

Brain-Controlled Robotic Arm System Based on Multi-Directional CNN-BiLSTM Network Using EEG Signals

Ji-Hoon Jeong^{ID}, Kyung-Hwan Shim^{ID}, Dong-Joo Kim^{ID}, and Seong-Whan Lee^{ID}, *Fellow, IEEE*

Abstract— Brain-machine interfaces (BMIs) can be used to decode brain activity into commands to control external devices. This paper presents the decoding of intuitive upper extremity imagery for multi-directional arm reaching tasks in three-dimensional (3D) environments. We designed and implemented an experimental environment in which electroencephalogram (EEG) signals can be acquired for movement execution and imagery. Fifteen subjects participated in our experiments. We proposed a multi-directional convolution neural network-bidirectional long short-term memory network (MDCBN)-based deep learning framework. The decoding performances for six directions in 3D space were measured by the correlation coefficient (CC) and the normalized root mean square error (NRMSE) between predicted and baseline velocity profiles. The grand-averaged CCs of multi-direction were 0.47 and 0.45 for the execution and imagery sessions, respectively, across all subjects. The NRMSE values were below 0.2 for both sessions. Furthermore, in this study, the proposed MDCBN was evaluated by two online experiments for real-time robotic arm control, and the grand-averaged success rates were approximately 0.60 (± 0.14) and 0.43 (± 0.09), respectively. Hence, we demonstrate the feasibility of intuitive robotic arm control based on EEG signals for real-world environments.

Index Terms— Brain-machine interface (BMI), electroencephalogram (EEG), motor imagery, intuitive robotic arm control, deep learning.

Manuscript received July 11, 2019; revised November 19, 2019 and February 19, 2020; accepted March 12, 2020. Date of publication March 18, 2020; date of current version May 8, 2020. This work was supported in part by the Institute of Information & Communications Technology Planning & Evaluation (IITP) funded by the Korea Government through the Development of Non-Invasive Integrated BCI SW Platform to Control Home Appliances and External Devices by User's Thought via AR/VR Interface under Grant 2017-0-00432, in part by the IITP funded by the Korea Government through the Development of BCI based Brain and Cognitive Computing Technology for Recognizing User's Intentions using Deep Learning under Grant 2017-0-00451, and in part by the IITP funded by the Korea Government through the Department of Artificial Intelligence, Korea University, under Grant 2019-0-00079. (Corresponding author: Seong-Whan Lee.)

Ji-Hoon Jeong, Kyung-Hwan Shim, and Dong-Joo Kim are with the Department of Brain and Cognitive Engineering, Korea University, Seoul 02841, South Korea (e-mail: jh_jeong@korea.ac.kr; kh_shim@korea.ac.kr; dongjookim@korea.ac.kr).

Seong-Whan Lee is with the Department of Artificial Intelligence, Korea University, Seoul 02841, South Korea (e-mail: sw.lee@korea.ac.kr).

This article has supplementary downloadable material available at <http://ieeexplore.ieee.org>, provided by the authors.

Digital Object Identifier 10.1109/TNSRE.2020.2981659

I. INTRODUCTION

BRAIN-MACHINE interface (BMI) technology is a promising tool for the recognition of user intention and for communicating with external devices. Noninvasive BMIs aim to translate human thoughts, decoded by brain activity into control signals, for external applications without brain implant surgery. These interfaces typically decode the cortical correlates of movement parameters such as velocity/position of limb movements or muscular activity to generate control sequences [1], [2]. Over the past decades, BMIs have been developed to control devices such as wheelchairs [3], powered exoskeletons [4], [5], and robotic arms [6]–[9] not only facilitating the recovery of movement functions for spinal cord injury (SCI) and amyotrophic lateral sclerosis (ALS) patients, but also supporting the abilities of healthy people [10]–[13].

One type of BMI application, the robotic arm could provide alternative upper extremity function to both healthy people and motor-disabled patients, allowing them to perform high-level tasks, such as drinking water or moving objects. Some BMI groups have investigated scalp electroencephalogram (EEG)-based robotic arm control systems using various BMI paradigms, such as steady-state visual evoked potentials [14], P300 potentials [15], error-related potentials [8], motor imagery [6], [7], and shared control [16], [17].

To emulate the natural control of a brain-to-robot arm, motor imagery is an important topic that reflects the thought of users about moving specific body parts without any external stimuli. When a user imagines the motion of a limb, event-related desynchronization/synchronization (ERD/ERS) features, which are referred to as a sensory-motor rhythm (SMR), are induced in EEG signals over the primary sensorimotor area. These phenomena can be induced from the mu-band (8–12 Hz) and the beta-band (13–30 Hz). However, due to the low signal-to-noise ratio (SNR) of EEG signals, understanding the brain dynamics and accurately decoding various types of user intention for controlling human assistant robots is very challenging [18]–[20]. For reliable motor imagery decoding, a few groups have designed effective experiments that can comparatively easily induce endogenous potentials through movement execution sessions [21], [22]. Therefore, to control robotic and prosthetic arms, previous researchers have attempted to decode various upper extremity

TABLE I
THE CONVENTIONAL STUDIES FOR DECODING UPPER EXTREMITY MOVEMENTS FROM EEG SIGNALS

Authors	Experimental paradigm	Number of subjects	Tasks	EEG features	Decoding methods	Decoding performances
Kim et al. [9]	Arm movement following an infinity symbol	Ten	ME ^a , MI ^b	BP ^c	MLR ^d , KRR ^e	<i>r</i> -value: 0.46, NRMSE \leq 0.10 (ME) <i>r</i> -value: 0.41, NRMSE \leq 0.15 (MI)
Ofner et al. [21]	Forearm pronation/supination, elbow extension/flexion, hand open/close	Fifteen	ME, MI	BP	sLDA ^f	Accuracy (6-class): 55.0% (ME), 27.0% (MI)
Schwarz et al. [23]	Palmar grasp, pincer grasp, lateral grasp	Fifteen	ME	BP, PCA ^g	sLDA	Accuracy (2-class): 72.5%
Yong et al. [24]	Grasping, elbow extension/flexion	Twelve	ME, MI	CSP ^h , FBCSP ⁱ , BP	LDA, LR ^j , SVM ^k	Accuracy (3-class): 66.9% (ME), 60.7% (MI)
Iturrate et al. [25]	Power grasp, precision grasp	Ten	ME	TF ^l , PCA	sLDA	Accuracy (2-class): 70.0%
Edelman et al. [26]	Forearm pronation/supination, elbow extension/flexion	Five	MI	TF	MD ^m	True positive rate (elbow flexion): 82.2%
Li et al. [27]	Hand open/close, wrist pronation/supination	Four	ME	TD ⁿ , WL ^o , ZC ^p , SSC ^q	LDA	Accuracy: 87.0%
Chung et al. [28]	Elbow extension/flexion following three angles	Sixteen	ME	TF	MPA ^r	-
Antelis et al. [29]	Arm reaching for one of the three targets	Six	ME	PSD ^s	SVM, HMM ^t	Accuracy: 78.0%
Shiman et al. [30]	Center-out arm reaching for Five targets	Twelve	ME	FBCSP	sLDA	Accuracy (6-class): 50.3%
Handiru et al. [31]	Center-out arm reaching for four directions	Five	ME	SFA ^u	FLD ^v	Accuracy (4-class): 71.0%
Korik et al. [32]	Center-out arm reaching for four targets	Twelve	ME, MI	BP	BTS ^x	<i>r</i> -value: 0.40
Ubeda et al. [33]	Center-out arm reaching for eight targets	Five	ME	BP	MLR	Accuracy (8-class): 29.0%

^aMovement execution, ^bMotor imagery, ^cBand power, ^dMultiple linear regression, ^eKernel ridge regression, ^fShrinkage linear discriminant analysis,

^gPrincipal component analysis, ^hCommon spatial pattern, ⁱFilter-bank common spatial pattern, ^jLogistic regression, ^kSupport vector machine,

^lTime-frequency, ^mMahalanobis distance, ⁿTime-domain, ^oWaveform length, ^pZero crossings, ^qSlope sign changes, ^rMeasure projection analysis,

^sPower spectral density, ^tHidden markov model, ^uSupervised factor analytic, ^vFisher's linear discriminant, ^xBand power-based time-series

movements using motor imagery and movement execution. Table I summarizes the conventional studies for EEG-based upper extremity decoding.

Ofner *et al.* [21] decoded the different types of upper extremity motions for motor imagery and movement execution. They categorized six movements and a rest class and obtained significant average classification accuracies of 27% (movement vs. movement) and 73% (movement vs. rest) for motor imagery, and 55% and 87% for movement execution. Schwarz *et al.* [23] investigated the classification of hand motions according to various objects (e.g., palmar grasp, pincer grasp, and lateral grasp). They showed binary classification accuracies of 72.4% between grasp types and the peak classification accuracy could be achieved 93.5% for grasps vs. no-movement condition. Yong and Menon [24] proposed a multi-class BCI system that could be classified for a resting state, a hand grasping imagery state, and an elbow movement imagery state; the states were classified using discriminating EEG patterns. The averaged classification accuracies were 66.9% in movement execution task and 60.7% in movement imagery tasks for the 3-class problem across 12 subjects. Iturrate *et al.* [25] had investigated the existence of neural correlates of grasping from EEG signals in self-paced conditions, and demonstrated neural correlates of precision and power grasps differentially evolve as grasps unfold. They decoded the power and precision hand grasping types with 70.0% performance with a standardized low-resolution brain electromagnetic tomography (sLORETA). Edelman *et al.* [26] decoded four different wrist movements using power modulation in the mu-band based on an EEG source imaging method. They compared the classification accuracy between sensor-based and source-based methods. The source-based method showed a 12.7% increase in accuracy for the overall classification when compared with that of the sensor-based method. Li *et al.* [27] classified four different upper extremity movements such as wrist pronation/supination and hand open/close using

EEG-surface electromyography (sEMG) combination method. They finally obtained a classification accuracy of 87.0% using a channel optimization method (sequential forward selection). Chung *et al.* [28] explored the beta-bands activity and connectivity of arm reaching tasks with low and high visual gain according to the target distance (12°, 36°, and 72°). They used the EEG source localization based on measure projection analysis (MPA) and discovered that the high visual gain decreased movement errors with increased beta-band desynchronization in the parietal cortex and contralateral sensorimotor cortex.

