Motor Imagery Classification Using Inter-Task Transfer Learning via a Channel-Wise Variational Autoencoder-Based Convolutional Neural Network

Do-Yeun Lee, Ji-Hoon Jeong, *Associate Member, IEEE*, Byeong-Hoo Lee, and Seong-Whan Lee, *Fellow, IEEE*

***Abstract*— Highly sophisticated control based on a brain-computer interface (BCI) requires decoding kinematic information from brain signals. The forearm is a region of the upper limb that is often used in everyday life, but intuitive movements within the same limb have rarely been investigated in previous BCI studies. In this study, we focused on various forearm movement decoding from electroencephalography (EEG) signals using a small number of samples. Ten healthy participants took part in an experiment and performed motor execution (ME) and motor imagery (MI) of the intuitive movement tasks (Dataset I). We propose a convolutional neural network using a channel-wise variational autoencoder (CVNet) based on inter-task transfer learning. We approached that training the reconstructed ME-EEG signals together will also achieve more sufficient classification performance with only a small amount of MI-EEG signals. The proposed CVNet was validated on our own Dataset I and a public dataset, BNCI Horizon 2020 (Dataset II). The classification accuracies of various movements are confirmed to be 0.83 (±0.04) and 0.69 (±0.04) for Dataset I and II, respectively. The results show that the proposed method exhibits performance increases of approximately 0.09∼0.27 and 0.08∼0.24 compared with the conventional models for Dataset I and II, respectively. The outcomes suggest that the training model for decoding imagined movements can be performed using data from ME and a small number of data samples from MI. Hence, it is presented the feasibility of BCI learning strategies that can sufficiently learn deep learning with a few amount of calibration dataset and time only, with stable performance.**

***Index Terms*— Brain-computer interface, electroencephalogram, motor imagery, motor execution, deep learning.**

I. Introduction

DECODING movement intention from brain signals is one of the essential issues for controlling external devices. There are two types of brain-computer interfaces (BCIs) for approaching the issue: invasive methods and non-invasive methods [1]. Invasive BCI has used signal electrodes implanted in the brain which have high signal-to-noise ratios (SNRs). However, invasive BCIs require surgery for implanting electrodes, which exhibit long-term stability. Although non-invasive BCIs have low SNRs and low cost, they entail little risk of infection and are easy to set up for collecting brain signals. Non-invasive BCI systems often rely on electroencephalography (EEG) signals [2]. Therefore, non-invasive BCI studies could address the control of a variety of external devices, such as exoskeletons [3], wheelchairs [4], and robotic arms [5], [6]. These devices can help recover the motor functions of people with spinal cord injuries (SCI) or amyotrophic lateral sclerosis (ALS); further, they can also support the daily life of healthy people [7]. Recently, EEG-based BCIs have been extracted user intentions from EEG signals, several paradigms have been established, including steady-state visually evoked potential (SSVEP) [8], error-related potential (ErrP) [9], movement-related cortical potential (MRCP) [10], [11], as well as motor imagery (MI) and motor execution (ME) [12]–[14]. In particular, the MI paradigm has endogenous characteristics without external stimulation when performing tasks. Therefore, it can be adopted to perform tasks more intuitively than using many other EEG-based BCI paradigms.

The most common approach for decoding movement intentions from EEG signals is the use of sensorimotor rhythm (SMR) based methods, which use voluntary modulation of SMRs during imagined or real movement. Some studies [11], [15], [16] used a paradigm that decodes the user's

Manuscript received April 20, 2021; revised September 5, 2021, November 7, 2021, and December 29, 2021; accepted January 12, 2022. Date of publication January 18, 2022; date of current version January 31, 2022. This work was supported in part by the Institute of Information and Communications Technology Planning and Evaluation (IITP) Grant by the Korean Government under Grant 2017-0-00432, Grant 2017-0-00451, and Grant 2019-0-00079. *(Do-Yeun Lee and Ji-Hoon Jeong contributed equally to this work.) (Corresponding author: Seong-Whan Lee.)*

This work involved human subjects in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board of Korea University under Application No. KUIRB-2020-0013-01.

