted of training and test phases. Fig. 2 displays a flowchart of the framework that constructs the input images from six different CNN-BLSTM networks and then uses the input images to train the second-stage network. In the training phase, we trained the two deep learning architectures for the classification of

MI by estimating the muscle synergy activation coefficient (SAC) [48], [52], [53] and converting a group of the SACs into an image that represented a kinematic representation related to a specific grasp class, as described in Algorithm 1. In the test phase, we reconstructed the SAC by decoding the EEG signals using trained CNN-BLSTM. The second-stage for the final classification process was trained with the SAC images collected during the training phase. The trained *SiamNet* was applied in the test phase to classify the input image into the correct class.

1) *First-Stage Architecture*: In the first stage, each CNN used the corresponding signal information from the six muscles as a label to extract features from the EEG. The baseline was determined as an averaged value of rectified EMG signals using root mean square (RMS) to create a binary label representing muscle activation at a specific time interval. Subsequently, we could prepare the binary label in advance, which represents the muscle activation in a particular time interval by designating a positive value when the rectified and averaged value is greater than the baseline and a negative value when it is less than the baseline [48], [54]. As illustrated

TABLE I  
SPECIFICATIONS OF THE NEURAL NETWORK: DETAILS OF PARAMETERS AND LAYER IMPLEMENTATIONS

Stage	Layer	Type	Parameter	Output size	Activation
I	1	Input	-	$1 \times 20 \times 400$	
	2	Convolution	Filter size: $1 \times 10$ Stride size: $1 \times 1$ Kernels: 25	$25 \times 20 \times 391$	ReLU
		BatchNorm	-		
	3	Convolution	Filter size: $20 \times 1$ Stride size: $1 \times 1$ Kernels: 25	$25 \times 1 \times 391$	ReLU
		BatchNorm	-		
	4	Average pooling	Filter size: $1 \times 3$ Stride size: $1 \times 3$	$25 \times 1 \times 130$	
	5	Convolution	Filter size: $1 \times 10$ Stride size: $1 \times 1$ Kernels: 50	$50 \times 1 \times 121$	ReLU
		BatchNorm	-		
	6	Average pooling	Filter size: $1 \times 3$ Stride size: $1 \times 3$	$50 \times 1 \times 40$	
	7	Fully connected	-	$1 \times 200$	Softmax
II	8	BLSTM	Hidden units: 400	$6 \times 400$	
	1	Input	-	$1 \times 28 \times 28$	
	2	Convolution	Filter size: $3 \times 3$ Stride size: $1 \times 1$ Kernels: 10	$10 \times 28 \times 28$	ReLU
		BatchNorm	-		
	3	Convolution	Filter size: $3 \times 3$ Stride size: $1 \times 1$ Kernels: 14	$14 \times 28 \times 28$	ReLU
		BatchNorm	-		
	4	Convolution	Filter size: $3 \times 3$ Stride size: $1 \times 1$ Kernels: 18	$18 \times 28 \times 28$	ReLU
		BatchNorm	-		
	5	Fully connected	-	$1 \times 256$	
	6	Fully connected	-	$1 \times 1$	Sigmoid

in Fig. 2, each CNN in the first-stage extracts convolution features from segmented EEG using the binary labels. Using the CNN architecture, the network was trained and the EEG signals for each corresponding muscle could then represent the positive or negative labels using the baseline EMG signals according to the activation. Each CNN extracts coarse features by considering the activation of each muscle.

The input data for training and test phases were EEG signals with a fixed size of  $20 \times 400$  (channel  $\times$  time sample), as shown in Table I. The data were sequentially passed through two convolution layers, one had a  $1 \times 10$  kernel size as a temporal filter and the other had a  $20 \times 1$  kernel size as a spatial filter. An average pooling layer was then used to resize the convolution features, which had a  $1 \times 3$  kernel size with a stride of  $1 \times 3$ . Another pooling layer, which had the same specification, was applied after the last convolution layer with a  $1 \times 10$  kernel size. Using the muscle activation information as labels, the trained CNN extracted 200 convolution features from each data segment. In a manner that did not violate the information of the temporal domain, the convolution features extracted from the segment were connected and reconstructed into a vector of size  $1 \times 200$ .

Finally, each of the 200 features from the activation of the six muscles became an input feature of each cell of the BLSTM network ( $6 \times 200$ ). A sequence of the features represented in  $x_s$  ( $s = SAC$  of the muscle synergy 1-6) and the output of the BLSTM was trained recursively by activating

the units in the network with the following equations for each activation coefficient from the six muscles:

$$i_s = \sigma(W_{xi}x_s + W_{hi}h_s + W_{ci}c_s + b_i) \quad (1)$$

$$f_s = \sigma(W_{xf}x_s + W_{hf}h_s + W_{cf}c_s + b_f) \quad (2)$$

$$g_s = \varphi(W_{xg}x_s + W_{hg}h_s + b_g) \quad (3)$$

$$c_s = f_s \odot c_s + i_s \odot g_s \quad (4)$$

$$o_s = \sigma(W_{xo}x_s + W_{ho}h_s + W_{co}c_s + b_o) \quad (5)$$

$$h_s = o_s \odot \varphi(c_s) \quad (6)$$

where  $h_s$  are the hidden vectors with the subscription  $s$  and  $i_s$ ,  $f_s$ ,  $c_s$ , and  $o_s$  denote the activation vectors of the input gate, forget gate, memory cell, and output gate, respectively.  $\sigma$  is the sigmoid function defined as  $\sigma_x = 1$  and  $\varphi$  is the hyperbolic tangent function. The variables  $W$  and  $b$  are the weight matrices and biases, respectively, which must be trained. Feature vectors were extracted according to the time dependence using the BLSTM network. Each of the 200 extracted features from the kinematic data of the six muscles was input into six memory cells in a forward and backward LSTM layer. Then, the BLSTM network was trained based on the coarse-to-fine approach. The features of the FDS appeared at the beginning of the forward LSTM training, and the properties from the ED to the FDI were sequentially trained into memory cells. The backward LSTM layer trained the features of each EMG signal in the same manner, but in the reverse order.

The BLSTM network can be trained using six EMG signals from the corresponding muscles to predict the SAC. The forward and backward LSTM layers were set at a dropout ratio of 0.5. Each LSTM cell retrieved 400 hidden units; therefore, the output of each forward and backward LSTM had a size of  $6 \times 400$ , which included all of the kinematic data. A fully connected layer with a rectified linear unit was adopted between the connected layers to extract nonlinear features. Finally, the estimated activation coefficient for each synergy was predicted using a regression layer. We performed 500 iterations for the model training process and saved the model weights and hyperparameters that produced the least test data loss.

2) *Input Image Construction:* Input image construction is an embedding process to generate the inputs for the second stage architecture using the outputs from the first stage. The preprocessed EMG and estimated kinematic data were used to calculate synergies using non-negative matrix factorization (NMF), which is a common matrix factorization algorithm that has been used in related studies for decades [55], [56]. Muscle synergy and its activation coefficient have been assumed to be the components used by the central nervous system to generate motions. According to this, the accomplishment of different motor tasks relies on the ability of the motor system to recruit a small set of synergies on a single-trial basis and combine them in a task-dependent manner [57].