Especially, for decoding of arm reaching movement using EEG signals, Antelis *et al.* [29] demonstrated the possibility of the continuous decoding of a natural attempt in upper limb movement for stroke patients from the EEG signals of the unaffected contralesional motor cortex. They proposed a continuous decoder based on a support vector machine using a hidden Markov model, and detected the hand movement itself with 78.0% accuracy as well as the attempt at movement with 76.0% accuracy, respectively. Shiman *et al.* [30] investigated the decoding of upper extremity reaching movements in multiple directions with a robotic exoskeleton according to a horizontal 2D plane. Handiru *et al.* [31] also classified multi-directional arm movements using EEG source space analysis in the 2D space (north, west, east, and south). Some research groups have conducted studies estimating arm movement trajectories and velocity profiles from EEG signals [9], [32], [33].

These related studies have investigated to the possibility of decoding various types of upper extremity movement from EEG signals. However, sufficient decoding performance for real-time BMI application control has not yet been demonstrated. In this study, we first focused on a robust decoding method using a deep learning technique, which has demonstrated high performance in various research fields (e.g., computer vision and signal processing). Recent deep learning

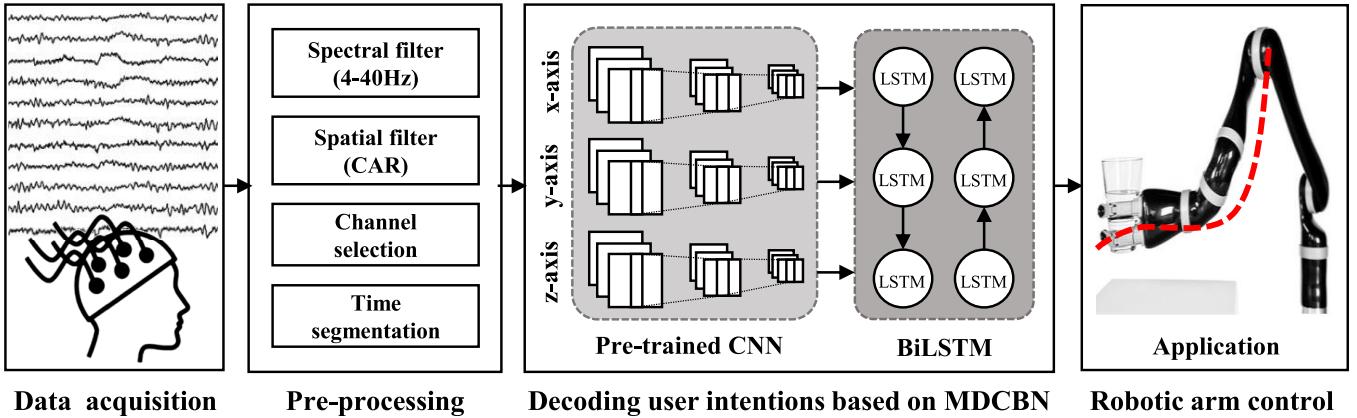


Fig. 1. The overall system architecture for BMI-based robotic arm control using the proposed MDCBN framework. The MDCBN was designed to consider 3D axes (x-, y-, and z-axis) for the multi-directional arm reaching.

schemes have also been applied in the BMI field to better understand brain imaging and to recognize various user intention [34], [35].

Furthermore, we attempted to decode intuitive imagery for 3D robotic arm control with a high degree of freedom (DoF). Recent advances in motor imagery-based robotic arms have utilized simplified imagery, such as a left hand, a right hand, feet, and both hands, as control commands. For example, Meng *et al.* [6] and Edelman *et al.* [36] demonstrated the feasibility of a robotic arm that performed reach-and-grasp tasks in a real-time environment. As control tasks, they adopted a 2D cursor control using motor imagery of the left hand, the right hand, both hands, and resting state. Penaloza and Nishio [7] applied the concept of a third arm in the process of grasp-release imagination of a bottle. This study demonstrated that the user's two arms could perform a specific task while a robot arm, called the third arm, is simultaneously manipulated by imagining a different movement. In this study, we approached 3D robotic arm control using intuitive imagery without artificial matching between the robotic arm and the user's intention. To the best of our knowledge, this is the first time that intuitive imagery of 3D multi-direction (left, right, forward, backward, up, and down) has been encoded for arm reaching. We designed a BMI experiment to obtain EEG signals for movement execution and motor imagery. To facilitate the robust decoding of intuitive imagery for multi-directional arm reaching, we proposed a deep learning framework based on the multidirectional convolution neural network (CNN)-bidirectional long short-term memory (BiLSTM) network. We evaluated the intuitive decoding performance to robotic arm controls using more challenging real-time scenarios involving high-level tasks. These tasks include the reach-and-grasp of objects in 3D space and the drinking a glass of water or beverage.

II. MATERIALS AND METHODS

We designed an experimental environment that can acquire EEG data according to various arm reaching tasks in a movement session and an imagery session. Additionally, we proposed a deep learning framework for decoding user intention and evaluated our overall system through an online

experiment. A JACO arm (Kinova Robotics Inc., Canada), a seven DoF human-like robotic arm with three fingers, was used as the BMI actuator for the subjects during the online experiments. Our overall system architecture for controlling the robotic arm from EEG signals is shown in Fig. 1.

A. Subjects

Fifteen healthy volunteers (eleven males and four females, all right-handed, age: 25-31 years) participated in our experiment. All volunteers were BMI naïve. Before the experiment, the subjects were required to be in normal health, get enough sleep (approximately 8 hours), avoid alcohol, and strenuous exercise. The subjects were informed about the experimental protocols and were provided their given informed consent with the Declaration of Helsinki and data user agreement. All experimental protocols and environments for ethical issues from human subjects were reviewed and approved by the Institutional Review Board at Korea University (1040548-KU-IRB-17-181-A-2).

B. Experimental Protocols and Setup

The tasks performed in our experiment involved center-out right-handed reaching movement and imagery (left, right, forward, backward, up, and down) in 3D space. The experimental setup is depicted in Fig. 2(a) and 2(b). The subjects sat in a chair in front of a desk. A display monitor was placed at a comfortable viewing distance of approximately 60 cm from the subject on the desk. Visual cues were shown in the form of a plus sign for rest and appropriate directional signs for reaching tasks. The experiment comprised two sessions: movement session and imagery session. In the movement session, the subjects performed the actual movement, such as center-out arm reaching in one of the directions. In the imagery session, the subjects were asked to perform motor imagery only. Each session included 240 trials for each subject (40 trials per six directions). The experimental paradigm is depicted in Fig. 2(c). Each trial began with a resting state of four seconds. During this period, the subjects were asked to minimize eye blinks and body movements. After being shown

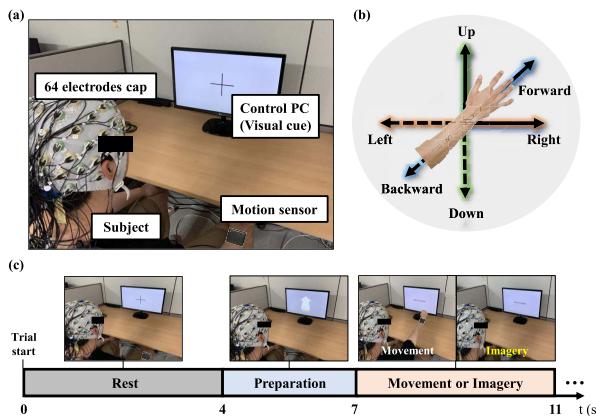


Fig. 2. Experimental setup and protocols for EEG data collection. (a) Experimental environments including the 64 EEG electrodes, a monitor for displaying visual cues, and a motion sensor for detecting kinematic information. (b) Arm reaching tasks in multi-direction (left, right, forward, backward, up, and down). (c) Experimental paradigms of the two sessions: movement session and imagery session.

a random visual instruction for three seconds, the subjects performed movement or imagery for four seconds.

C. Data Acquisition

EEG signals from 64 channels were continuously recorded according to an international 10-20 electrode position system using BrainAmp devices (BrainProduct GmbH, Germany). The ground and reference electrodes were placed on the FPz and FCz channels, respectively. The impedance of all electrodes was maintained below $15\text{ k}\Omega$. The signals were digitally sampled at 1,000 Hz and a 50 Hz notch filter was applied to reduce power supply noise. Additionally, kinematic data were recorded using an OPAL sensor (OPAL, APDM Inc., Canada). The sensor included an accelerometer, a gyroscope, and a magnetometer to provide the kinematic data of natural movement in the three axes. In our experiment, the sensor was strapped to the subject's right arm and a robotic arm was used to obtain the exact arm position and movement velocity.

D. EEG and Kinematic Data Preprocessing

The EEG data were bandpass filtered in the range from 4 to 40 Hz [37] using Hamming-windowed zero phase finite impulse response (FIR) filters with an optimized order ($N = 50$) [38]. Despite applying the digital filter to EEG, contaminated data still exists, because such data is generated by eye blinks and head movements. To obtain cleaned EEG data, we removed the contamination factors by an independent component analysis (ICA) which is commonly used to decompose the brain signals into statistically independent components (ICs). The data were transformed by the ICA mixing matrix. The contaminated ICs that had similar patterns of the Fp1 or Fp2 channels were automatically removed [39]–[41]. The remaining ICs were projected back into the scalp channel space to be reconstructed as the cleaned EEG data. Next, the EEG data were re-referenced using a common average reference (CAR) spatial filter to consider spatial distribution related to the upper extremity. The CAR filter,

which has been used to decrease volume conduction effects for scalp EEG functional connectivity, is mostly used to attenuate common artifacts arising at all EEG channels at the same time [23], [42], [43]. After spatial filtered, twenty EEG channels (FC5, FC3, FC1, FC2, FC4, FC6, C5, C3, C1, C2, C4, C6, CP5, CP3, CP1, CPz, CP2, CP4, and CP6) near the primary/supplementary motor cortices were selected. Those motor cortices located in central brain area were activated when the subjects executed movements and imagined their muscle movement [10], [44], [45]. Hence, we designed a deep learning architecture as the input for selecting only twenty related channels to avoid weights from learning poorly due to unnecessary information from all channels. The EEG data were downsampled from 1,000 to 100 Hz.

The kinematic data from the OPAL sensor were used as the movement profile. In our experiment, the robotic arm performed reaching tasks according to the directions given in order to acquire the baseline velocity profile after the sessions. Fig. 3(a) represents the arm trajectory information for 3D multi-direction (left, right, forward, backward, up, and down), obtained via the kinematic data of the robotic arm movement. The trajectory information was used to compare the reconstructed trajectory pathway with an arm trajectory information for real-time control commands. Fig. 3(b) shows the baseline velocity profiles for each axis and each direction. The variable t denotes the time elapsed within the 4 s interval and v is the value of measured velocity (cm/s) that ranges from -5 to 5 cm/s. We adopted the velocity profile of the kinematic information as the labels of the decoding model for training.