Do-Yeun Lee and Byeong-Hoo Lee are with the Department of Brain and Cognitive Engineering, Korea University, Anam-dong, Seongbuk-gu, Seoul 02841, South Korea (e-mail: doyeun_lee@korea.ac.kr; bh_lee@korea.ac.kr).

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

Digital Object Identifier 10.1109/TNSRE.2022.3143836

intentions as the task imagined, such as reaching in various directions, hand grasping, and wrist twisting. To perform more natural movements, decoding kinematic information from EEG signals is required. It is already known from previous studies that the EEG signals recorded during an MI session and the EEG signals recorded during an ME session are similar [17]. Since MI tasks entail the imagining of muscle movements [18], the user's feeling of fatigue is high, and it is impossible to confirm whether the correct imagination is performed according to the instructions given. Therefore, this paradigm is intuitive but has the disadvantage of poor signal quality. By contrast, ME tasks are easier to perform because they entail consistent motions; further, the resultant EEG signals are consistent.

In this study, we focus on deciphering various upper limb movement tasks within the same limb. In order to decode various upper limb movements, participants performed actual and imagined movements corresponding to four different tasks (forearm extension, hand grasp, wrist supination, and rest). Further, we propose a deep neural network based on a variational autoencoder (VAE) using each task transfer learning for decoding various upper limb movements. Using the ME task can reduce participant fatigue and result in more consistent and robust signal characteristics than using the MI task. Therefore, the ME-EEG signals can be applied as an advantage to replace a part of the MI task. We adopted an inter-task transfer learning method to calibrate the training data using a small number of task samples. We demonstrate that the channel-wise VAE method approach can be used to compensate for the shortcomings of the ME samples and to reconstruct them similarly to the MI samples, obtaining useful spatial information. In addition, this study confirms that the decoding performance is maintained even when the number of training samples decreases because the ME samples are used. Based on this, the proposed method shows the possibility of being used for real-time control and practical BCI systems in the future.

II. Related Work

Several methods related to upper limb decoding are summarized herein with regard to prior studies performed using the MI task. Recently, there have been studies on applying various machine learning and deep learning techniques to BCI systems based on brain signals [19]–[24]. Wu *et al.* [25] decoded three different MI tasks (left: left hand, right: right hand, down: both feet) using simplified Infomax-based spatial filters to investigate the potential of independent component analysis (ICA). Borra *et al.* [26] proposed a shallow CNN (Sinc-shallowNet) that is lightweight and interpretable by using a temporal sinc-convolutional layer.

MI-based BCI studies have focused on classifying intuitive movement tasks. Jeong *et al.* [6] classified six different reaching directions for the same limb using an MDCBN-based deep learning framework considering multiple directions in 3D. Ofner *et al.* [15] analyzed the encoding of six different upper limb movements in the time-domain of low-frequency EEG signals (<4 Hz). They also decoded imagined horizontal and vertical movements using low-frequency band EEG signals in 2D space. Korik *et al.* [16] investigated a power spectral density (PSD)-based band power time-series to improve the trajectory decoding accuracy of kinesthetically imagined 3D hand movement tasks. Edelman *et al.* [27] decoded four different types of wrist movements (flexion, extension, supination, and pronation) within the right hand using the EEG source imaging (ESI) method.

Furthermore, in recent years, studies on reducing the number of training sessions in the BCI domain have been conducted owing to participants' fatigue and time consumption issues [28]–[30] for applications. Several conventional studies have focused on EEG classification with a low number of training samples. Lu *et al.* [31] demonstrated EEG signal classification using the regularization and aggregation techniques in conjunction with the common spatial pattern (CSP) method using a small sample setting (SSS). Tu and Sun [32] proposed a subject transfer learning framework that uses data from other participants to account for long calibration times and new participants. Recently, Liang and Ma [33] also proposed a transfer learning algorithm to achieve improved results with small datasets. The multi-source fusion transfer learning (MFTL) algorithm for MI classification is executed by collecting a small EEG dataset from a target participant and using information from other participants. Jia *et al.* [34] proposed a self-adaptively denoised event-related desynchronization (ERD)-based motor intention recognition algorithm (DeERD) to train systems with a small dataset.