Briefly, this algorithm for muscle synergy extraction ( $W$ ) and the relative activation of those synergies ( $H$ ) from the original matrix ( $V$ ) of the EMG data is such that

$$V^{N \times M} \approx W^{N \times R} \cdot H^{R \times M} \quad (7)$$

where  $V$  is the original data matrix with  $N$  as the number of muscles,  $M$  as the number of time samples,  $W$  is the synergy

**Algorithm 1:** Training Procedure of NeuroGrasp

- 
- Input:** Preprocessed EEG data and kinematic ground truth
- $X = \{x_i\}_{i=1}^D$ ,  $\{x_i\} \in \mathbb{R}^{C \times T}$ : motor imagery training data, where  $D$  is total number of trials with  $C$  channels and  $T$  sample points
  - (1)  $Y = \{y_i\}_1^G$ ,  $\{y_i\} \in \mathbb{R}^{C \times T}$ : kinematic ground truth (EMG data), where  $G$  is corresponding class with  $C$  channels and  $T$  time points
- Output:** Predicted class label of  $X$
- Step 1:** Train CNN using muscle activation labels
- Input  $X_{tr}$ : a training set of EEG data
  - Input  $\{Y_i\}_{i=1}^E$ : a binary label corresponding to the  $X_{tr}$ , where  $E$  is total number of muscles; calculated from (1)
  - Output  $X_N$ : features extracted by convolution layers
- Step 2:** Train BLSTM with synergy activation coefficient
- (2) Input  $X_N$ : a set of features from trained CNN
  - (3) Input  $Y_N$ : synergy activation coefficient as labels which were calculated at (1)
  - (4) Output  $\widehat{Y}_N$ : predicted synergy activation coefficient
  - Generate loss value by calculating differences between  $\widehat{Y}_N$  and  $Y_N$  using (2) to construct images for the input in Step 4
- Step 3:** Resize constructed images
- $I_N$ : a set of images by resizing (3)
  - $I_{tr}$ : a set of images by resizing (4)
- Step 4:** Train convolutional *SiamNet*
- $I_N$  and  $I_{tr}$  inputs for convolutional layers
  - Initialize the parameters of convolutional layers to random values
  - Generate loss value by calculating differences between convolutional layers output of  $I_{tr}$  and ground truth  $I_N$  using contrastive loss
  - Rank the similarity scores by the representative image  $I_N$  for each class
- 

matrix, and  $H$  is the coefficient matrix.  $W$  is  $N \times R$ , which is a matrix with  $R$  synergies,  $N$  is the number of muscles, and  $H$  is an  $R \times M$  matrix with  $R$  synergies and  $M$  is the number of time samples. Thus, each row of  $H$  represents the extent to which the corresponding synergy is activated or used to generate force, simply the activation level of each synergies over a single grasping action [58].

3) *Second-Stage Architecture*: As the first-stage architecture constructed an image from the estimated SAC, the second-stage architecture is to classify the images into one of the four classes. In the second stage, the convolution layers extract coarse features by considering the estimated SAC. The SAC image was resized to  $28 \times 28$  before being input into *SiamNet*. This process ensures that the vertical- and horizontal-axes lengths of the data matrix are the same so that features extracted through the convolution layer are not biased in a particular domain. The purpose of the resizing was to ensure that the spatial and temporal features of the synergies and activation coefficients were equally employed by equalizing the size of the vertical and horizontal axes of the image data.

The *SiamNet* had the same structure as the one above, as indicated in Fig. 2. We can obtain distinct features for two input data through the convolution layers of the *SiamNet*, and the L1 vector means a vector that represents the difference between the extracted features as absolute distance (the L1 norm). In the equation below,  $\alpha_j$  is a parameter that is automatically learned in the training process, multiplied by the distance value for each feature.  $\sigma$  is a sigmoid active function

$$\sigma \left( \sum_j \alpha_j |h_{1,L-1}^{(j)} - h_{2,L-1}^{(j)}| \right). \quad (8)$$

The features extracted through the CNN are passed through the fully connected layer to be used for determining a single value using a sigmoid function. If the predicted value is one, it is the same class; if the predicted value is zero, it is another class. The function of this next step is to obtain the loss using binary cross-entropy for the predicted value and true label

$$L(x_1, x_2, t) = t \cdot \log(p(x_1 \circ x_2)) + (1 - t) \cdot \log(1 - p(x_1 \circ x_2)) + \lambda \cdot \|w\|_2. \quad (9)$$

The loss function used in the learning process is described above. The test image created through the first stage is fed into the network as an input, and the ground-truth image is fed into the other network as a comparison target to train the network. These networks classify the images with the highest probability values into the same class. These characteristics of *SiamNet* were used to classify the four classes. Comparing a certain test image with four ground-truth images, which are sequentially inserted one after another, resulted in similarity scores based on the probability of similarity. These scores were compared to estimate the correct class by finding a test image that is most similar to the ground-truth image. For example, if the presented test image is probabilistically similar to the first of the four ground-truth images compared sequentially, it is classified as Class 1 (Cyl). The equation is as follows:

$$C(\hat{x}, S) = \operatorname{argmax}_c P(\hat{x} \circ x_c), x_c \in S. \quad (10)$$

We first calculate the L1 vector of the two reflections that have passed through the CNN for the two images. In short, to implement *SiamNet* for the second-stage architecture, we calculate the L1 vector of the two reflections that have passed through the CNN for the two images. Then, the L1 vector is passed through the hidden layer and a sigmoid conversion is performed on the output layer. We used a loss function based on binary cross-entropy for model training. The models export outputs with values between zero and one. This value is called similarity because the larger value, the more similar the image is. We used this similarity to perform the final classification.

**F. Evaluation**

1) *Offline Session*: The offline session was performed to compare the classification accuracy of the proposed method against those of comparable methods and evaluate the capacity of the proposed method to provide performance enhancements. The EEG data consisted of 50 trials per class. The classification results were evaluated using the  $10 \times 5$ -fold cross-validation. The four grasp-type classification used 50 independent trials per class to train the model. The two-grasp category classification was the result of new training and testing of the model by combining data from the cylindrical (Cyl) and the spherical (Sph) classes into a single power grasp class and combining the lateral (Lat) and pincer (Pin) classes into the precision grasp class.

Comparable methods were selected as representative algorithms in conventional BCI studies: CSPs and LDA (CSP-LDA) [59], ShallowConvNet [60], DeepConvNet [60], and EEGNet [61]. We conducted a comparative evaluation at the model level on how well the data were classified using

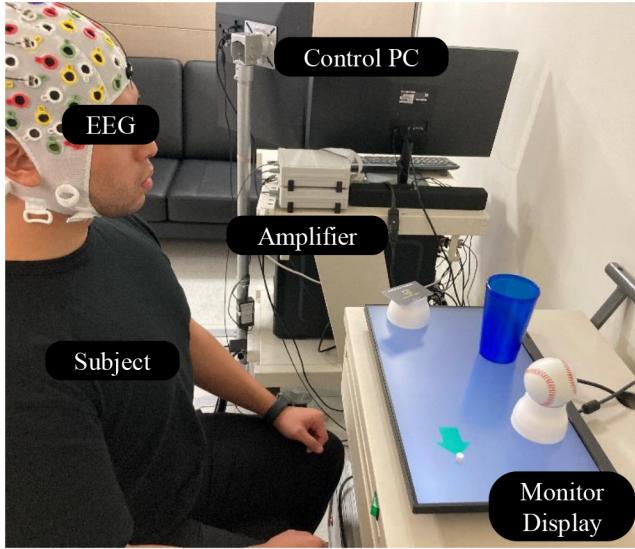


Fig. 3. Experimental setup and environment for online session. The control PC decodes EEG signals in real time and presents visual feedback of the classification result on the monitor display. Offline session is performed in the same experimental environment and setting.

the same input data. As a result, the proposed NeuroGrasp showed better classification performance than other models. In the study concerning EEG decoding based on deep learning, the most comparable models are EEGNet and DeepConvNet, and we have assigned the labels to 1 (Cyl), 2 (Sph), 3 (Lat), and 4 (Pin) in the classification layer. NeuroGrasp requires EMG data in the training phase, but like other comparable models, once the model is trained, it can produce classification results only via EEG data without the need to acquire additional EMG data in the test phase. Hence, we believe that the comparison should be with the above-mentioned models to verify the performance and stability of the proposed model before moving on to the online experiment.

2) *Online Session*: Fig. 3 describes the experimental setup for the online session. The calibration for the training data acquisition and real-time decoding tests was performed in the above environment. Similar to the offline experiment, subjects were asked to imagine the four grasp motions. The number of trials was also the same as in the offline session. Fifty trials were collected for each class ( $50 \times 4 = 200$  trials). Accordingly, the calibration data were recorded and the proposed model was retrained before running the online session. In addition, the proposed framework was trained using data from an offline session, and then the framework was tested using the online session data. Offline and online sessions were conducted on different recording days to confirm the effectiveness of the session-independent method using the proposed model. After the calibration time, the test trial consisted of 24 trials ( $12 \text{ trials} \times 2 \text{ runs}$ ). The 12 trials in one run could obtain results when four classes were tried three times, as depicted in Table IV. In the binary case, 12 trials in the run consisted of two classes and six times each. We then provided a run twice to help the subject adapt to the real-time experiment.