E. Multi-Directional CNN-BiLSTM Network (MDCBN)

As the decoding method, we proposed the multi-directional CNN-BiLSTM network (MDCBN)-based deep learning framework considering 3D multi-direction. The proposed framework took the form of a hybrid deep learning model using the CNN and the BiLSTM (Fig. 4 and Fig. 5). Conventional hybrid deep learning frameworks [46]–[48] have been commonly trained to extract spatial features of brain activities using a CNN and have been trained with temporal information using LSTM networks [47], [49]. In contrast, the proposed MDCBN framework was designed using a CNN architecture to train the multi-direction information per axis as pretraining and it used the BiLSTM network for training the relationships in the 3D space (x-, y-, and z-axes). Furthermore, we designed the MDCBN framework based on adopting a subject-dependent BCI approach in the training stage. The intra-subject approach is one of those model generating strategies based on machine learning [18], [50], [51] and deep learning [47], [48], [52]–[54] that lead to dramatic BCI performance improvement according to each subject since a large individual difference of EEG signals [18], [55]. Hence, the proposed MDCBN model could train according to each subject and evaluated the decoding performance in the offline analysis and real-time experiment using each individual model.

Initially, we randomly selected 80% of the trials as a training set and used the remaining 20% as a test set for

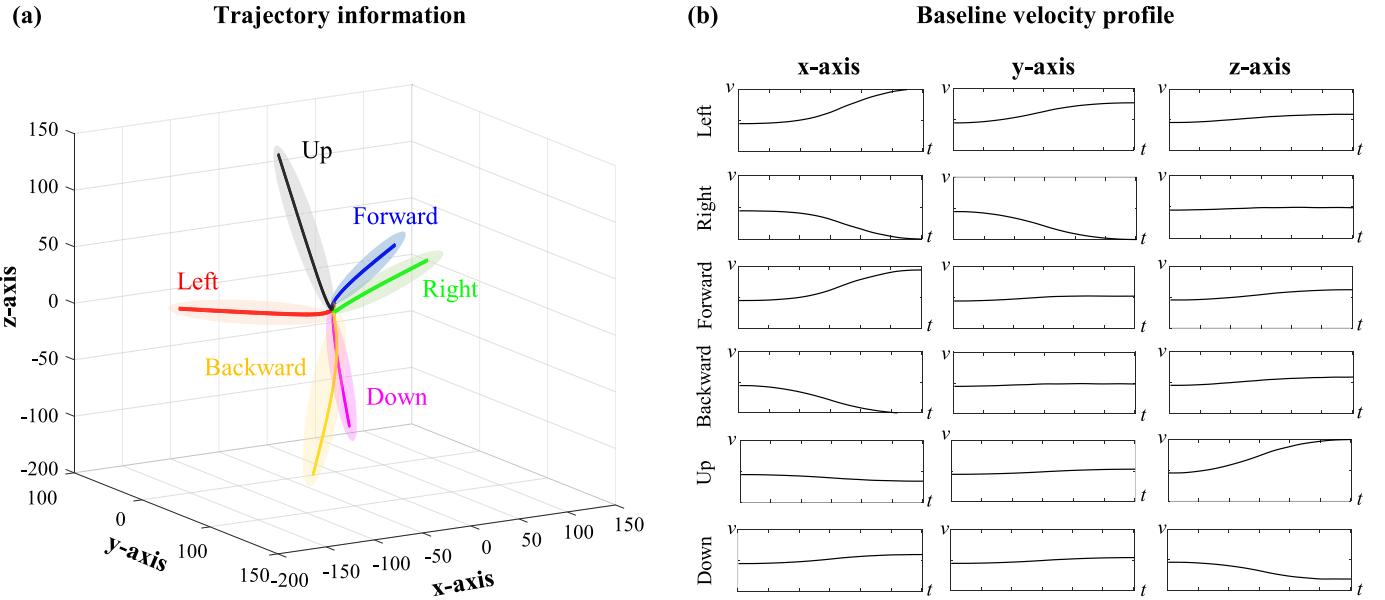


Fig. 3. Baseline kinematic information for arm trajectory and velocity profile. (a) Trajectory information using the robotic arm movement profile. (b) Velocity profiles for six directions per each axis.

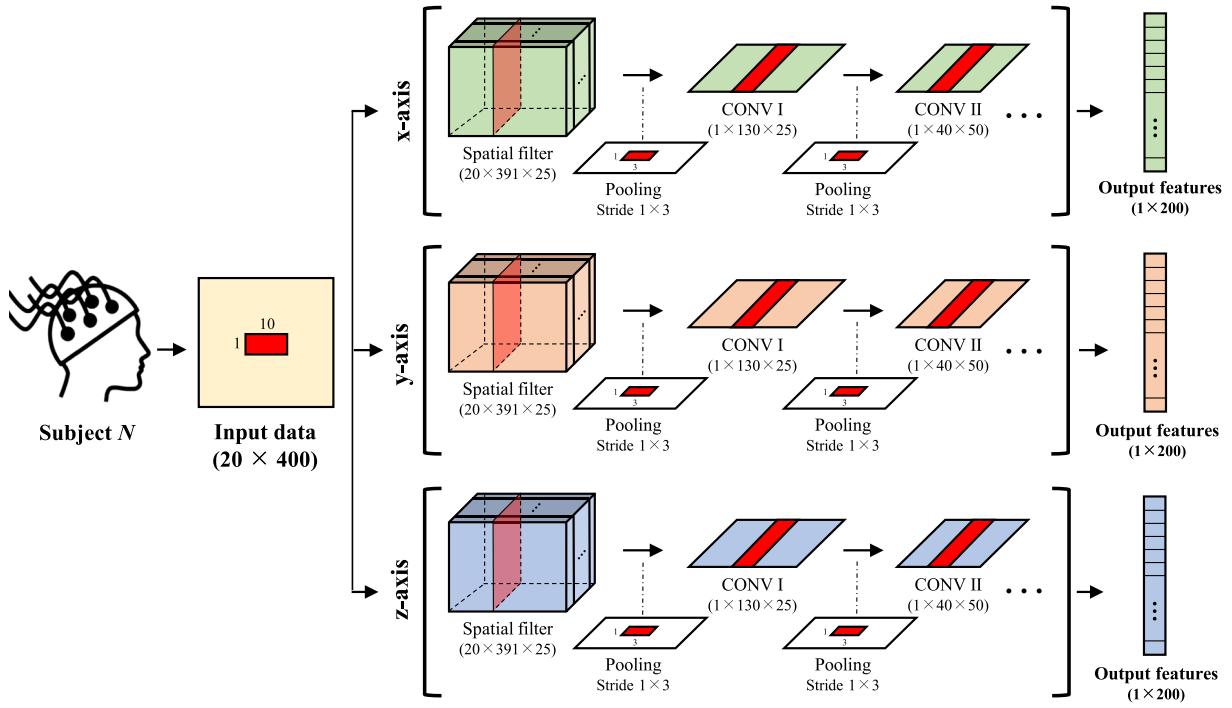


Fig. 4. The flowchart of the CNN architecture pretraining the velocity profile for the representative subject N . The CNN trains for a positive or negative direction according to the x-, y-, z-axis using the baseline velocity profile. The output features were flattened to become an input of each BiLSTM cell for each axis.

performance validation. The entire dataset included 240 trials comprising by 40 trials per each class. To avoid overfitting the learning problem for a particular class, we selected the data in a fair manner by using 80% of the data in each class as training data and the remaining 20% as test data. Hence, the data from 192 trials (i.e., 32 trials \times 6 classes) were assigned for training, and the data from the remaining 48 trials were designated as test data for the improvement of decoding performance. By use of the CNN architecture, the MDCBN was pretrained,

and the EEG signals per axis were able to represent the positive or negative directions using the baseline velocity profile according to the directions. Each axis of the CNN adopted the principle of the DeepConvNet architecture [56], which was modified to extract coarse features considering each axis (Fig. 4).

During pretraining, the input data were EEG signals with a fixed-size 20×400 (channel \times time). The data were passed through a temporal filter with a receptive field of 1×10 size

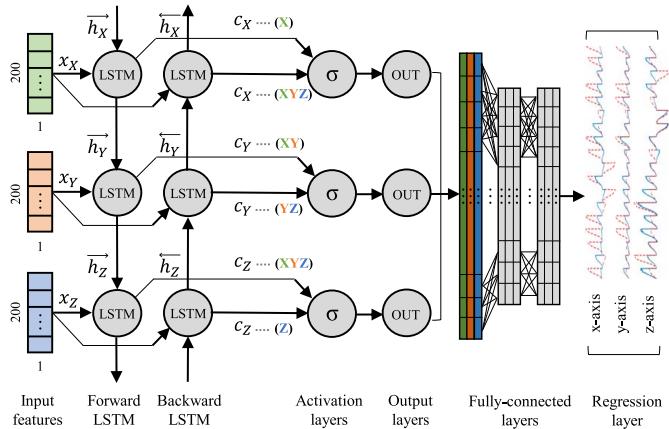


Fig. 5. The flowchart of the BiLSTM network for predicting the velocity profile of multi-direction. The BiLSTM network comprises LSTM cells according to each axis, and the LSTM cells are able to reflect the properties of the xyz-plane in 3D space due to the BiLSTM characteristics.

and a spatial filter using a 20×1 receptive field. Then, an average pooling layer was used to resize the convolution, which had a 1×3 kernel size with a stride of 1×3 . The convolutional layers were set using the 1×10 receptive field. Sequentially, the three convolutional layers (CONV I, CONV II, CONV III) and the pooling layers were used to pre-train the weight. The dense layer was set to classify the positive and negative directions across each axis. Finally, each of the 1×200 features from the x-, y-, and z-axes become an input feature of each cell of the BiLSTM network. A sequence of the features is represented on the x_{axis} ($axis = X, Y$, and Z) and the output of the BiLSTM is trained recursively by activation of the units in the network with the following equations for each axis:

$$i_{axis} = \sigma(W_{xi}x_{axis} + W_{hi}h_{axis} + W_{ci}c_{axis} + b_i), \quad (1)$$

$$f_{axis} = \sigma(W_{xf}x_{axis} + W_{hf}h_{axis} + W_{cf}c_{axis} + b_f), \quad (2)$$

$$c_{axis} = f_{axis}c_{axis} + i_{axis}\varphi(W_{xc}x_{axis} + W_{hc}h_{axis} + b_c), \quad (3)$$

$$o_{axis} = \sigma(W_{xo}x_{axis} + W_{ho}h_{axis} + W_{co}c_{axis} + b_o), \quad (4)$$

$$h_{axis} = o_{axis}\varphi(c_{axis}) \quad (5)$$

where h_{axis} is hidden vectors with the subscription $axis$ and i_{axis} , f_{axis} , c_{axis} , o_{axis} denote the activation vectors of the input gate, the forget gate, the memory cell, and the output gate, respectively. σ is the sigmoid function defined as $\sigma_x = \frac{1}{1+\exp^{-x}}$ and φ is the hyperbolic tangent function. The variable W and b are the weight matrix and bias that need to be trained.