Generative models have recently begun to be applied to time series data. Hartmann *et al.* [35] developed a framework for generating brain signals by applying generative adversarial networks (GANs) to EEG signals. It has shown the potential of various studies such as data augmentation, EEG super-sampling, or restoration of corrupted data segments in the BCI domain. Jiao *et al.* [36] used GAN to augment the insufficient training data set. The proposed network was effective in data augmentation and classifier performance. Luo *et al.* [37] also proposed a method of augmenting EEG training dataset to improve the performance of the model. They conducted experiments with three methods, conditional Wasserstein GAN (cWGAN), selective VAE (sVAE), and selective WGAN (sWGAN). The resulting augmented training dataset significantly improved the performance of the EEG-based emotion recognition model.

In this study, we have also focused on using a small number of training samples, including those corresponding to real movements, to develop a practically applicable BCI system based on brain signals. This method provides a new solution for calibrating EEG features in MI classification, and it is expected to contribute to practical BCI advances.

III. Materials and Methods

A. Datasets

In this study, the datasets comprised ME and MI samples. We used an EEG dataset that was collected ourselves and a public dataset to validate the proposed method. These datasets are described in the following sections.

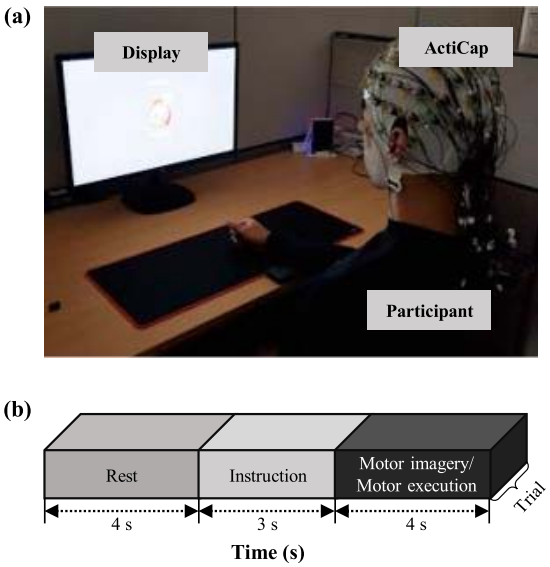

Fig. 1. Experimental environment and paradigm. (a) Environmental configuration of data acquisition. (b) Experimental paradigm. The participants perform various upper limb movements through execution and imagery sessions.

1) Dataset I (Our Dataset): Ten participants (7 males and 3 females, right-handed, age: 24-31 years) took part in the experiment. None of the participants had a history of psychiatric or motor impairment. This study was approved by the institutional review board (IRB) of Korea University [KUIRB-2020-0013-01] (January 2, 2020), and written consent according to the Declaration of Helsinki was obtained from all participants.

Fig. 1(a) and 1(b) show the experimental environment and paradigm. Participants were seated in a comfortable chair in front of a desk and asked to perform a task. The LCD display was set to a distance of approximately 80 cm from the participants [38]. The experimental paradigm was composed of ME and MI sessions. Each session was conducted as described below. First, a rest cue with a fixation cross was presented for 4 seconds at the beginning of each trial. Next, 3 seconds were provided as a time for reading instructions, which were presented via a visual cue. Participants were asked to perform the upper limb tasks, one of the three different movements (forearm extension, hand grasp, or wrist supination), following the visual cues for 4 seconds. The visual cue was randomly presented in the 50 trials per task, 150 trials in total. In the ME session, participants performed actual movements of the upper limb movement tasks. On the other hand, for the MI session, they imagined the muscle movement for the displayed task without performing any movement only. The EEG signals were measured using 60 EEG channels on the whole cortex according to the international 10-20 system with the ActiCap system and BrainAmp amplifiers (Brain Products GmbH, Germany) and MATLAB 2019b software. The ground and reference electrodes were placed at AFz and FCz, respectively. The sampling rate was 1000 Hz, and the impedance of all electrodes was maintained below 15 kΩ. The notch filter frequency rate was set at 60 Hz to reduce the DC power supply noise.