TABLE II  
PERFORMANCE EVALUATION RESULTS ON OFFLINE SESSION.  
GRAND-AVERAGED CLASSIFICATION ACCURACY OF FOUR GRASP TYPES  
USING THE PROPOSED AND COMPARABLE METHODS

Subject	10×5-fold cross-validation				
	CSP+LDA	ShallowNet	DeepConvNet	EEGNet	NeuroGrasp
Sub01	0.46(±0.09)	0.61(±0.11)	0.68(±0.10)	0.72(±0.08)	0.74(±0.08)
Sub02	0.38(±0.06)	0.46(±0.05)	0.44(±0.05)	0.55(±0.09)	0.78(±0.11)
Sub03	0.45(±0.12)	0.50(±0.07)	0.52(±0.08)	0.60(±0.11)	0.71(±0.12)
Sub04	0.27(±0.02)	0.46(±0.03)	0.42(±0.11)	0.43(±0.08)	0.65(±0.09)
Sub05	0.42(±0.09)	0.63(±0.15)	0.62(±0.11)	0.63(±0.07)	0.60(±0.07)
Sub06	0.48(±0.11)	0.71(±0.09)	0.72(±0.08)	0.76(±0.08)	0.79(±0.10)
Sub07	0.34(±0.04)	0.61(±0.06)	0.59(±0.06)	0.67(±0.09)	0.71(±0.13)
Sub08	0.28(±0.01)	0.45(±0.13)	0.50(±0.08)	0.71(±0.04)	0.80(±0.09)
Sub09	0.49(±0.10)	0.41(±0.12)	0.41(±0.11)	0.44(±0.12)	0.64(±0.09)
Sub10	0.45(±0.06)	0.59(±0.08)	0.58(±0.07)	0.49(±0.06)	0.69(±0.09)
Sub11	0.39(±0.02)	0.61(±0.06)	0.64(±0.12)	0.54(±0.09)	0.57(±0.10)
Sub12	0.33(±0.03)	0.35(±0.02)	0.36(±0.02)	0.39(±0.03)	0.51(±0.09)
Mean	<b>0.40(±0.08)</b>	<b>0.53(±0.11)</b>	<b>0.54(±0.12)</b>	<b>0.58(±0.12)</b>	<b>0.68(±0.09)</b>

TABLE III  
PERFORMANCE EVALUATION RESULTS ON OFFLINE SESSION.  
GRAND-AVERAGED CLASSIFICATION ACCURACY OF TWO-CLASS GRASP  
CATEGORIES USING PROPOSED AND COMPARABLE METHODS

Subject	10×5-fold cross-validation				
	CSP+LDA	ShallowNet	DeepConvNet	EEGNet	NeuroGrasp
Sub01	0.61(±0.09)	0.79(±0.12)	0.82(±0.08)	0.88(±0.11)	0.86(±0.09)
Sub02	0.66(±0.11)	0.77(±0.09)	0.72(±0.08)	0.69(±0.06)	0.89(±0.11)
Sub03	0.61(±0.07)	0.71(±0.05)	0.69(±0.09)	0.73(±0.08)	0.81(±0.11)
Sub04	0.54(±0.09)	0.73(±0.12)	0.74(±0.07)	0.77(±0.06)	0.90(±0.09)
Sub05	0.63(±0.08)	0.79(±0.10)	0.69(±0.12)	0.73(±0.07)	0.86(±0.09)
Sub06	0.60(±0.09)	0.74(±0.11)	0.75(±0.04)	0.78(±0.06)	0.88(±0.08)
Sub07	0.55(±0.06)	0.80(±0.12)	0.81(±0.11)	0.73(±0.05)	0.87(±0.06)
Sub08	0.62(±0.04)	0.79(±0.08)	0.72(±0.09)	0.67(±0.06)	0.91(±0.07)
Sub09	0.59(±0.07)	0.74(±0.05)	0.72(±0.11)	0.68(±0.08)	0.88(±0.06)
Sub10	0.57(±0.07)	0.83(±0.07)	0.71(±0.05)	0.73(±0.08)	0.85(±0.06)
Sub11	0.59(±0.08)	0.79(±0.08)	0.77(±0.11)	0.71(±0.09)	0.83(±0.12)
Sub12	0.44(±0.07)	0.59(±0.05)	0.60(±0.08)	0.62(±0.09)	0.78(±0.08)
Mean	<b>0.58(±0.06)</b>	<b>0.76(±0.06)</b>	<b>0.73(±0.06)</b>	<b>0.73(±0.07)</b>	<b>0.86(±0.04)</b>

### III. EXPERIMENTAL RESULTS

In this section, we present the overall performance according to the experiments (offline and online) using the proposed method and comparison methods.

#### A. Offline Session Performance

The overall performances of the offline session using the proposed model and comparison models are summarized in Tables II and III. We analyzed the classification performance obtained when using training images from different sources according to each classification group (e.g., 4-class and 2-class). The proposed method contributed to improving the classification performance for MI-BCI using the images of the grasp kinematic representation, the SAC corresponding to each grasping action as the input. The images were the result of decoding EEG signals to the corresponding grasp class; hence, even a simple CNN architecture could learn distinctive features from the images and precisely classify the test data.

1) *Four Types of Hand Grasping*: In Table II, for four grasp-tasks classifications, the proposed method demonstrated a 0.28 higher classification performance than the CSP-LDA

based on a machine learning technique. Among the conventional methods using a deep learning framework, EEGNet provided the best classification accuracy of 0.58, a 0.10 difference compared with the proposed method. The highest performance was recorded at 0.80 in Sub08, and the lowest performance was 0.51 in Sub12 in the four-class classification. In addition, to quantitatively compare the differences in classification performance between the proposed method and conventional methods, a statistical analysis was conducted employing a paired *t*-test with the Bonferroni correction [11]. Initially, the normality and homoscedasticity of each comparative group (e.g., EEGNet versus Neurograsp) were verified due to the small number of samples [3]. Both the normality (Shapiro-Wilk test) and homoscedasticity (Levene's test) were satisfied when comparing each conventional method and the proposed method. The proposed model, Neurograsp, showed significant differences in decoding performance ( $p < 0.005$ ) compared to other models (CSP+LDA, ShallowNet, DeepConvNet, and EEGNet).

2) *Two Categories of Hand Grasping:* Table III also displays the performance evaluation results for the offline session for the two-grasp categories. We divided the grasp types into power grasp and precision grasp, as depicted in Fig. 1. The proposed method demonstrated a classification performance of 0.86 across all subjects. Compared with the conventional methods, a difference of approximately 0.10~0.28 was recorded. The highest performance (0.91) was obtained by Sub08; the lowest performance (0.78) was recorded by Sub12, similar to the four-grasp-type case. Statistical differences in model performance were measured using a paired *t*-test with the Bonferroni correction for multiple comparisons. A significant difference was observed between the proposed and conventional methods ( $p < 0.005$ ).

3) *Confusion Matrix Comparison:* Fig. 4 displays the confusion matrix of the proposed model for classifying each grasping type and category (4-class and 2-class) across all subjects. Each column of the matrix represents the target class, and each row represents the predicted class. All true-positive values were greater than the true-negative and false-negative values for the tasks. Furthermore, the configuration of the multiclass comprised four different grasp types. As depicted in the confusion matrices, the true positive of the grasp category had a greater value than the grasp types. However, the proposed method confused the classification of similar grasp types that were in the same category.

### B. Online Session Performance

The real-time classification performances of the online session using the proposed method are summarized in Tables IV and V. We conducted an online session across six subjects who recorded offline performance greater than 0.68 [62], [63].