For considering the 3D space properties of the multi-direction, we used the BiLSTM network to extract the features according to the axis dependency rather than the time-dependence (Fig. 5). Each 1×200 extracted features from the x-, y-, and z-axes were fed into three memory cells in a forward LSTM layer and a backward LSTM layer, respectively. Next, the BiLSTM network was trained by the principle known as a coarse-to-fine approach. The features of the x-axis appear at the beginning of the forward LSTM training, and the properties of the y-axis and z-axis are sequentially trained into the memory cells. Hence, the first memory cell of the forward LSTM layer can train the features of the x-plane (1D) and,

sequentially, the memory cells reflect the properties of the xy-plane (2D) and xyz-plane (3D). The backward LSTM layer trained the features of each axis in the same manner but in the reverse order.

In this manner, the BiLSTM network can be trained by the 3D information to predict the velocity profile according to each axis. The forward and backward LSTM layers were set with a 0.5 dropout ratio. Each LSTM cell retrieves 300 hidden units, therefore, the output of each forward and backward LSTM has a size of 3×300 , which includes all of the 3D space. We used the fully connected layer with the rectified linear unit to extract nonlinear features that were adopted between the connected layers. Finally, we predict the velocity profiles for each axis using a regression layer.

We performed 500 iterations (epochs) for the model training process and saved the model weights and hyper-parameters that produced the lowest loss of the test data.

F. Continuous Decoding Strategy for Online Experiments

For real-time robotic arm control in the online experiment, we modified the MDCBN framework for the continuous decoding of user intention. We first downsampled the EEG data to 100 Hz and selected twenty EEG channels to be used as model training. The spectral and spatial filters were applied for each time window. The filters were the same as those used in the model training. We set the sliding window length to 4 s, and shifted it every 0.5 s. In the preprocessing step, we first removed the artifact signals with respect to electrooculography (EOG) activity from Fp1 and Fp2 channels using ICA. Additionally, the parameters used in the model training were equally applied for continuous decoding.

The predicted velocity profile of each time window was validated using the baseline velocity profiles (Fig. 3(b)) of the six directions. We calculated each Pearson correlation coefficient (CC) between the predicted velocity profile and baseline velocity profiles of the six directions. Each correlation coefficient (r) was transformed using a sigmoid kernel function ($S(r_i)$) that can assign a high weight to values with a high correlation coefficient (above 0.5). The $S(r_i)$ is computed by

$$S(r_i) = \frac{1}{1 + \exp^{-(order \times (r_i - 0.5))}} \quad (6)$$

where i of the r_i indicated the order of six directions, $i = 1$ (left), 2 (right), 3 (forward), 4 (backward), 5 (up), and 6 (down). The $order$ is eight, and the range of r_i is $[-1, 1]$ such that a criterion of the function is 0.5. The kernel function maximizes the likelihood of high correlation coefficient values so that the sigmoid value of the directions for each window were determined. Next, we computed the output probability (Out_i) for decoding six directions using softmax functions according to each window.

$$Out_i = \frac{\exp^{S(r_i)}}{\sum_{i=1}^6 \exp^{S(r_i)}} \quad (7)$$

To obtain the final decision to control commands, we calculated the average values of output probabilities for three successive windows depending on each direction. Finally,

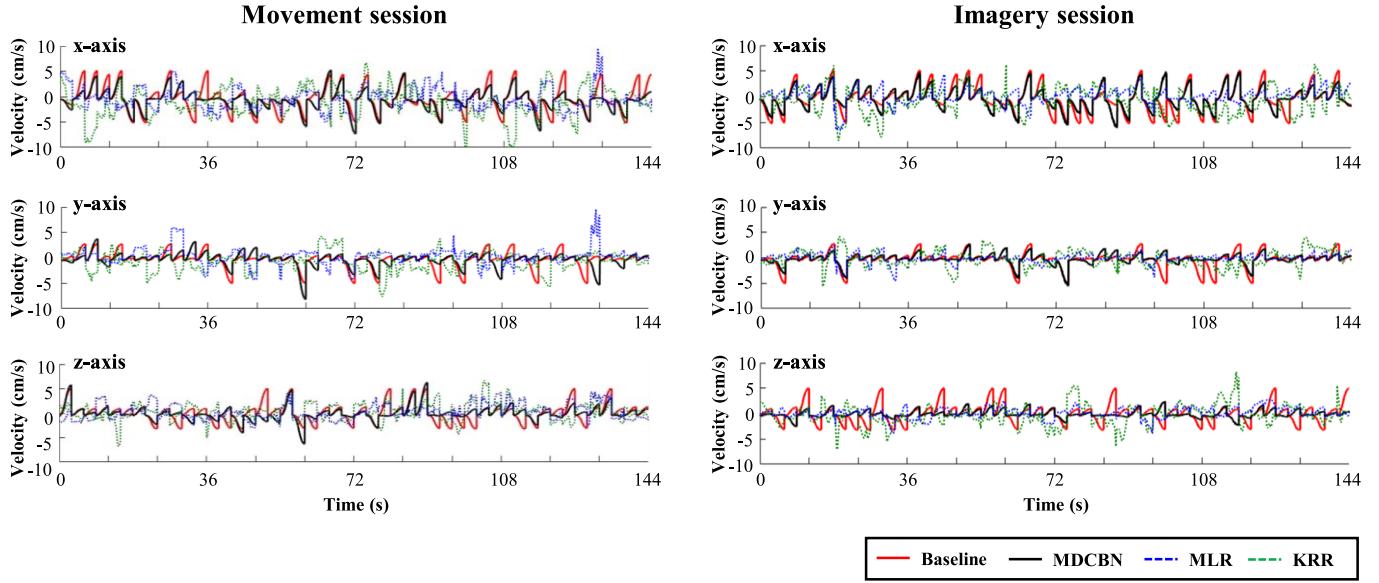


Fig. 6. An example comparison between the baseline velocity (red line) and the decoded velocity (MDCBN: black line, KRR: green dashed line, and LSTM: blue dashed line) for each movement session and imagery session from subject S1. Each decoded velocity was tested in 48 trials that were not included in the model training.

we selected the direction with a maximum probability as the final real-time robotic arm control command.

$$avgOut_i = \frac{1}{3} \sum_{window=1}^3 Out_{i,window} \quad (8)$$

$$Final\ decision = \max(avgOut_i) \quad (9)$$

III. RESULTS

A. Velocity Profiles Decoding for 3D Multi-Direction

Table II shows the decoding performance comparison on the test data using the proposed MDCBN method and conventional methods: multiple linear regression (MLR) [57], KRR [9], LSTM [58], ShallowConvNet [56], and EEG-Net [59]. We measured the CC and the normalized root mean square error (NRMSE) per axis for both movement session and imagery session. The CC (r -value) is computed between the predicted velocity profile using each method and the baseline velocity profile. In addition, we measured the decoding performances (CC and NRMSE) for 3D axes, including the x-, y-, and z-axes, to consider 3D control. For a better quantitative comparison between each method and the baseline, we also performed a statistical analysis using the paired t -test. All p -values were below 0.005 in our experiments.

Using the proposed method, the grand-averaged CC and the NRMSE of the 3D axes were 0.4712 and 0.1780, respectively, for the six directions in the movement session. In the imagery session, the averaged CC and NRMSE were 0.4575 and 0.1685 across all subjects. In particular, subject S1 showed the maximum CC values of 0.7650 and 0.7096 for each session, thus reflecting a strong positive correlation between the predicted value and the baseline value. Subjects S5 and S4 showed the minimum performances for each movement session (CC: 0.3198) and imagery session (CC: 0.2243), respectively. However, most subjects showed a coefficient

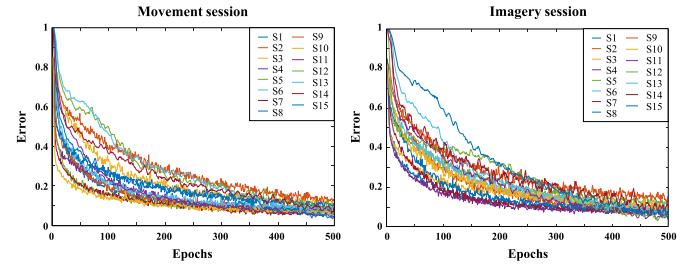


Fig. 7. Convergence curve of the training error across all subjects for each movement session and imagery session. The errors were calculated to extract the optimized parameters during the network training.

value that corresponded to greater than a moderate positive correlation ($r \geq 0.4$). Compared to conventional methods, the MDCBN method showed relatively high decoding performances for both sessions. The ShallowConvNet and EEGNet exhibited a high correlation among the conventional methods, and the MDCBN exhibited a performance difference of approximately 0.2 on the average ($p \leq 0.01$). MLR, a traditional nonlinear method, and KRR, a linear method for regression analysis, showed very weak correlations of less than 0.1. The LSTM method also exhibited low decoding performance of less than 0.1 except for a few subjects in both sessions. Additionally, the decoding performances using MLR, KRR, and LSTM were significantly different when compared to the performances of the proposed method ($p \leq 0.005$).