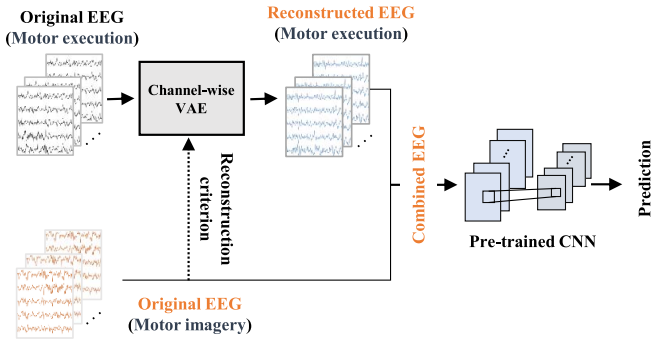

Fig. 2. Illustration of the channel-wise variational autoencoder (VAE) based on CNN (CVNet) for MI-EEG classification. The architecture comprises channel-wise VAE, feature extraction block, and a classification block.

2) Dataset II (BNCI Horizon 2020 Dataset): For upper limb movement decoding from EEG, a dataset was acquired from 15 healthy participants (age 27±5, 9 female, 1 left-handed) by Ofner *et al.* [15] and made freely available, via the BNCI Horizon 2020 database, at http://bnci-horizon-2020.eu/database/data-sets. The experiment was performed with ME in the first session and MI in the second session. The dataset comprised EEG data from 61 channels covering the frontal, central, parietal, and temporal areas using active electrodes and four 16-channel amplifiers (g.tec medical engineering GmbH, Austria). The ground and reference channels were placed on the AFz and the right mastoid, respectively. The sampling rate was 512 Hz and suppressed with a notch filter at 50 Hz because of power line interference. Each participant performed six movement types for 3 seconds: elbow flexion/extension, forearm supination/pronation, and hand open/close, all with the right upper limb. In this dataset, the recorded data corresponded to 10 runs with 42 trials per run. Overall of six movement classes and a rest class, we selected four classes (elbow extension, wrist supination, hand open, and hand close).

B. Proposed Method

The proposed method uses both ME and MI samples to classify four different upper limb tasks. We adapted the concept of transfer learning to classify the MI class using the ME dataset as part of the training data [39]. We utilize the VAE with a deep neural network in order to apply the reconstruction method to the MI classification as depicted in Fig. 2. We attempted to use the spatial features of the ME samples. However, because the noise and amplitude levels of the signals were different, we used the VAE to scale the ME samples such that the aforenoted levels were similar to those of the MI samples. In the following, the datasets are described together with the pre-processing applied to obtain the trials used to train and test the channel-wise VAE based on the convolutional neural network (CVNet) depicted in Fig. 3.

1) Data Preprocessing: In Dataset I, the recorded signal was downsampled from 1000 Hz to 250 Hz. Band-pass filtering was then performed at 4-40 Hz based on a finite impulse response (FIR) filter, and a (0∼4) s time interval