1) *Four Types of Hand Grasping:* Table IV displays continuous decisions during the online session for the four-grasping tasks. Sub01 demonstrated the highest performance on average in Run I and Run II, 0.75. The majority of the subjects indicated a greater probability of success in Run II than in Run I (e.g., Sub01, Sub06, Sub08, and Sub10). Moreover, the model

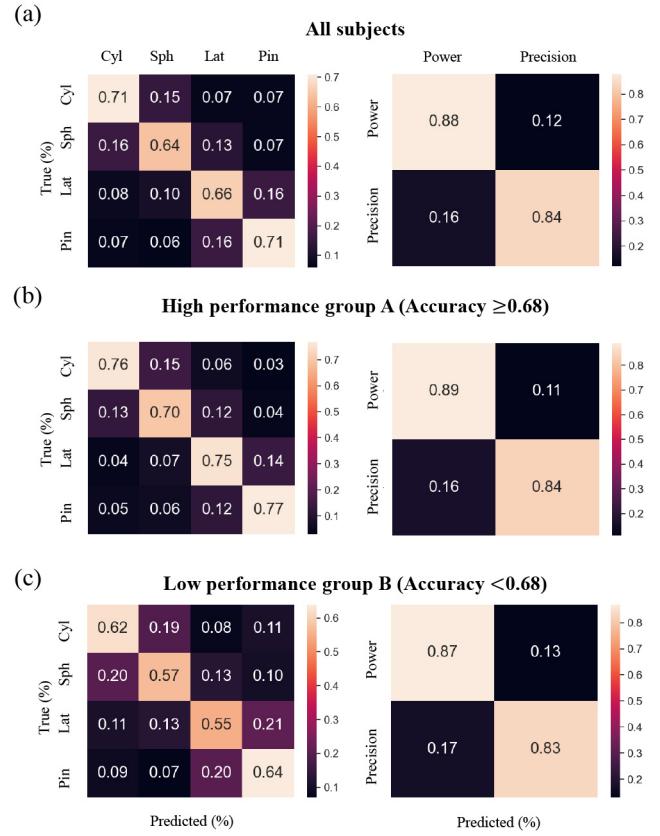


Fig. 4. Confusion matrices of each class for the representative subject groups. (a) Horizontal plane for high-performance subject group. (b) Vertical plane for low-performance subject group (\*Cyl: cylindrical grasp, \*Sph: spherical grasp, \*Lat: lateral grasp, \*Pin: pincer grasp.).

demonstrated a tendency to misclassify the grasping types of similar hand gestures, such as Cyl versus Sph and Lat versus Pin. The proposed model of Sub10 was confused regarding the hand-grasp types between Sph and Pin.

2) *Two Categories of Hand Grasping:* As mentioned in Table IV, the proposed model was confused regarding similar hand-grasp types during the experiment. Table V displays the continuous decisions during MI between the power grasp and precision grasp. The overall grand-average performance indicated a high probability of 0.78 ( $\pm 0.09$ ). Sub02 demonstrated the lowest classification performance in both sessions; Sub01 presented the highest performance probabilities with an average of 0.91 (Run I: 0.83 and Run II: 1.00). However, a more stable model with low intersubject variations is required.

3) *Constructed Images Comparison and Classification:* Fig. 5 describes the EEG data from each sliding window in an online session and describes the list of images generated. In the online session, we provided the subjects with an MI duration time of 6 s. The subject was asked to perform a single imagination when he wished. Because the pretrained model was trained using data in a 4-s size segment, the sliding window was 4 s in size and moved with a step size of 0.4 s. The SAC images represent muscle synergies and their activation coefficients. In a SAC image, the vertical axis represents a total of six muscle synergies, and the horizontal axis represents the time samples. The color of each bin in the SAC

**TABLE IV**  
ONLINE SESSION RESULTS ACROSS REPRESENTATIVE SUBJECTS. TRIALS WERE RANDOMLY GIVEN AMONG THE FOUR CLASSES. THE NUMBER FOR EACH CLASS WAS DISTRIBUTED EQUALLY

Sub01	Trials		1	2	3	4	5	6	7	8	9	10	11	12	Accuracy	
	Run I	True labels	Sph	Pin	Cyl	Lat	Cyl	Sph	Lat	Cyl	Pin	Pin	Sph	Lat	0.67	
		Predicted	Sph	Pin	Cyl	Pin	Cyl	Sph	Pin	Cyl	Cyl	Cyl	Sph	Lat		
	Run II	True labels	Sph	Pin	Lat	Pin	Cyl	Cyl	Cyl	Sph	Pin	Sph	Lat	Lat	0.83	
		Average														<b>0.75</b>
Sub02	Trials		1	2	3	4	5	6	7	8	9	10	11	12	Accuracy	
	Run I	True labels	Cyl	Lat	Pin	Sph	Sph	Pin	Cyl	Cyl	Lat	Lat	Pin	Sph	0.50	
		Predicted	Cyl	Cyl	Lat	Sph	Pin	Sph	Lat	Lat	Lat	Pin	Cyl			
	Run II	True labels	Sph	Cyl	Lat	Pin	Lat	Cyl	Lat	Sph	Cyl	Pin	Pin	Sph	0.50	
		Average														<b>0.50</b>
Sub03	Trials		1	2	3	4	5	6	7	8	9	10	11	12	Accuracy	
	Run I	True labels	Pin	Sph	Cyl	Lat	Pin	Sph	Lat	Pin	Cyl	Lat	Cyl	Sph	0.67	
		Predicted	Pin	Sph	Cyl	Cyl	Pin	Sph	Pin	Cyl	Lat	Pin	Cyl			
	Run II	True labels	Pin	Lat	Pin	Sph	Pin	Cyl	Sph	Cyl	Sph	Lat	Lat	Cyl	0.67	
		Average														<b>0.67</b>
Sub06	Trials		1	2	3	4	5	6	7	8	9	10	11	12	Accuracy	
	Run I	True labels	Pin	Sph	Cyl	Lat	Sph	Cyl	Pin	Cyl	Lat	Lat	Sph	Pin	0.67	
		Predicted	Pin	Sph	Sph	Pin	Lat	Cyl	Pin	Cyl	Cyl	Lat	Sph	Pin		
	Run II	True labels	Lat	Pin	Pin	Lat	Sph	Pin	Sph	Lat	Cyl	Cyl	Sph		0.75	
		Average														<b>0.71</b>
Sub08	Trials		1	2	3	4	5	6	7	8	9	10	11	12	Accuracy	
	Run I	True labels	Cyl	Lat	Sph	Cyl	Lat	Sph	Cyl	Pin	Lat	Sph	Pin	Pin	0.50	
		Predicted	Cyl	Lat	Cyl	Lat	Lat	Cyl	Cyl	Cyl	Cyl	Pin	Pin			
	Run II	True labels	Pin	Lat	Pin	Sph	Pin	Cyl	Sph	Cyl	Sph	Lat	Cyl	Cyl	0.67	
		Average														<b>0.58</b>
Sub10	Trials		1	2	3	4	5	6	7	8	9	10	11	12	Accuracy	
	Run I	True labels	Pin	Lat	Pin	Sph	Pin	Cyl	Sph	Cyl	Sph	Lat	Lat	Cyl	0.58	
		Predicted	Pin	Lat	Sph	Pin	Lat	Cyl	Lat	Sph	Lat	Lat	Cyl	Cyl		
	Run II	True labels	Cyl	Lat	Sph	Cyl	Lat	Sph	Cyl	Pin	Lat	Sph	Pin	Pin	0.75	
		Average														<b>0.67</b>

**TABLE V**  
ONLINE SESSION RESULTS ACROSS REPRESENTATIVE SUBJECTS: REORGANIZED RESULTS OF 4-CLASS INTO BINARY

Sub01	Trials 1 2 3 4 5 6 7 8 9 10 11 12 Accuracy														Sub06	Trials 1 2 3 4 5 6 7 8 9 10 11 12 Accuracy																	
	Run I	True labels	Pr		Run I	True labels	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	0.83														
		Predicted	Pr		Run II	True labels	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	0.75														
	Run II	True labels	Pr			True labels	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	0.91														
		Average																	Average														<b>0.79</b>
Sub02	Trials 1 2 3 4 5 6 7 8 9 10 11 12 Accuracy														Sub08	Trials 1 2 3 4 5 6 7 8 9 10 11 12 Accuracy																	
	Run I	True labels	Pr		Run I	True labels	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	0.75														
		Predicted	Pr		Run II	True labels	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	0.58														
	Run II	True labels	Pr			True labels	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	0.58														
		Average																	Average														<b>0.66</b>
Sub03	Trials 1 2 3 4 5 6 7 8 9 10 11 12 Accuracy														Sub10	Trials 1 2 3 4 5 6 7 8 9 10 11 12 Accuracy																	
	Run I	True labels	Pr		Run I	True labels	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	0.83														
		Predicted	Pr		Run II	True labels	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	0.83														
	Run II	True labels	Pr			True labels	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	Pr	0.91														
		Average																	Average														<b>0.74</b>

images represents the relative activation of those synergies; therefore, a bright color indicates a strong relation, and a dark color indicates a weak relation of the corresponding synergies over time. The six SAC images generated for each class corresponded to the EEG data obtained from each sliding window. Only one image was selected based on the similarity scores from six image candidates and used in the final classification. This strategy to increase performance in online sessions

has positively affected the performance improvements. The generation of each image was completed in the first stage of the proposed framework; in the second stage, the image was classified using the *SiamNet* network. The result was four similarity scores corresponding to each class. The method selected the image with the largest gap of the four similarity scores among the six images, and the class to which the image corresponded as the highest score of the four scores, namely, the