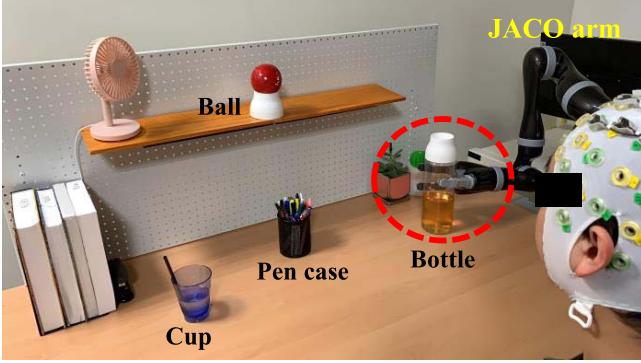
Fig. 6 reports the decoded velocity profile of the axes (x-, y-, and z-axis) for a representative subject S1 performing both movement session and imagery session. The predicted velocity profiles were evaluated using the test data set that included all directions (20% of the entire dataset). The grand-averaged velocity of the movement session was used as the baseline velocity profile for each subject. In the example shown in Fig. 6, the decoding coefficients of the velocity profile between the baseline and the proposed MDCBN

TABLE II

DECODING PERFORMANCE COMPARISON USING MLR, KRR, LSTM, SHALLOWCONVNET, EEGNET, AND MDCBN FOR BOTH SESSIONS

		Movement session							Imagery session											
Subject	Methods	CC			NRMSE	CC			NRMSE	Subject	Methods	CC			NRMSE					
		x-axis	y-axis	z-axis		3D axes	3D axes	3D axes				x-axis	y-axis	z-axis						
S1	MLR	-0.0867	-0.1045	0.1896	0.0027	0.2062	S1	MLR	-0.0497	-0.0621	-0.0180	-0.0248	0.2020	S1	MLR	-0.0497	-0.0621	-0.0180	-0.0248	0.2020
	KRR	-0.1444	-0.0022	0.0859	-0.0293	0.2933		KRR	0.0657	-0.0827	0.0364	0.0331	0.2693		KRR	0.0657	-0.0827	0.0364	0.0331	0.2693
	LSTM	-0.0627	-0.0313	0.0903	-0.0025	0.2592		LSTM	0.0419	0.1387	0.1600	0.0940	0.2117		LSTM	0.0419	0.1387	0.1600	0.0940	0.2117
	ShallowConvNet	0.7741	0.6225	0.4546	0.6389	0.1355		ShallowConvNet	0.5139	0.5802	0.4647	0.4040	0.1684		ShallowConvNet	0.5139	0.5802	0.4647	0.4040	0.1684
	EEGNet	0.5888	0.3719	0.5727	0.7261	0.1412		EEGNet	0.4139	0.3075	0.4754	0.424	0.2360		EEGNet	0.4139	0.3075	0.4754	0.424	0.2360
	MDCBN	0.7458	0.7933	0.6027	0.7650	0.1150		MDCBN	0.9631	0.6364	0.4691	0.7096	0.1382		MDCBN	0.9631	0.6364	0.4691	0.7096	0.1382
S2	MLR	-0.0002	0.0466	-0.0891	0.0164	0.1923	S2	KRR	-0.0032	0.2579	-0.1759	0.0126	0.2837	S2	MLR	-0.0097	0.0563	-0.0152	0.0390	0.1931
	KRR	-0.0032	0.2579	-0.1759	0.0126	0.2837		KRR	0.0338	0.0319	-0.1124	0.0257	0.2738		KRR	0.0338	0.0319	-0.1124	0.0257	0.2738
	LSTM	0.1247	0.1442	-0.0910	0.0729	0.2148		LSTM	0.2776	0.0006	-0.0667	0.1153	0.2034		LSTM	0.2776	0.0006	-0.0667	0.1153	0.2034
	ShallowConvNet	0.5447	0.2703	0.2260	0.3697	0.1844		ShallowConvNet	0.0776	-0.0597	0.0801	0.0135	0.2104		ShallowConvNet	0.0776	-0.0597	0.0801	0.0135	0.2104
	EEGNet	0.4101	0.4902	0.2599	0.4401	0.2387		EEGNet	0.0690	0.1411	0.2756	0.0124	0.4071		EEGNet	0.0690	0.1411	0.2756	0.0124	0.4071
	MDCBN	0.5789	0.4597	0.3752	0.5316	0.1721		MDCBN	0.5207	0.7335	0.6562	0.4943	0.1529		MDCBN	0.5207	0.7335	0.6562	0.4943	0.1529
S3	MLR	-0.2482	-0.1230	-0.0796	-0.1539	0.2018	S3	KRR	-0.0254	-0.0902	-0.1105	-0.0567	0.2698	S3	KRR	-0.2158	-0.0301	-0.0180	-0.1042	0.2839
	LSTM	0.1221	0.1502	-0.0993	0.0779	0.2628		LSTM	0.0374	-0.1222	0.0168	0.0238	0.2087		LSTM	0.0374	-0.1222	0.0168	0.0238	0.2087
	ShallowConvNet	0.0850	0.2524	0.1386	0.1996	0.2035		ShallowConvNet	0.4597	0.2770	0.0443	0.3381	0.1855		ShallowConvNet	0.4597	0.2770	0.0443	0.3381	0.1855
	EEGNet	0.1422	0.3097	0.2041	0.2112	0.2756		EEGNet	0.3231	0.3435	0.2049	0.2994	0.3058		EEGNet	0.3231	0.3435	0.2049	0.2994	0.3058
	MDCBN	0.3640	0.2794	0.3347	0.5854	0.1802		MDCBN	0.8002	0.4374	0.8999	0.6661	0.1362		MDCBN	0.8002	0.4374	0.8999	0.6661	0.1362
	MLR	0.0989	0.2178	-0.1260	0.0667	0.1910		KRR	0.1221	0.1502	-0.0993	0.0779	0.2628		KRR	0.0348	-0.1795	-0.0137	-0.0620	0.2894
S4	LSTM	0.2308	0.3990	0.1174	0.2193	0.2015	S4	MLR	-0.0621	-0.1908	0.1095	-0.0505	0.2259	S4	MLR	-0.0621	-0.1908	0.1095	-0.0505	0.2259
	ShallowConvNet	-0.0380	0.0576	0.0891	0.0051	0.2076		EEGNet	0.1324	0.2564	0.1325	0.2071	0.3527		EEGNet	0.1324	0.2564	0.1325	0.2071	0.3527
	EEGNet	0.2647	0.1896	0.0725	0.1294	0.2851		MDCBN	0.1835	0.2381	0.1416	0.2243	0.1975		MDCBN	0.1835	0.2381	0.1416	0.2243	0.1975
	MDCBN	0.3199	0.4852	0.1007	0.3365	0.2112		MLR	0.1611	-0.0040	-0.0576	0.0991	0.1885		MLR	0.1611	-0.0040	-0.0576	0.0991	0.1885
	MLR	-0.2116	-0.2127	0.0798	-0.0876	0.1988		KRR	0.1325	-0.0591	-0.0520	0.0555	0.2881		KRR	0.1325	-0.0591	-0.0520	0.0555	0.2881
	KRR	-0.1028	-0.1252	-0.0867	-0.0976	0.3192		LSTM	0.0272	0.2699	0.0548	0.2200	0.2046		LSTM	0.0272	0.2699	0.0548	0.2200	0.2046
S5	LSTM	0.0273	0.2447	-0.0040	0.0623	0.2043	S5	ShallowConvNet	0.3725	0.0449	0.0142	0.3240	0.1856	S5	ShallowConvNet	0.3725	0.0449	0.0142	0.3240	0.1856
	ShallowConvNet	0.2956	0.4229	0.2811	0.3131	0.1902		EEGNet	0.2745	0.6544	0.3698	0.3605	0.2536		EEGNet	0.2745	0.6544	0.3698	0.3605	0.2536
	EEGNet	0.3469	0.3087	0.3040	0.3083	0.2375		MDCBN	0.5333	0.5425	0.3767	0.5485	0.1466		MDCBN	0.5333	0.5425	0.3767	0.5485	0.1466
	MDCBN	0.2993	0.4538	0.3633	0.3198	0.1883		MLR	0.0375	0.0542	-0.0008	0.0605	0.1920		MLR	0.0375	0.0542	-0.0008	0.0605	0.1920
	MLR	0.0442	0.1372	-0.0767	0.0609	0.1938		KRR	0.0853	0.0232	-0.0675	0.0291	0.2795		KRR	0.0853	0.0232	-0.0675	0.0291	0.2795
	KRR	0.0558	0.3003	0.2219	0.1805	0.2588		LSTM	0.0257	0.1509	0.0408	0.0621	0.2101		LSTM	0.0257	0.1509	0.0408	0.0621	0.2101
S6	LSTM	0.0389	0.1442	0.1872	0.0973	0.2033	S6	ShallowConvNet	0.1958	0.2072	0.3142	0.2267	0.1879	S6	ShallowConvNet	0.1958	0.2072	0.3142	0.2267	0.1879
	ShallowConvNet	0.3751	0.3383	0.3396	0.3669	0.1775		EEGNet	0.0207	0.1808	0.3381	0.0872	0.3402		EEGNet	0.0207	0.1808	0.3381	0.0872	0.3402
	EEGNet	0.3074	0.2835	0.2661	0.2794	0.2655		MDCBN	0.3985	0.2770	0.6492	0.4773	0.1694		MDCBN	0.3985	0.2770	0.6492	0.4773	0.1694
	MDCBN	0.3189	0.2931	0.1335	0.4435	0.1903		MLR	-0.1081	-0.0247	0.0503	-0.0351	0.1950		MLR	-0.1081	-0.0247	0.0503	-0.0351	0.1950
	MLR	-0.1013	-0.0293	-0.1176	-0.0528	0.2004		KRR	-0.0603	-0.0193	0.1087	0.0107	0.2712		KRR	-0.0603	-0.0193	0.1087	0.0107	0.2712
	KRR	-0.0400	-0.0637	-0.0277	-0.0330	0.2916		LSTM	0.0224	0.1354	0.0599	0.0625	0.2106		LSTM	0.0224	0.1354	0.0599	0.0625	0.2106
S7	LSTM	-0.0030	-0.0373	-0.1918	-0.0507	0.2330	S7	ShallowConvNet	-0.0177	0.2090	0.2568	0.0416	0.2175	S7	ShallowConvNet	-0.0177	0.2090	0.2568	0.0416	0.2175
	ShallowConvNet	0.3753	0.4549	0.6664	0.4107	0.1743		EEGNet	0.0166	0.0298	0.2530	0.0614	0.3487		EEGNet	0.0166	0.0298	0.2530	0.0614	0.3487
	EEGNet	0.4108	0.1007	0.5432	0.4829	0.2165		MDCBN	0.1539	0.4164	0.4173	0.3733	0.2018		MDCBN	0.1539	0.4164	0.4173	0.3733	0.2018
	MDCBN	0.6214	0.2275	0.5846	0.4453	0.1852		MLR	0.0576	-0.0113	-0.0325	0.0584	0.1888		MLR	0.0576	-0.0113	-0.0325	0.0584	0.1888
	MLR	0.0551	-0.0446	0.0471	0.0866	0.1882		KRR	0.2333	0.0545	-0.1303	0.0988	0.2887		KRR	0.2333	0.0545	-0.1303	0.0988	0.2887
	KRR	-0.0357	-0.0317	0.0872	0.0060	0.2840		LSTM	0.2553	0.2665	-0.0674	0.1780	0.2066		LSTM	0.2553	0.2665	-0.0674	0.1780	0.2066
S8	LSTM	-0.1127	-0.1171	0.1259	-0.0222	0.2302	S8	ShallowConvNet	0.1201	0.2201	0.0621	0.0626	0.1991	S8	ShallowConvNet	0.1201	0.2201	0.0621	0.0626	0.1991
	ShallowConvNet	-0.0227	0.2630	0.1321	-0.0023	0.1977		EEGNet	-0.0282	0.1804	0.2282	-0.0002	0.3692		EEGNet	-0.0282	0.1804	0.2282	-0.0002	0.3692
	EEGNet	0.2123	0.3035	0.1525	0.1334	0.2879		MDCBN	0.3802	0.4232	<b									