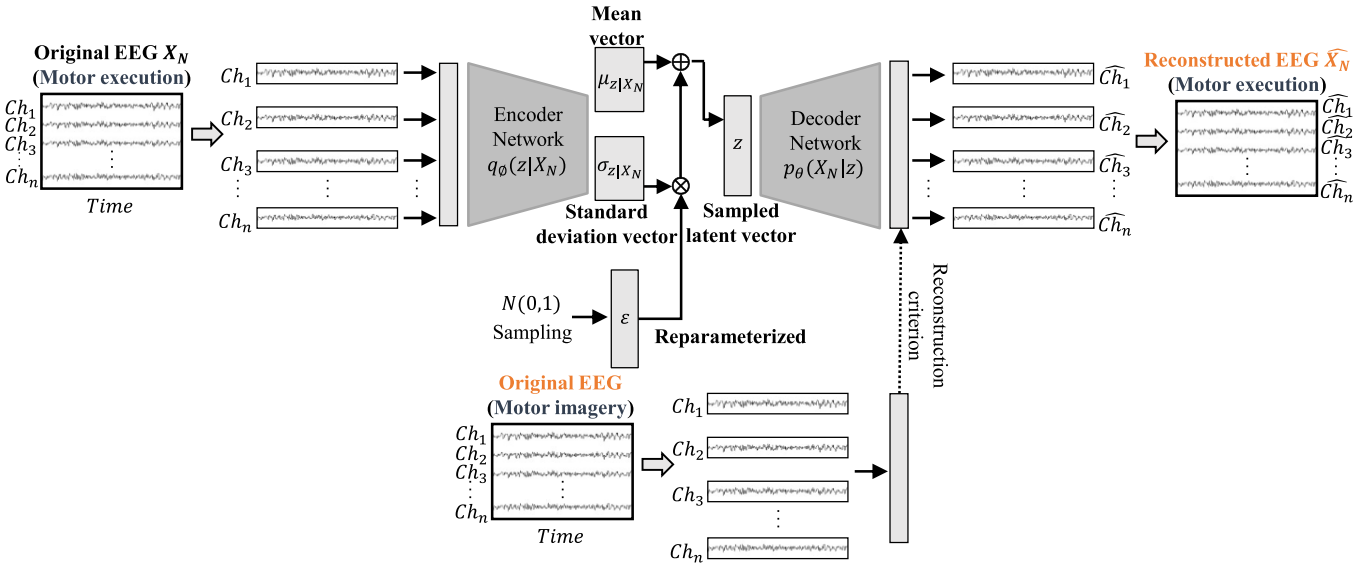

Fig. 3. Flowchart of the channel-wise variational autoencoder (VAE) for reconstructing ME-EEG signals based on original MI-based EEG signals. The channel-wise VAE framework comprises depth-wise convolution and fully-connected layers. The output of the channel-wise VAE is reconstructed ME samples.

was selected for analysis [6]. Infomax independent component analysis (ICA) was adopted to make clean EEG data. The infomax ICA is useful to decompose brain signals into statistically independent components (ICs). It could robustly remove artifacts, such as eye and head movement artifacts from the EEG signals [40], [41]. After the transforming ICA mixing matrix, the remaining ICs were projected back into the original EEG channel space to be reconstructed as the clean data. Although the signals were recorded in all areas, we selected 27 channels (F1-4, Fz, FC1-4, FCz, C1-6, Cz, CP1-4, CPz, P1-4, Pz, and POz) nearby the primary motor cortex and supplementary motor area (SMA) on the scalp. The motor cortex is related to generating neural impulses that pass down to the spinal cord and control the execution of movement. In addition, SMA is related to functions including the internally generated planning of movement and the coordination of the two sides of the body such as in bi-manual coordination. Therefore, movement-related EEG recording has been acquired in the selected channels following the 10/20 international electrode placement [42]. On the other hand, the data from other brain areas such as the prefrontal and temporal were unused owing to artifacts arising from sound and facial or eyeball movements [43]. In the case of Dataset I, the combined EEG samples (original EEG + reconstructed EEG) as shown in Fig. 2, we segmented filtered ME-EEG and MI-EEG into the training sets (80% of the entire dataset, with each individual training set corresponding to 40% of the entire dataset) and a test set (20% of the entire MI-EEG dataset). The data set consisting only of MI samples for comparison was used as a training set of 80% and a test set of 20% of the MI-EEG data. In the process of separating the entire data set, the training data set and the test data set were separated and configured not to overlap.

Regarding Dataset II, EEG samples were preprocessed using the same steps as those applied for Dataset I. We used all channels to analyze the dataset [15]. To evaluate the proposed method on Dataset II, we randomly divided the data of ME-EEG and MI-EEG samples into training sets. The combined EEG samples were comprised 80% of ME-EEG and 80% of MI-EEG, for training and 20% of the entire MI-EEG data as a test set.

2) Network Architecture: We designed CVNet as the MI classification model. The proposed method is composed of two main steps: channel-wise VAE for ME sample reconstruction and a deep neural network for MI classification.

VAE aims at learning probability distributions as a generative model and reconstructing data [44], [45]. An autoencoder (AE) [46], which is the background of VAEs, typically performs unsupervised learning and aims at dimension reduction. However, the AE method does not entail the probability distribution based on which the data is generated. In contrast with the VAE method, the AE simply re-creates a vector via dimension reduction of the data; however, the VAE determines the underlying probability distribution of the original data to produce similar data.