## Sub01-RunII

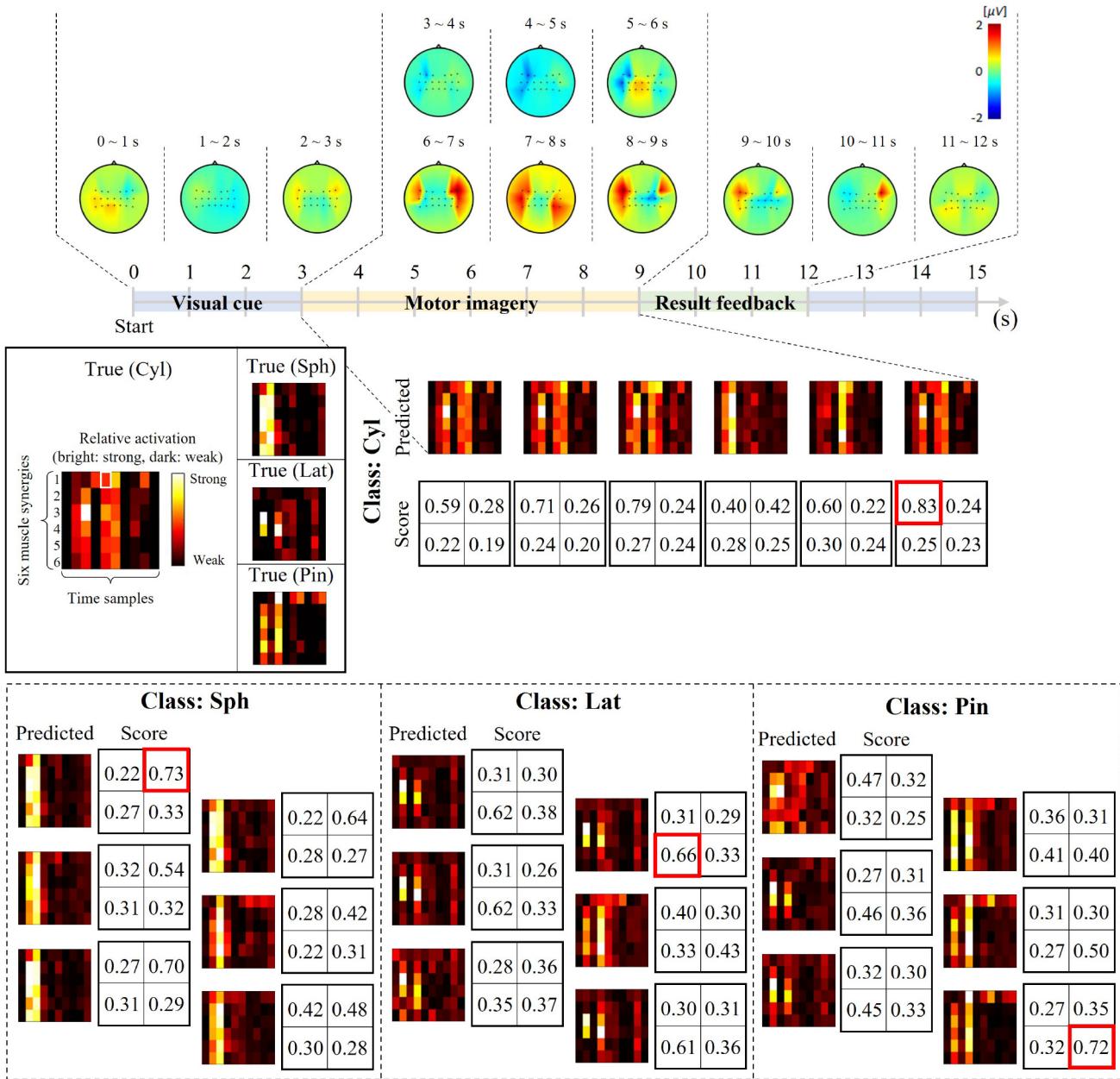


Fig. 5. Visualization of true and predicted muscle SAC images by class. The score indicates the similarity with the true image on the left-hand side of the predicted image. The proposed framework has the ability to visually identify the MI, which clearly indicates the muscle movement in the middle stage, allowing for the selective acquisition of only the clearly indicated segment information for use in the classification. For example, in the online session, we selected one of the six SAC images with the greatest standard deviation of similarity score by class and used the single image for the final classification. Also, scalp plots suggest neurophysiological evidence by representing amplitude changes of EEG signals over time.

one most similar to the ground-truth image of a particular class. This was presented as the final classification result.

## IV. DISCUSSION

In this study, we presented an online BCI system to decode natural grasp classes based on the proposed dual-stage deep learning framework, called NeuroGrasp. The proposed framework could classify natural grasp actions into four types. Simultaneously, these four grasp types are divided into two grasp categories, enabling two-class classification as well. We

define MI associated with the multiple hand grasp categories and the detailed grasp types that are performed within a single arm as the high-level MI. Consequently, it is possible to develop a four-class classification model that demonstrates stable classification performance in offline and online sessions for practical use in the future. Based on the results, we will continue to investigate the decoding of more grasping motions with brain signals. For example, hand grasping motions, such as “hook” and “tripod,” which are useful in real-world environments, will also be further experimented with, to confirm

the decoding possibility. Furthermore, the method can decode EEG data from MI to generate images representing the grasp kinematic based on SAC, providing them a unique advantage by ensuring that they are highly relevant to motion imagination. On the other hand, conventional decoding methods are not clear to demonstrate that actually classifying the EEG data evoked by the MI. Furthermore, compared to the other four conventional methods, the proposed deep learning framework provided higher classification performance in the offline session. The strength of the proposed model is that it maintains robust performance in the online session. The performances of all other models decreased significantly in the online session; in contrast, the proposed method maintained its performance relatively well.

One of the advantages of the proposed framework requiring emphasis is the decoding of clear MI-related EEG signals by constructing the estimated SAC. The limitation of MI decoding, which other related studies have identified, is that there are no effective means to clearly prove whether MI resulted from the analysis of the EEG data evoked by the imagination of continuous muscle movements of the designated motion. The traditional approach to prevent MI from being classified by other external factors, such as visual or auditory stimulation, during the experiment involves preparing a well-designed experimental protocol or noise removal method. Even then, the other BCI paradigms, such as visual imagery, can replace MI [64], [65]. For example, the conventional MI-BCIs can classify the EEG data evoked by recalling images of a target object or situation instead of the signals evoked by imagining continuous muscle movements to perform a particular action.

We conducted neurophysiological analysis to show whether the subjects actually performed the MI during the predetermined time period. The scalp plots shown in Fig. 5 represent the activation of EEG channels that change over time during an online experiment. Using the rest state as baseline, each plot shows the average of EEG amplitude ( $\mu\text{V}$ ) collected for 1 s. For example, we can see the sequential changes of the amplitude while the subject performing the MI compared to the cue period for 1–3 s. In particular, the activation became stronger around the motor cortex area from 5 s, and appeared strongest during 6–9 s. Then, the activation was gradually stabilized close to the level at the rest state during the result feedback period. We performed a baseline correction with a period of  $[-1 \ 0]$  s as the baseline. We can observe the sparse red spots at the corner estimated as irregular artifacts in several scalp plots, which can be interpreted as the proposed method having a stable performance despite unexpected noise or irregular amplitude changes in certain scalp regions that can easily be encountered in online sessions. Among the predicted six SAC image samples, the last prediction result shows the highest decoding performance, which uses the data of a 5–9 s segment that may include the irregular artifacts. However, stable decoding performance can also be observed for the second and third cases (using the data of 3.4–7.4 s and 3.8–7.8 s segments), achieving the similarity scores of 0.71 and 0.79, respectively.

The scalp topographies presented in Fig. 5 were created to show the amplitude changes of EEG signals over time

during the online experiment. We focused on managing the experimental protocol and channel selection to minimize and correct the EMG-correlated artifacts that hinder EEG-related analysis. Peripheral channels around the motor cortex were excluded to reduce unnecessary noise by movements, especially the EMG-related artifacts. We also tried to minimize unnecessary artifacts by not using low-frequency bands below 3 Hz and high-frequency bands above 40 Hz. For this reason, we only used the bandwidth between 4 and 40 Hz, which is known to be associated with MI.