**(a) Online experiment I:
Reach-and-grasp some objects**



**(b) Online experiment II:
Drinking a glass of water or beverage**

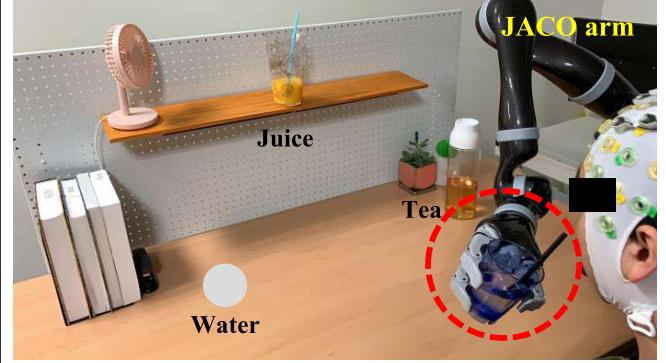


Fig. 8. A brain-controlled robotic arm control system for two different online experiments. **(a)** Reach-and-grasp some objects (ball, cup, pen case, and bottle) located in different directions. **(b)** Drinking a glass of water or beverage placed in the ‘left’, ‘forward’, and ‘right’ directions. The ‘backward’ direction was used for bringing cups or bottles to the mouth.

method had values of approximately 0.7458 (x-axis), 0.7933 (y-axis), and 0.6027 (z-axis) in the movement session. In the imagery session, the values of the decoding coefficients were approximately 0.9631 (x-axis), 0.6364 (y-axis), and 0.4691 (z-axis) per axis. For comparison with representative regression methods, the MLR method (blue dotted line) also showed decoding coefficient values of the velocity profiles as being less than 0.1 for all axes in both sessions. The KRR method (green dotted line) showed the low coefficient values for the velocity profile of -0.1444 (x-axis), -0.0022 (y-axis), and 0.0859 (z-axis) in the movement session and 0.0657 (x-axis), -0.0827 (y-axis), and 0.0364 (z-axis) in the imagery session.

B. Convergence Process of Training Model

To observe the convergence process of the MDCBN for training decoding model, the error change curves across all subjects per session are shown in Fig. 7 along with the epochs. In the training stage, the MDCBN learned the decoding information of six directions from EEG data. The data for model training and verification were collected from only 240 trials per each subject. However, the proposed MDCBN comprised much shallow architecture than the other existing deep neural networks. Therefore, despite using a single subject’s EEG data, the proposed model could converge along with the learning iteration. In this study, most of the subjects exhibited convergence within 300~400 epochs in the movement session and imagery session.

C. Real-Time Robotic Arm Control Based on EEG

To validate the possibility of real-time robotic arm control using the MDCBN model, we conducted two online experiments. Fig. 8 presents the experimental environments for each online scenario. In the first experiment, experiment I, the subjects were asked to perform reach-and-grasp tasks using objects placed at various locations. The objects (a bottle, a cup, a ball, and a pen case) were located along each base direction of the home position for the robotic arm (Fig. 8(a)). In the

experiment II, the subjects were asked to drink a glass of water or beverage. The cups and a bottle containing water and other beverages were also placed at the various locations along each direction. In this case, the subjects performed ‘backward’ imagination to bring the cups or bottle to their mouth (Fig. 8(b)). A movement calibration process for the pre-assigned positions was performed using the robotic arm before the online experiments. All grasping motions of the robotic arm were carried out automatically after the reaching motions. Additionally, the scenarios of the experiments did not require a ‘down’ command as a control command. However, for a fair evaluation, the training model learned by data from all six directions was used so that the ‘down’ decision was evaluated as a misclassification.

To cover the real-time decoding performance of subjects who showed insufficient decoding performance in the offline analysis, we adopted a synchronized real-time paradigm according to a single-trial. The subjects could focus on just the single-trial that controls the robotic arm without having to constantly concentrate during the experiment. In addition, as mentioned in the ‘F. Continuous Decoding Strategy for Online Experiments’ of II. MATERIALS AND METHODS section, the final decisions as control commands were assigned by using the class which was selected as having the highest output probability over three sub-trial. Therefore, if the wrong class showed the highest probability as one of the decoding output, we could consider additional decoding outputs with the highest probability within the rest of the sub-trial as the final decision. The success rate of the entire experiment was defined as the number of times that the task was performed thoroughly without missing a motion during three runs (10 trials for each run). Tables III and IV show the success rates of two online experiments across fifteen subjects. In online experiment I, the subjects showed a grand-averaged success rate of 0.60 (± 0.14) across all subjects. The success rate of some subjects (S2, S3, S5, S6, S8, S12, S13, and S14) improved as the number of runs increased. Subject S3 had the best success rate of the experiment and showed only one misclassification

TABLE III
EVALUATION PERFORMANCE FOR ONLINE EXPERIMENT I

Subjects	Success rate			
	Run I	Run II	Run III	Average
S1	0.7 (7/10)	0.6 (6/10)	0.7 (7/10)	0.67 (± 0.06)
S2	0.4 (4/10)	0.6 (6/10)	0.8 (8/10)	0.60 (± 0.20)
S3	0.8 (8/10)	0.9 (9/10)	0.9 (9/10)	0.87 (± 0.06)
S4	0.3 (3/10)	0.5 (5/10)	0.4 (4/10)	0.40 (± 0.10)
S5	0.6 (6/10)	0.8 (8/10)	0.8 (8/10)	0.74 (± 0.12)
S6	0.5 (5/10)	0.7 (7/10)	0.7 (7/10)	0.64 (± 0.12)
S7	0.5 (5/10)	0.4 (4/10)	0.6 (6/10)	0.50 (± 0.10)
S8	0.4 (4/10)	0.5 (5/10)	0.5 (5/10)	0.47 (± 0.06)
S9	0.5 (5/10)	0.6 (6/10)	0.4 (4/10)	0.50 (± 0.10)
S10	0.5 (5/10)	0.4 (4/10)	0.3 (3/10)	0.40 (± 0.10)
S11	0.6 (6/10)	0.8 (8/10)	0.7 (7/10)	0.70 (± 0.10)
S12	0.6 (6/10)	0.7 (7/10)	0.7 (7/10)	0.67 (± 0.06)
S13	0.5 (5/10)	0.7 (7/10)	0.7 (7/10)	0.63 (± 0.12)
S14	0.4 (4/10)	0.5 (5/10)	0.5 (0/10)	0.47 (± 0.06)
S15	0.6 (6/10)	0.8 (8/10)	0.7 (0/10)	0.70 (± 0.10)

TABLE IV
EVALUATION PERFORMANCE FOR ONLINE EXPERIMENT II

Subjects	Success rate			
	Run I	Run II	Run III	Average
S1	0.4 (4/10)	0.6 (6/10)	0.6 (6/10)	0.53 (± 0.12)
S2	0.3 (3/10)	0.5 (5/10)	0.5 (5/10)	0.43 (± 0.12)
S3	0.6 (6/10)	0.5 (5/10)	0.6 (6/10)	0.57 (± 0.06)
S4	0.2 (2/10)	0.4 (4/10)	0.3 (3/10)	0.30 (± 0.10)
S5	0.3 (3/10)	0.4 (4/10)	0.5 (5/10)	0.40 (± 0.10)
S6	0.2 (2/10)	0.4 (4/10)	0.4 (4/10)	0.33 (± 0.12)
S7	0.4 (4/10)	0.3 (3/10)	0.5 (5/10)	0.40 (± 0.10)
S8	0.3 (3/10)	0.4 (4/10)	0.3 (3/10)	0.33 (± 0.06)
S9	0.4 (4/10)	0.5 (5/10)	0.3 (3/10)	0.40 (± 0.10)
S10	0.4 (4/10)	0.3 (3/10)	0.3 (3/10)	0.33 (± 0.06)
S11	0.5 (5/10)	0.6 (6/10)	0.5 (5/10)	0.53 (± 0.06)
S12	0.4 (4/10)	0.6 (6/10)	0.5 (5/10)	0.50 (± 0.10)
S13	0.3 (3/10)	0.7 (7/10)	0.6 (6/10)	0.53 (± 0.21)
S14	0.4 (4/10)	0.3 (3/10)	0.3 (3/10)	0.33 (± 0.06)
S15	0.4 (4/10)	0.5 (5/10)	0.5 (5/10)	0.47 (± 0.06)

for each Run II and Run III. In the online experiment II, the subjects showed a low success rate of less than 0.5 across all subjects (Table IV). In this experiment, if the two reaching tasks were executed by all correct commands, it was counted as a success. The ‘left’, ‘up’, and ‘right’ tasks were used in arm reaching, and the ‘backward’ task was used for raising the cup or bottle to the mouth. The subjects presented a grand-averaged success rate of 0.43 (± 0.09), and subject S13 showed the highest success rate of 0.7 in Run II. However, most subjects did not show a significant improvement in the success rate over 30 trials, and they stated that it was difficult to continuously imagine two motions involving for high-level tasks in experiment II.