Selecting and classifying useful spatial information from MI-based EEG samples still remains a difficult problem. In particular, it is very difficult for a participant to imagine consistently when performing high-level tasks within the same limb. To address this problem, we propose a channel-wise VAE-based deep learning model optimized to obtain spatial features that are important for classification using EEG signals obtained via ME samples. The channel-wise VAE is composed of one input layer, five hidden layers, and one output layer (Fig. 3).

We denote the format of the input data as:

$$X_N = \{(T, C_N)\}, \tag{1}$$

where T is the time of one epoch; C is the channel; and N is the number of channels. Encoder $q_\phi\ (z \mid X_N)$ learns a sampled

latent representation z from the input data X_N, while decoder $p_\theta(X_N \mid z)$ aims to reconstruct the data X_N from the learned representation z (Fig. 3).

The mean and variance of the distribution of latent vector z values are output when an X_N input is given to the encoder. The distribution of features from here follows a Gaussian normal distribution. Therefore, the latent variable values in the latent space represent the standard deviation σ_N and mean μ_N, which approximates the true distribution of $q_\phi(X_N)$.

Next, z is sampled with a reparameterization trick, namely

$$z_N = \mu_N + \sigma_N \odot \varepsilon \tag{2}$$

where $\epsilon \sim N(0, 1)$ and $\odot$ refers to the element-wise product. The value for ϵ is randomly extracted from the normal distribution $N(0, 1)$, multiplied by the standard deviation σ_N, and then added to the mean μ_N to form a value of z as previously defined.

The z value is entered into the input of the decoder, and an output $\widehat{x_N}$ with the original input size is produced. Because the distribution of $p_\theta(X_N)$ is set to Bernoulli, the output has a value between 0 and 1. Therefore, the activation function was set to a sigmoid function.

The loss function l_N in the channel-wise VAE framework, which is known as the evidence lower bound, is given by:

$$l_N(X_N; \phi, \theta) = -\mathbb{E}_{z\sim q_\phi}\left[log p_\phi(x_N \mid z)\right] \tag{3}$$
$$+ D_{KL}\left(q_\phi(z \mid x_N) \,||\, p(z)\right) \tag{4}$$

where the first term (3) is the reconstruction loss of the decoder. In the case of a generative model, it is necessary to consider the relationship between the created output and the MI because new data x has to be created. The decoder follows the Bernoulli distribution, and thus, it finds the cross-entropy between the two. The ME signal was designed to be similar to the MI signal distribution using the reconstruction criterion while passing through the loss function (Fig. 3). The second term (4) is the regularization loss. This is the loss term for the distribution of a function that applies the true distribution to ensure that X_N learns to exhibit the same distribution as the original distribution. In this case, the loss determines the probability distribution distance between the two. The channel-wise VAE is trained through stochastic gradient descent (SGD) and the backpropagation method. Owing to this approach, the ME samples reconstructed based on the MI samples exhibit low noise, and their signal amplitude is rescaled such that it is similar to that of the MI-based EEG samples; more importantly, useful spatial information is obtained through this method. In addition, in the BCI domain, the spatial information of each channel is crucial. Therefore, the architecture was designed to reconstruct each channel at the input and output of the VAE. In the case of evaluation using only MI samples, the loss function was not modified because the conventional methods were used.

A deep neural network consists of three primary types of layers: convolution, pooling, and fully-connected layers. Table I describes the CNN network used in our experiment. In the input layer, the reconstructed ME and MI samples are fed as the input data of the feature extraction step. At this time, the sizes of the two datasets are the same (C, T).

TABLE I
DETAILED SPECIFICATION OF THE CNN ARCHITECTURE COMPRISING CONVOLUTIONAL LAYERS, POOLING LAYERS, AND FULLY CONNECTED LAYERS