Additional neurophysiological analysis was performed and presented in Fig. 6. When we analyzed EEG signals using CSP, spatial feature differences on scalp topography are not significant for each class. Meaningful activation is observed in the MI area on the scalp. Since it is the imagery of the right hand movements, the channels located on the left-hand side usually exhibit stronger activation than the channels located on the right-hand side. This neurophysiological result shows that MI for hand grasp tasks that are practically done with a single arm is difficult to classify with spatial features only. However, event-related spectral perturbation (ERSP) maps by each class changing over frequency can present relatively significant differences compared to CSP-based scalp topographies. We used the EEGLAB toolbox running under the MATLAB environment for ERSP computation [66]. First, we set the parameters to extract 200-time points from a window size of  $-2\text{--}4.5$  s at a frequency below 50 Hz. Detailed parameters followed the default settings of the EEGLAB software v2020.0. For example, we selected the baseline between  $-2$  and  $0$  s and adopted the divisive baseline method provided by the software to obtain the ERSP maps.

We could not see similar patterns with different subjects for each hand grasp class. If the subject focused well on the imagination of moving the arm and finger muscles sequentially to perform certain hand movements as we instructed beforehand, the detailed procedure would vary from subject to subject. On the other hand, significant differences between classes were identified within-subject. Therefore, we believe that neurophysiological features for the classification of hand grasp types are highly subject dependent.

In addition, the proposed framework estimates SAC through EEG decoding. SAC is the information that represents the level of activation of the group of the corresponding six muscles with time variation. This information is reconstructed by BLSTM using extracted features from a trained CNN based on the labels of the six independent muscle activities. Hence, we can argue that the classification results of the proposed framework are derived exactly by the evoked EEG signals by the imaged muscle movement over the natural grasp tasks, which is called MI. In this study, we focused on showing the feasibility of decoding multiple grasping actions using the MI; however, we will perform in-depth research on the relations between EEG and EMG signals and what features of EMG influenced to improve the performance. At the same time, we will analyze whether the model can actually reconstruct patterns for grasping actions via decoding of EEG signals in future works.

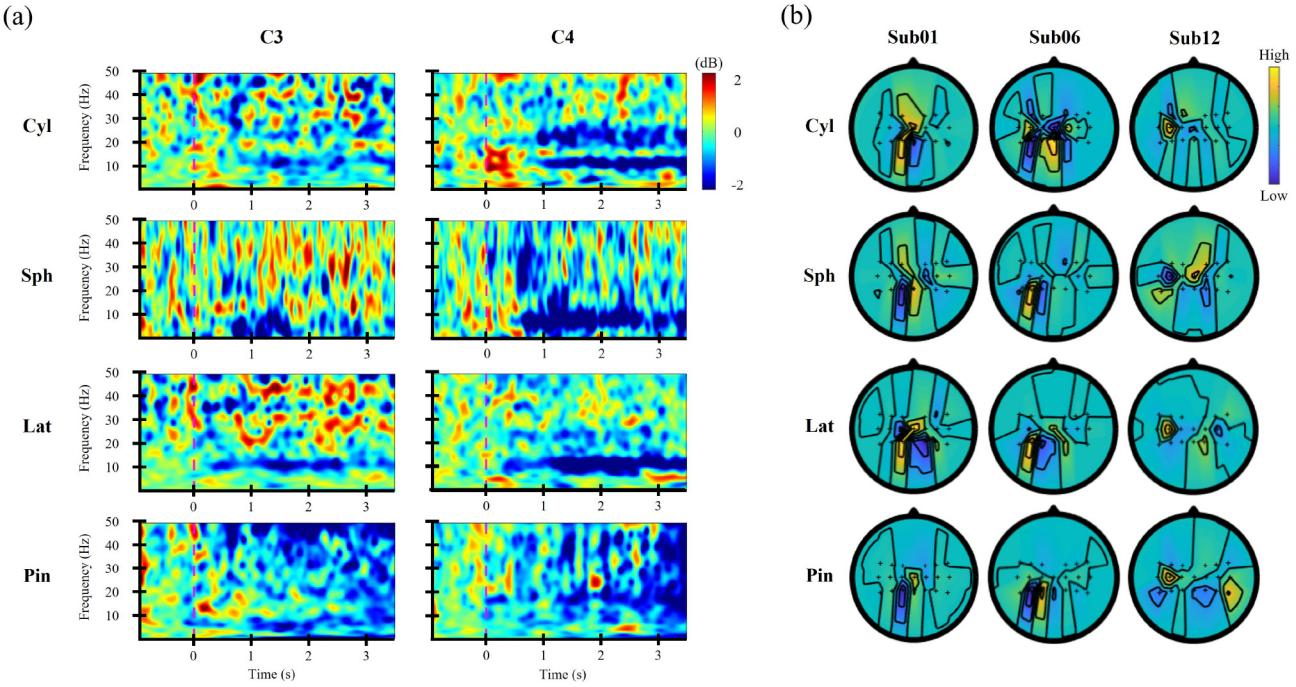


Fig. 6. Illustration of CSPs from the channels over the motor cortex region and ERSP maps were obtained based on C3 and C4 channels when subjects performed MI for each type of hand grasp. (a) Examples of ERSP maps for averaged trials of each class for Sub01. (b) CSPs for multiple subjects who showed similar and dissimilar patterns, respectively.

Recent BCI advances have adopted the deep learning framework, which can present the accurate decoding of user thought [67]. For example, the MI-BCI system demonstrated high decoding performance such as classifying upper extremity movements in an offline session [11]. However, only limited studies have demonstrated a real-time BCI system based on a deep learning framework because it requires more model training time in the calibration session compared with the conventional machine learning approaches. In general, to increase decoding accuracy, deep learning frameworks require more training samples and deeper layers; however, this requires an unacceptable amount of time for real-time BCI system users. Herein, we focused on online sessions with stable system performance despite using a deep learning framework. Hence, we designed a framework that required approximately  $20(\pm 5)$  min to train a model for each subject. Moreover, the subject is permitted a sufficient rest period of several minutes after the calibration session for data acquisition. In addition, we plan to design a shallower framework to further reduce the learning time and to conduct experiments by applying advanced machine learning algorithms for few-trial learning, reducing calibration time in future work.

Furthermore, the proposed model adopts a subject-dependent approach to consider training individual characteristics according to each subject. The development of a BCI system that considers individual differences for all subjects is an important advance direction; however, developing a complete subject-independent BCI system remains a challenge due to the lack of data [68]–[70]. The subject-independent BCI could address the rapid adaptation of a set of new subject samples and characteristics. To extend the proposed model, we plan to collect a large data sample with a different day

session and adopt advanced algorithms related to intersubject learning [71]–[73]. This is essential for practical BCI development as it can also evolve into a session-independent decoding model that does not require additional collection of data needed for calibration in online sessions.

## V. CONCLUSION

Decoding natural grasping from EEG signals is challenging. Conventional approaches have attempted to extract more features from the raw signals or to optimize the decoding algorithm. Limits exist in these methods; they are limited to the analysis of EEG signals associated with the movement of the different muscles. An approach to effectively analyze natural grasping should be able to exploit muscle synergies relating to grasping.

In this article, we proposed a method to improve the decoding performance and practicality of MI-BCIs such that users can control neuroprosthetics or robotic hands with sufficient accuracy. The proposed method, referred to as a dual-stage deep learning framework, is based on a CNN-BLSTM architecture and convolutional *SiamNet* with a proposed novel approach to creating input images. This approach improved the classification performance in both the offline and online sessions. Based on the outcome of the acceptable performance confirmed through the offline and online sessions, we are confident that the proposed method can efficiently increase the classification performance, even in BCI systems operated in an online environment. Therefore, we will focus on developing high-performance and fully asynchronous online BCI systems by attempting to overcome the limitations identified in this study in future work.

## ACKNOWLEDGMENT

The authors would like to thank B.-H. Kwon, B.-H. Lee, D.-Y. Lee, and D.-H. Lee for their assistance with the database construction and to thank Prof. C. Guan for useful discussions regarding the experiments.