Fig. 9 presents the trajectory pathway reconstruction of the predicted velocity profile through the online experiment for representative subject S3. We mapped the velocity profile to the 3D position dimension (3D axes) to represent the arm trajectory pathway according to the direction. Subject S3 achieved decoding performance of 0.6661 (CC) and 0.1362

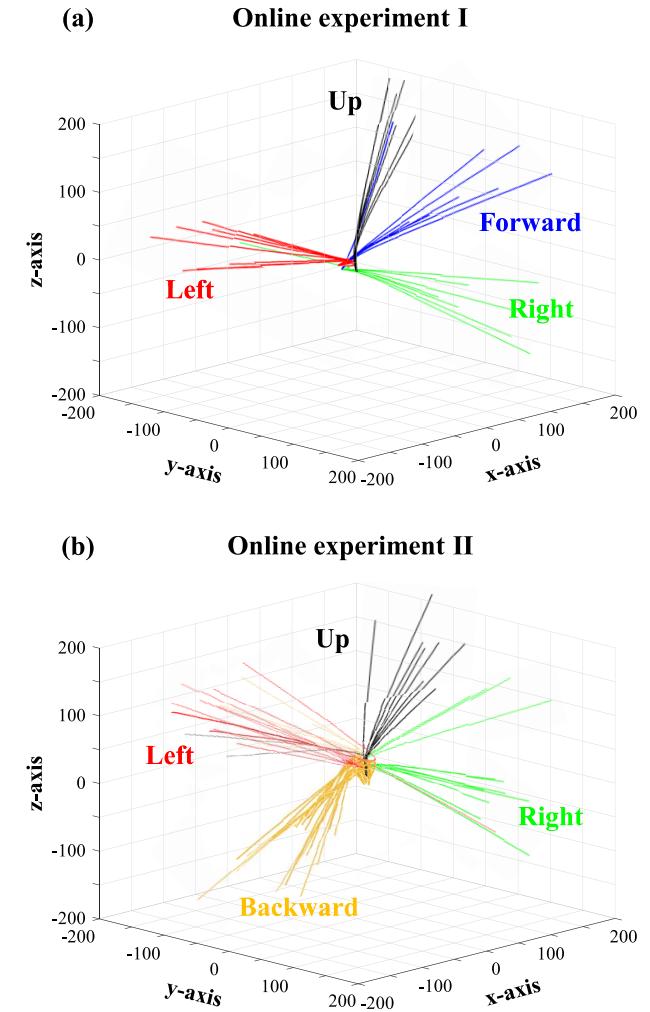


Fig. 9. The decoded arm trajectory pathway of two online experiments for subject S3. The trajectory pathway is indicated by different colors per direction (left: red, right: green, forward: blue, up: black, backward: yellow).

(NRMSE) in the imagery session. Furthermore, in the online experiments, subject S3 showed success rates of 0.87 and 0.56, respectively. The decoded trajectory pathways were well distinguished according to the directions. Only one trial for ‘forward’ and one trial for ‘right’ were misclassified (Fig. 8(a)), and several trajectory estimations of the ‘backward’ task tended to be confused with the ‘left’ task (Fig. 8(b)).

IV. DISCUSSION

In this study, we demonstrated the feasibility of healthy subjects operating a brain-controlled robotic arm that employed a deep learning technique with a high DoF. The proposed deep learning framework is based on the MDCBN that can robustly decode multi-directional arm reaching from EEG signals. In the offline analysis, the experimental results showed that the proposed framework decoded the intuitive imagery more accurately with higher performance than other methods for both the movement session and the imagery session (Table II). In addition, the success rates of the online experiments were approximately 0.60 (± 0.14) and 0.43 (± 0.09), respectively,

across fifteen subjects (Table III and IV). To the best of our knowledge, this system is the first to adopt a deep learning framework for EEG-based robotic arm control in real-time 3D environments, and the first to employ intuitive imagery decoding for performing high-level tasks using the same upper extremity.

Recently, BMI studies have attempted to decode user intention using a deep learning approach for several paradigms, such as motor imagery classification [47], [52], [54], [56], [60], event-related potentials (ERPs) detection [48], mental workload [49], [61], [62], and emotion recognition [63]. For robust motor imagery decoding of the natural interaction between the brain and machine, Lei *et al.* [60] proposed an effective walking imagery evaluation method based on the virtual environment to elicit reliable brain signal activities. Furthermore, a multiview multilevel deep polynomial network (MMDPN) demonstrated discrimination performances comparable to networks using other deep learning frameworks in the decoding of walking imagery. Zhang *et al.* [54] classified binary motor imagery tasks (left hand and right hand) using dataset III from BCI Competition II. They proposed a novel deep learning approach with an empirical mode decomposition method to effectively train two-layer deep networks (CNNs and wavelet neural networks). Similarly, Wang *et al.* [47] proposed a classification framework based on LSTM networks with one dimension-aggregate approximation. They evaluated the proposed method for four motor imagery tasks (left hand, right hand, both feet, and tongue). To ensure a fair comparison, they focused on a binary classification for each group comprising the combinations for each class. Likewise, Schirrmeister *et al.* [56] also used the BCI Competition IV dataset for four classes. They proposed the Deep and Shallow ConvNet frameworks for robust motor imagery decoding and visualization. The framework considers not only the spatial information of brain activity but also the spectral range.

These deep learning techniques have contributed to the robust EEG decoding of different upper or lower extremity movements despite poor signal to noise and spontaneous potentials [64]. In contrast, in this study, we verified that deep learning techniques could also be used for decoding various intuitive imageries using only a single upper extremity. Decoding various extremity movements from the single limb using EEG signals is one of the novel and challenging issues in the BMI field. A few recent studies have investigated multiclass approaches and trajectory prediction approaches using machine learning techniques [30], [32]. Although these studies have not achieved acceptable decoding performances for real-time control of BMI systems, they showed the possibility to achieve high DoF and intuitive BMI control with an artificial interaction that matched user intention with specific robotic movements. Although it was impossible to directly compare the present findings with previous works because of different protocols and environments, our results demonstrated a high decoding performance for various types of arm movements. We achieved these results by decoding EEG signals corresponding to the same extremity (Table II). The proposed MDCBN framework could possibly train the EEG signals by simultaneously considering the velocity profiles of 3D space

information (Fig. 4 and Fig. 5). Additionally, we attempted to adopt intuitive imagery decoding of the brain-controlled robotic arm with a high DoF in real-time environments. However, the results revealed that higher decoding performance are required to improve the accuracy of the online experiments. Especially, the subjects S4, S8, and S14 showed a low CC for 3D axes in the offline analysis compared to the other subjects. Although the insufficient decoding performance, they performed a real-time robotic arm control according to the experiment I and experiment II. As results depicted in Table III and Table IV, they showed the average success rate as approximately 0.4 and 0.3 in both experiment scenarios. In the real-time BCI environment, unreliable performance is one of the critical problems. Especially, in the motor imagery paradigm, the reliable decoding performance from subjects is still difficult owing to without external stimuli. Hence, in this paper, we designed a continuous decoding strategy for improving real-time robotic arm control performance as depicted in II. MATERIALS AND METHODS, F. *Continuous Decoding Strategy for Online Experiments*. The strategy could consider the nonstationary EEG characteristics during continuous decoding by calculating average output probabilities of sub-trial according to the six-direction. Therefore, we could avoid the outlier decision of output probability for the final real-time robotic arm control command. Actually, according to a few success-trial, not all three sub-trial were determined as the same with the true label but the final decision is correct. The continuous decoding strategy has affected on improving the system performance. However, it still needs more enhancing decoding performance for all subjects in real-time environments. We plan to adopt advanced deep architectures into the proposed model for guaranteeing robust continuous decoding performance when using more short time-series data.

In addition, a better training environment should be designed, as the current training environment imposed a high cognitive workload on the subjects due to the long training time. Approximately 3~4 hours is needed to collect EEG data for calibration, including the training time for a deep learning model. For this reason, some subjects performed well in the training sessions but had a low success rate because of their blurred attention in the real-time control sessions (subjects S5, S7, and S15). This problem is quite common in online BMI experiments. Thus, a few groups have attempted to use a pre-trained decoding model on another day to reduce the calibration time in using a large amount of EEG data [65]–[67] or to design the training model using only a few data samples [54]. Hence, we will investigate the design of a deep learning architecture to apply adaptation learning using only a few training data samples to solve this online BMI problem. As one of the different approaches to better training time for the online BCI system, inter-subject learning has been developing. The inter-subject BMI system could be addressed the rapid adaptation of a set of new subject's samples. However, the problems of low performance and the necessity of large dataset for reliable model performance exists. In this study, one of the ultimate goals is the development of an inter-subject brain-controlled robotic arm system with robust decoding performance. For the investigation, we have evaluated

the decoding performance for inter-subject using leave-one-subject-out (LOSO) cross-validation in motor imagery data. The results showed a low decoding performance as 0.08 for grand average CC in 3D axes. Therefore, we plan to adopt an advanced algorithm by referring to the state-of-the-art [55], [67] of inter-subject learning.

V. CONCLUSIONS AND FUTURE WORKS

In this study, we presented a brain-controlled robotic arm system based on the MDCBN-based deep learning framework. We decoded the intuitive imagery of users with respect to the arm reaching in 3D multi-direction (left, right, forward, backward, up, and right) from EEG signals. The proposed MDCBN framework was designed to predict the velocity profiles considering the 3D axes (x-, y-, and z-axis) per direction. We evaluated the framework in movement sessions and imagery sessions across fifteen subjects. The experimental results showed that the proposed network could predict the velocity profiles of the arm reaching imageries in 3D directions and could significantly improve its decoding performance as compared to the other methods. Furthermore, we demonstrated the feasibility of BMI-based robotic arm control through online experiments in a real-time environment. This feasibility has the potential to enhance rehabilitation effects and better support daily life for patients by intuitive BMI control.

In future work, we plan to confirm that the proposed framework can achieve stable decoding performance for many sessions over multiple days. Moreover, we will modify the proposed framework to allow training using fewer EEG data samples so that the calibration time required for in real-time BMI control is reduced. We believe that the brain-controlled robotic arm system will contribute to developing real-world BMI application techniques and to supporting daily life and rehabilitation for elderly and motor-disabled patients.

ACKNOWLEDGMENT

The authors thank Prof. C. Guan and Dr. K.-T. Kim for the useful discussion of the data analysis and Mr. J.-H. Cho, Mr. B.-H. Kwon, Mr. B.-H. Lee, and Ms. D.-Y. Lee for their help with the EEG database construction.