## REFERENCES

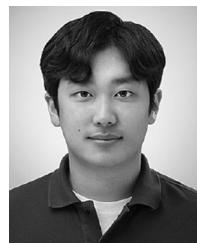
- [1] K.-T. Kim, H.-I. Suk, and S.-W. Lee, "Commanding a brain-controlled wheelchair using steady-state somatosensory evoked potentials," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 3, pp. 654–665, Mar. 2018.
- [2] K. LaFleur, K. Cassady, A. Doud, K. Shadie, E. Rogin, and B. He, "Quadcopter control in three-dimensional space using a noninvasive motor imagery-based brain-computer interface," *J. Neural Eng.*, vol. 10, no. 4, 2013, Art. no. 046003.
- [3] J.-H. Jeong, N.-S. Kwak, C. Guan, and S.-W. Lee, "Decoding movement-related cortical potentials based on subject-dependent and section-wise spectral filtering," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 3, pp. 687–698, Mar. 2020.
- [4] D.-O. Won, H.-J. Hwang, D.-M. Kim, K.-R. Müller, and S.-W. Lee, "Motion-based rapid serial visual presentation for gaze-independent brain-computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 2, pp. 334–343, Feb. 2018.
- [5] N.-S. Kwak, K.-R. Müller, and S.-W. Lee, "A lower limb exoskeleton control system based on steady state visual evoked potentials," *J. Neural Eng.*, vol. 12, no. 5, 2015, Art. no. 056009.
- [6] A. Schwarz, P. Ofner, J. Pereira, A. I. Sburlea, and G. R. Müller-Putz, "Decoding natural reach-and-grasp actions from human EEG," *J. Neural Eng.*, vol. 15, no. 1, 2017, Art. no. 016005.
- [7] P. Ofner, A. Schwarz, J. Pereira, and G. R. Müller-Putz, "Upper limb movements can be decoded from the time-domain of low-frequency EEG," *PLoS One*, vol. 12, no. 8, 2017, Art. no. e0182578.
- [8] H. A. Agashe, A. Y. Paek, Y. Zhang, and J. L. Contreras-Vidal, "Global cortical activity predicts shape of hand during grasping," *Front. Neurosci.*, vol. 9, p. 121, Apr. 2015.
- [9] J.-H. Jeong, K.-H. Shim, D.-J. Kim, and S.-W. Lee, "Brain-controlled robotic arm system based on multi-directional CNN-BiLSTM network using EEG signals," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 5, pp. 1226–1238, May 2020.
- [10] M. Zhao, M. Marino, J. Samogin, S. P. Swinnen, and D. Mantini, "Hand, foot and lip representations in primary sensorimotor cortex: A high-density electroencephalography study," *Sci. Rep.*, vol. 9, no. 1, pp. 1–12, 2019.
- [11] J.-H. Jeong, B.-H. Lee, D.-H. Lee, Y.-D. Yun, and S.-W. Lee, "EEG classification of forearm movement imagery using a hierarchical flow convolutional neural network," *IEEE Access*, vol. 8, pp. 66941–66950, 2020.
- [12] D. Achancaray and M. Hayashibe, "Decoding hand motor imagery tasks within the same limb from EEG signals using deep learning," *IEEE Trans. Med. Robot. Bionics*, vol. 2, no. 4, pp. 692–699, Nov. 2020.
- [13] R. Alazrai, H. Alwanni, Y. Baslan, N. Alnuman, and M. I. Daoud, "EEG-based brain-computer interface for decoding motor imagery tasks within the same hand using Choi-Williams time-frequency distribution," *Sensors*, vol. 17, no. 9, p. 1937, 2017.
- [14] B. J. Edelman *et al.*, "Noninvasive neuroimaging enhances continuous neural tracking for robotic device control," *Sci. Robot.*, vol. 4, no. 31, p. eaaw6844, 2019.
- [15] C. I. Penalosa and S. Nishio, "BMI control of a third arm for multitasking," *Sci. Robot.*, vol. 3, no. 20, p. eaat1228, 2018.
- [16] N.-S. Kwak and S.-W. Lee, "Error correction regression framework for enhancing the decoding accuracies of Ear-EEG brain-computer interfaces," *IEEE Trans. Cybern.*, vol. 50, no. 8, pp. 3654–3667, Aug. 2020.
- [17] X. Li, O. W. Samuel, X. Zhang, H. Wang, P. Fang, and G. Li, "A motion-classification strategy based on sEMG-EEG signal combination for upper-limb amputees," *J. Neuroeng. Rehabil.*, vol. 14, no. 1, p. 2, 2017.
- [18] L. Yao, J. Meng, D. Zhang, X. Sheng, and X. Zhu, "Combining motor imagery with selective sensation toward a hybrid-modality BCI," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 8, pp. 2304–2312, Aug. 2014.
- [19] Y. Chen *et al.*, "A high-security EEG-based login system with RSVP stimuli and dry electrodes," *IEEE Trans. Inf. Forensics Security*, vol. 11, pp. 2635–2647, 2016.
- [20] T. Yu *et al.*, "Enhanced motor imagery training using a hybrid BCI with feedback," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 7, pp. 1706–1717, Jul. 2015.
- [21] M.-H. Lee, J. Williamson, D.-O. Won, S. Fazli, and S.-W. Lee, "A high performance spelling system based on EEG-EOG signals with visual feedback," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 7, pp. 1443–1459, Jul. 2018.
- [22] O. Özdenizci, S. Y. Günay, F. Quivira, and D. Erdoğmuş, "Hierarchical graphical models for context-aware hybrid brain-machine interfaces," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Honolulu, HI, USA, 2018, pp. 1964–1967.
- [23] E. López-Larraz, N. Birbaumer, and A. Ramos-Murguialday, "A hybrid EEG-EMG BMI improves the detection of movement intention in cortical stroke patients with complete hand paralysis," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, 2018, pp. 2000–2003.
- [24] B. Lei *et al.*, "Walking imagery evaluation in brain computer interfaces via a multi-view multi-level deep polynomial network," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 3, pp. 497–506, Mar. 2019.
- [25] S. Sun, C. Zhang, and D. Zhang, "An experimental evaluation of ensemble methods for EEG signal classification," *Pattern Recognit. Lett.*, vol. 28, no. 15, pp. 2157–2163, 2007.
- [26] B. Wang *et al.*, "Common spatial pattern reformulated for regularizations in brain-computer interfaces," *IEEE Trans. Cybern.*, vol. 51, no. 10, pp. 5008–5020, Oct. 2021.
- [27] F. Qi *et al.*, "Spatiotemporal-filtering-based channel selection for single-trial EEG classification," *IEEE Trans. Cybern.*, vol. 51, no. 2, pp. 558–567, Feb. 2021.
- [28] P. Zhang, X. Wang, W. Zhang, and J. Chen, "Learning spatial-spectral-temporal EEG features with recurrent 3D convolutional neural networks for cross-task mental workload assessment," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 1, pp. 31–42, Jan. 2019.
- [29] Y. Zhang, C. S. Nam, G. Zhou, J. Jin, X. Wang, and A. Cichocki, "Temporally constrained sparse group spatial patterns for motor imagery BCI," *IEEE Trans. Cybern.*, vol. 49, no. 9, pp. 3322–3332, Sep. 2019.
- [30] Z. Tayeb *et al.*, "Validating deep neural networks for online decoding of motor imagery movements from EEG signals," *Sensors*, vol. 19, no. 1, p. 210, 2019.
- [31] M. Veres, M. Moussa, and G. W. Taylor, "Modeling grasp motor imagery through deep conditional generative models," *IEEE Robot Autom. Lett.*, vol. 2, no. 2, pp. 757–764, Apr. 2017.
- [32] T. Uktveris and V. Jusas, "Application of convolutional neural networks to four-class motor imagery classification problem," *Inf. Technol. Control*, vol. 46, no. 2, pp. 260–273, 2017.
- [33] D. Zhang, L. Yao, K. Chen, S. Wang, X. Chang, and Y. Liu, "Making sense of spatio-temporal preserving representations for EEG-based human intention recognition," *IEEE Trans. Cybern.*, vol. 50, no. 7, pp. 3033–3044, Jul. 2020.
- [34] L. Yun, Z. Lifeng, and Z. Shujun, "A hand gesture recognition method based on multi-feature fusion and template matching," *Procedia Eng.*, vol. 29, pp. 1678–1684, Jan. 2012.
- [35] 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, 2017, Art. no. 016003.
- [36] Z. Zhang *et al.*, "A novel deep learning approach with data augmentation to classify motor imagery signals," *IEEE Access*, vol. 7, pp. 15945–15954, 2019.
- [37] W. Huang, Y. Xue, L. Hu, and H. Liuli, "S-EEGNet: Electroencephalogram signal classification based on a separable convolution neural network with bilinear interpolation," *IEEE Access*, vol. 8, pp. 131636–131646, 2020.
- [38] X. Deng, B. Zhang, N. Yu, K. Liu, and K. Sun, "Advanced TSGL-EEGNet for motor imagery EEG-based brain-computer interfaces," *IEEE Access*, vol. 9, pp. 25118–25130, 2021.
- [39] 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, 2020, Art. no. 016025.
- [40] P.-Y. Jeng, C.-S. Wei, T.-P. Jung, and L.-C. Wang, "Low-dimensional subject representation-based transfer learning in EEG decoding," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 6, pp. 1915–1925, Jun. 2021.
- [41] O. Özdenizci, Y. Wang, T. Koike-Akino, and D. Erdoğmuş, "Learning invariant representations from EEG via adversarial inference," *IEEE Access*, vol. 8, pp. 27074–27085, 2020.
- [42] K. Zhang, N. Robinson, S.-W. Lee, and C. Guan, "Adaptive transfer learning for EEG motor imagery classification with deep convolutional neural network," *Neural Netw.*, vol. 136, pp. 1–10, Apr. 2021.