REFERENCES

- [1] T. M. Vaughan, "Guest editorial brain-computer interface technology: A review of the second international meeting," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 2, pp. 94–109, Jun. 2003.
- [2] R. Abiri, S. Borhani, E. W. Sellers, Y. Jiang, and X. Zhao, "A comprehensive review of EEG-based brain-computer interface paradigms," *J. Neural Eng.*, vol. 16, no. 1, Feb. 2019, Art. no. 011001.
- [3] 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.
- [4] Y. He, D. Eguren, J. M. Azorín, R. G. Grossman, T. P. Luu, and J. L. Contreras-Vidal, "Brain-machine interfaces for controlling lower-limb powered robotic systems," *J. Neural Eng.*, vol. 15, no. 2, Apr. 2018, Art. no. 021004.
- [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, Oct. 2015, Art. no. 056009.
- [6] J. Meng, S. Zhang, A. Bekyo, J. Olsoe, B. Baxter, and B. He, "Non-invasive electroencephalogram based control of a robotic arm for reach and grasp tasks," *Sci. Rep.*, vol. 6, no. 1, p. 38565, Dec. 2016.
- [7] C. I. Penalosa and S. Nishio, "BMI control of a third arm for multi-tasking," *Sci. Robot.*, vol. 3, no. 20, Jul. 2018, Art. no. eaat1228.
- [8] I. Iturrate, R. Chavarriaga, L. Montesano, J. Minguez, and J. D. R. Millán, "Teaching brain-machine interfaces as an alternative paradigm to neuroprosthetics control," *Sci. Rep.*, vol. 5, no. 1, p. 13893, Nov. 2015.
- [9] J.-H. Kim, F. Bießmann, and S.-W. Lee, "Decoding three-dimensional trajectory of executed and imagined arm movements from electroencephalogram signals," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 23, no. 5, pp. 867–876, Sep. 2015.
- [10] 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.
- [11] D.-O. Won, H.-J. Hwang, S. Dähne, K.-R. Müller, and S.-W. Lee, "Effect of higher frequency on the classification of steady-state visual evoked potentials," *J. Neural Eng.*, vol. 13, no. 1, Feb. 2016, Art. no. 016014.
- [12] D. Kuhner *et al.*, "A service assistant combining autonomous robotics, flexible goal formulation, and deep-learning-based brain–computer interfacing," *Robot. Auto. Syst.*, vol. 116, pp. 98–113, Jun. 2019.
- [13] M. Lee *et al.*, "Motor imagery learning across a sequence of trials in stroke patients," *Restorative Neurol. Neurosci.*, vol. 34, no. 4, pp. 635–645, Aug. 2016.
- [14] X. Chen, B. Zhao, Y. Wang, S. Xu, and X. Gao, "Control of a 7-DOF robotic arm system with an SSVEP-based BCI," *Int. J. Neural Syst.*, vol. 28, no. 8, Oct. 2018, Art. no. 1850018.
- [15] R. Spataro *et al.*, "Reaching and grasping a glass of water by locked-in ALS patients through a BCI-controlled humanoid robot," *J. Neurol. Sci.*, vol. 357, pp. e48–e49, Oct. 2015.
- [16] X. Chen, B. Zhao, Y. Wang, and X. Gao, "Combination of high-frequency SSVEP-based BCI and computer vision for controlling a robotic arm," *J. Neural Eng.*, vol. 16, no. 2, Apr. 2019, Art. no. 026012.
- [17] Y. Xu *et al.*, "Shared control of a robotic arm using non-invasive brain–computer interface and computer vision guidance," *Robot. Auto. Syst.*, vol. 115, pp. 121–129, May 2019.
- [18] 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.
- [19] B. He, B. Baxter, B. J. Edelman, C. C. Cline, and W. W. Ye, "Non-invasive brain-computer interfaces based on sensorimotor rhythms," *Proc. IEEE*, vol. 103, no. 6, pp. 907–925, Jun. 2015.
- [20] T.-E. Kam, H.-I. Suk, and S.-W. Lee, "Non-homogeneous spatial filter optimization for ElectroEncephaloGram (EEG)-based motor imagery classification," *Neurocomputing*, vol. 108, pp. 58–68, May 2013.
- [21] 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.
- [22] 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, Jun. 2011.
- [23] 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, Feb. 2018, Art. no. 016005.
- [24] X. Yong and C. Menon, "EEG classification of different imaginary movements within the same limb," *PLoS ONE*, vol. 10, no. 4, 2015, Art. no. e0121896.
- [25] I. Iturrate *et al.*, "Human EEG reveals distinct neural correlates of power and precision grasping types," *NeuroImage*, vol. 181, pp. 635–644, Nov. 2018.
- [26] B. J. Edelman, B. Baxter, and B. He, "EEG source imaging enhances the decoding of complex right-hand motor imagery tasks," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 1, pp. 4–14, Jan. 2016.
- [27] 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.
- [28] J. W. Chung, E. Ofori, G. Misra, C. W. Hess, and D. E. Vaillancourt, "Beta-band activity and connectivity in sensorimotor and parietal cortex are important for accurate motor performance," *Neuroimage*, vol. 144, pp. 164–173, 2017.
- [29] J. M. Antelis, L. Montesano, A. Ramos-Murgialday, N. Birbaumer, and J. Minguez, "Decoding upper limb movement attempt from EEG measurements of the contralateral motor cortex in chronic stroke patients," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 1, pp. 99–111, Jan. 2017.
- [30] F. Shiman *et al.*, "Classification of different reaching movements from the same limb using EEG," *J. Neural Eng.*, vol. 14, no. 4, Aug. 2017, Art. no. 046018.

- [31] V. Shenoy Handiru, A. P. Vinod, and C. Guan, "EEG source space analysis of the supervised factor analytic approach for the classification of multi-directional arm movement," *J. Neural Eng.*, vol. 14, no. 4, Aug. 2017, Art. no. 046008.
- [32] A. Korik, R. Sosnik, N. Siddique, and D. Coyle, "Decoding imagined 3D hand movement trajectories from EEG: Evidence to support the use of mu, beta, and low gamma oscillations," *Frontiers Neurosci.*, vol. 12, p. 130, Mar. 2018.
- [33] A. Úbeda, J. M. Azorín, R. Chavarriaga, and J. D. R. Millán, "Classification of upper limb center-out reaching tasks by means of EEG-based continuous decoding techniques," *J. Neuroeng. Rehabil.*, vol. 14, no. 1, pp. 1–14, Dec. 2017.
- [34] Y. Zhao *et al.*, "Automatic recognition of fMRI-derived functional networks using 3-D convolutional neural networks," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 9, pp. 1975–1984, Sep. 2018.
- [35] Y. Li, X.-R. Zhang, B. Zhang, M.-Y. Lei, W.-G. Cui, and Y.-Z. Guo, "A channel-projection mixed-scale convolutional neural network for motor imagery EEG decoding," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 6, pp. 1170–1180, Jun. 2019.
- [36] B. J. Edelman *et al.*, "Noninvasive neuroimaging enhances continuous neural tracking for robotic device control," *Sci. Robot.*, vol. 4, no. 31, Jun. 2019, Art. no. eaaw6844.
- [37] 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.
- [38] 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. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2013, pp. 3282–3285.
- [39] H. Zeng and A. Song, "Removal of EOG artifacts from EEG recordings using stationary subspace analysis," *Sci. World J.*, vol. 2014, pp. 1–9, Jan. 2014.
- [40] B. Singh and H. Wagatsuma, "A removal of eye movement and blink artifacts from EEG data using morphological component analysis," *Comput. Math. Methods Med.*, vol. 2017, pp. 1–17, 2017.
- [41] A. Jafarifarmand, M. A. Badamchizadeh, S. Khanmohammadi, M. A. Nazari, and B. M. Tazehkand, "A new self-regulated neuro-fuzzy framework for classification of EEG signals in motor imagery BCI," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 3, pp. 1485–1497, Jun. 2018.
- [42] E. Abdalsalam M, M. Z. Yusoff, D. Mahmoud, A. S. Malik, and M. R. Bahloul, "Discrimination of four class simple limb motor imagery movements for brain-computer interface," *Biomed. Signal Process. Control*, vol. 44, pp. 181–190, Jul. 2018.
- [43] C. A. Stefano Filho, R. Attux, and G. Castellano, "EEG sensorimotor rhythms' variation and functional connectivity measures during motor imagery: Linear relations and classification approaches," *PeerJ*, vol. 5, p. e3983, Nov. 2017.
- [44] B. Xu *et al.*, "Phase synchronization information for classifying motor imagery EEG from the same limb," *IEEE Access*, vol. 7, pp. 153842–153852, 2019.
- [45] P. Wei, W. He, Y. Zhou, and L. Wang, "Performance of motor imagery brain-computer interface based on anodal transcranial direct current stimulation modulation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 3, pp. 404–415, May 2013.
- [46] Y. R. Tabar and U. Halici, "A novel deep learning approach for classification of EEG motor imagery signals," *J. Neural Eng.*, vol. 14, no. 1, Feb. 2017, Art. no. 016003.
- [47] P. Wang, A. Jiang, X. Liu, J. Shang, and L. Zhang, "LSTM-based EEG classification in motor imagery tasks," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 11, pp. 2086–2095, Nov. 2018.
- [48] J. Li, Z. L. Yu, Z. Gu, W. Wu, Y. Li, and L. Jin, "A hybrid network for ERP detection and analysis based on restricted Boltzmann machine," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 3, pp. 563–572, Mar. 2018.
- [49] 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.
- [50] H.-I. Suk and S.-W. Lee, "A novel Bayesian framework for discriminative feature extraction in brain-computer interfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 2, pp. 286–299, Feb. 2013.
- [51] 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, Aug. 2015.
- [52] 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.
- [53] G. Xu *et al.*, "A deep transfer convolutional neural network framework for EEG signal classification," *IEEE Access*, vol. 7, pp. 112767–112776, 2019.
- [54] 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.
- [55] 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.
- [56] R. T. Schirrmeister *et al.*, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Hum. Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, Nov. 2017.
- [57] J. M. Antelis, L. Montesano, A. Ramos-Murguialday, N. Birbaumer, and J. Minguez, "On the usage of linear regression models to reconstruct limb kinematics from low frequency EEG signals," *PLoS ONE*, vol. 8, no. 4, 2013, Art. no. e61976.
- [58] K. M. Tsioris, V. C. Pezoulas, M. Zervakis, S. Konitsiotis, D. D. Koutsouris, and D. I. Fotiadis, "A long short-term memory deep learning network for the prediction of epileptic seizures using EEG signals," *Comput. Biol. Med.*, vol. 99, pp. 24–37, Aug. 2018.
- [59] 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, Oct. 2018, Art. no. 056013.
- [60] B. Lei *et al.*, "Walking imagery evaluation in brain computer interfaces via a new deep learning framework," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 3, pp. 497–506, 2019.
- [61] Z. Jiao, X. Gao, Y. Wang, J. Li, and H. Xu, "Deep convolutional neural networks for mental load classification based on EEG data," *Pattern Recognit.*, vol. 76, pp. 582–595, Apr. 2018.
- [62] J.-H. Jeong, B.-W. Yu, D.-H. Lee, and S.-W. Lee, "Classification of drowsiness levels based on a deep spatio-temporal convolutional bidirectional LSTM network using electroencephalography signals," *Brain Sci.*, vol. 9, no. 12, p. 348, 2019.
- [63] S. Alhagry, A. Aly, and R. A. El-Khoribi, "Emotion recognition based on EEG using LSTM recurrent neural network," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 10, pp. 355–358, 2017.
- [64] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: A review," *J. Neural Eng.*, vol. 16, no. 3, Jun. 2019, Art. no. 031001.
- [65] I. Hossain, A. Khosravi, I. Hettiarachchi, and S. Nahavandi, "Batch mode query by committee for motor imagery-based BCI," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 1, pp. 13–21, Jan. 2019.
- [66] 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, May 2019.
- [67] 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.*, pp. 1–14, 2019.