- [43] B. Xu *et al.*, "Phase synchronization information for classifying motor imagery EEG from the same limb," *IEEE Access*, vol. 7, pp. 153842–153852, 2019.
- [44] K. K. Ang and C. Guan, "EEG-based strategies to detect motor imagery for control and rehabilitation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 4, pp. 392–401, Apr. 2017.
- [45] S.-H. Park, D. Lee, and S.-G. Lee, "Filter bank regularized common spatial pattern ensemble for small sample motor imagery classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 2, pp. 498–505, Feb. 2018.
- [46] S. Sakhavi, C. Guan, and S. Yan, "Learning temporal information for brain-computer interface using convolutional neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5619–5629, Nov. 2018.
- [47] M. Miyakoshi, A. Delorme, T. Mullen, K. Kojima, S. Makeig, and E. Asano, "Automated detection of cross-frequency coupling in the electrocorticogram for clinical inspection," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, 2013, pp. 3282–3285.
- [48] M. D. K. Breteler, K. J. Simura, and M. Flanders, "Timing of muscle activation in a hand movement sequence," *Cereb. Cortex*, vol. 17, no. 4, pp. 803–815, 2007.
- [49] E. J. Weiss and M. Flanders, "Muscular and postural synergies of the human hand," *J. Neurophysiol.*, vol. 92, no. 1, pp. 523–535, 2004.
- [50] L. Bertinetto, J. Valmadre, J. F. Henriques, A. Vedaldi, and P. H. S. Torr, "Fully-convolutional siamese networks for object tracking," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 850–865.
- [51] G. Koch, R. Zemel, and R. Salakhutdinov, "Siamese neural networks for one-shot image recognition," in *Proc. ICML Deep Learn. Workshop*, vol. 2, 2015, pp. 1–30.
- [52] H. Yokoyama, N. Kaneko, T. Ogawa, N. Kawashima, K. Watanabe, and K. Nakazawa, "Cortical correlates of locomotor muscle synergy activation in humans: An electroencephalographic decoding study," *iScience*, vol. 15, pp. 623–639, May 2019.
- [53] A. D'Avella, P. Saltiel, and E. Bizzi, "Combinations of muscle synergies in the construction of a natural motor behavior," *Nat. Neurosci.*, vol. 6, no. 3, pp. 300–308, 2003.
- [54] M. Simão, N. Mendes, O. Gibaru, and P. Neto, "A review on electromyography decoding and pattern recognition for human-machine interaction," *IEEE Access*, vol. 7, pp. 39564–39582, 2019.
- [55] M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca, and R. J. Plemmons, "Algorithms and applications for approximate non-negative matrix factorization," *Comput. Stat. Data Anal.*, vol. 52, no. 1, pp. 155–173, 2007.
- [56] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, vol. 13, 2000, pp. 556–562.
- [57] I. Delis, B. Berret, T. Pozzo, and S. Panzeri, "A methodology for assessing the effect of correlations among muscle synergy activations on task-discriminating information," *Front. Comput. Neurosci.*, vol. 7, p. 54, May 2013.
- [58] S. Israely, G. Leisman, C. C. Machluf, and E. Carmeli, "Muscle synergies control during hand-reaching tasks in multiple directions post-stroke," *Front. Comput. Neurosci.*, vol. 12, p. 10, Feb. 2018.
- [59] S.-L. Wu, C.-W. Wu, N. R. Pal, C.-Y. Chen, S.-A. Chen, and C.-T. Lin, "Common spatial pattern and linear discriminant analysis for motor imagery classification," in *Proc. Symp. Comput. Intell. Cogn. Algorithms Mind Brain*, Singapore, 2013, pp. 146–151.
- [60] R. T. Schirrmeister *et al.*, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Human Brain Map.*, vol. 38, no. 11, pp. 5391–5420, 2017.
- [61] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, 2018, Art. no. 056013.
- [62] B. Blankertz *et al.*, "Neurophysiological predictor of SMR-based BCI performance," *NeuroImage*, vol. 51, no. 4, pp. 1303–1309, 2010.
- [63] L. Acqualagna, L. Botrel, C. Vidaurre, A. Kübler, and B. Blankertz, "Large-scale assessment of a fully automatic co-adaptive motor imagery-based brain computer interface," *PLoS One*, vol. 11, no. 2, 2016, Art. no. e0148886.
- [64] P. Chholak *et al.*, "Visual and kinesthetic modes affect motor imagery classification in untrained subjects," *Sci. Rep.*, vol. 9, no. 1, pp. 1–12, 2019.
- [65] R. L. Klatzky, "On the relation between motor imagery and visual imagery," *Behav. Brain Sci.*, vol. 17, no. 2, pp. 212–213, 1994.
- [66] 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, 2004.
- [67] 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, 2019, Art. no. 051001.
- [68] J.-H. Jeong *et al.*, "Multimodal signal dataset for 11 intuitive movement tasks from single upper extremity during multiple recording sessions," *GigaScience*, vol. 9, no. 10, p. gaa098, 2020.
- [69] H.-I. Suk and S.-W. Lee, "Subject and class specific frequency bands selection for multiclass motor imagery classification," *Int. J. Imag. Syst. Technol.*, vol. 21, no. 2, pp. 123–130, 2011.
- [70] M.-H. Lee, S. Fazli, J. Mehner, and S.-W. Lee, "Subject-dependent classification for robust idle state detection using multi-modal neuroimaging and data-fusion techniques in BCI," *Pattern Recognit.*, vol. 48, no. 8, pp. 2725–2737, 2015.
- [71] 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, 2019, Art. no. 026007.
- [72] O.-Y. Kwon, M.-H. Lee, C. Guan, and S.-W. Lee, "Subject-independent brain-computer interfaces based on deep convolutional neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 10, pp. 3839–3852, Oct. 2020.
- [73] M.-H. Lee *et al.*, "EEG dataset and OpenBMI toolbox for three BCI paradigms: An investigation into BCI illiteracy," *Gigascience*, vol. 8, no. 5, p. giz002, 2019.



**Jeong-Hyun Cho** received the M.S. degree in information and communication engineering from Ajou University, Suwon, South Korea, in 2017, with a graduate education agreement for military officers in South Korea. He is currently pursuing the Ph.D. degree with the Department of Brain and Cognitive Engineering, Korea University, Seoul, South Korea.

His research interests include haptic intelligence, machine learning, deep learning, and brain-computer interfaces.



**Ji-Hoon Jeong** (Associate Member, IEEE) received the B.S. degree in computer information science and the Ph.D. degree in brain and cognitive engineering from Korea University, Seoul, South Korea, in 2015 and 2021, respectively.

He is currently a Research Professor with the Department of Artificial Intelligence, Korea University. His research interests include machine learning, deep learning, and brain-computer interfaces.



**Seong-Whan Lee** (Fellow, IEEE) received the B.S. degree in computer science and statistics from Seoul National University, Seoul, South Korea, in 1984, and the M.S. and Ph.D. degrees in computer science from the Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 1986 and 1989, respectively.

He is currently the Head of the Department of Artificial Intelligence, Korea University, Seoul. His current research interests include artificial intelligence, pattern recognition, and brain engineering.

Dr. Lee is a Fellow of the International Association of Pattern Recognition and the Korea Academy of Science and Technology.