Block	Layer	Hyper-parameters	Output shape	Activation
-	Input		$(1, C^{a)}, T^{b)})$	
I	Tem-Conv2D	$K_1^{c)} = 30, F_1^{d)} = (1, 25), P_1^{e)} = (0, 0)$	(K_1, C, T_1)	Linear
	BatchNorm2D	m = 0.99	(K_1, C, T_1)	
II	DW-Conv2D	$K_2 = 30, F_2 = (C, 1), P_2 = (0, 0)$	$(K_2, 1, T_1)$	Linear
	BatchNorm2D	m = 0.99	$(K_2, 1, T_1)$	
	Activation	$\alpha = 1$	$(K_2, 1, T_1)$	ELU
	AvgPool2D	$F_{p1}^{f)} = (1, 4), S_{p1}^{g)} = (1, 4)$	$(K_2, 1, T_{p1})$	
	Dropout	$p = 0.5$	$(K_2, 1, T_{p1})$	
III	Sep-Conv2D	$K_3 = 40, F_3 = (1, 30), P_3 = (0, 0)$	$(K_3, 1, T_{p1})$	Linear
	BatchNorm2D	m = 0.99	$(K_3, 1, T_{p1})$	
	Activation	$\alpha = 1$	$(K_3, 1, T_{p1})$	ELU
	AvgPool2D	$F_{p2} = (1, 3), S_{p2} = (1, 3)$	$(K_3, 1, T_{p2})$	
	Dropout	$p = 0.5$	$(K_3, 1, T_{p2})$	
IV	Sep-Conv2D	$K_4 = 40, F_4 = (1, 30), P_4 = (0, 0)$	$(K_4, 1, T_{p2})$	Linear
	BatchNorm2D	m = 0.99	$(K_4, 1, T_{p2})$	
	Activation	$\alpha = 1$	$(K_4, 1, T_{p2})$	ELU
	AvgPool2D	$F_{p3} = (1, 3), S_{p3} = (1, 3)$	$(K_4, 1, T_{p3})$	
	Dropout	$p = 0.5$	$(K_4, 1, T_{p3})$	
V	AvgPool2D	$F_{p4} = (1, 3), S_{p4} = (1, 3)$	$(K_4, 1, T_{p4})$	
	Dropout	$p = 0.5$	$(K_4, 1, T_{p4})$	
	AvgPool2D	$F_{p5} = (1, 3), S_{p5} = (1, 3)$	$(K_4, 1, T_{p5})$	
	Dropout	$p = 0.5$	$(K_4, 1, T_{p5})$	
VI	Flatten		(K_4, T_{p5})	
	Fully-Connected	$N_c^{h)} = 250$	(N_c)	
	Fully-Connected	$N_c = 4$	(N_c)	
	Activation		(N_c)	Softmax

$^{a)}$The number of electrodes, $^{b)}$Time samples, $^{c)}$The number of convolutional kernels, $^{d)}$Kernel size, $^{e)}$Padding size, $^{f)}$Pool size, $^{g)}$Pool stride, $^{h)}$Number of the classes

The spatio-temporal feature block was a 2D temporal convolutional layer that trained $K = 30$ band-pass temporal filters with a low number of learnable parameters. A 2D spatial depthwise convolutional (DW-Conv2D) layer was then introduced to learn spatial filters of size $(C, 1)$ for each temporal feature map. Each convolutional layer in these blocks is characterized by the kernel size (F), padding size (P), pool size (F_p), and pool stride (S_p). Next, two 2D separable convolutional (Sep-Conv2D) layers were applied to reduce the number of fitting parameters and explicitly decouple the relationship within and across feature maps by first learning a kernel summarizing each feature map individually.

Batch normalization along the feature map dimension was applied to the neurons of each convolutional layer, followed by a non-linear activation function. In this study, exponential linear units (ELUs) were adopted as the activation function, as this non-linearity allows faster and more noise-robust learning than other non-linearities. The formulas are shown in (5) and (6) below.

$$f(x) = x \quad (x > 0) \tag{5}$$
$$f(x) = \alpha \cdot (exp(x) - 1) \quad (x \leq 0) \tag{6}$$

Furthermore, Schirrmeister *et al.* [47] reported better performance for shallow and deep CNNs applied to EEG motor

TABLE II
CLASSIFICATION PERFORMANCE OF PROPOSED AND CONVENTIONAL METHODS IN DATASET I ACROSS ALL PARTICIPANTS

Participants	Original EEG (only MI)				Combined EEG				
	CSP+LDA [48]	DeepConvNet [47]	ShallowConvNet [47]	EEGNet [49]	CSP+LDA [48]	DeepConvNet [47]	ShallowConvNet [47]	EEGNet [49]	CVNet
P01	0.50 (±0.07)	0.60 (±0.01)	0.70 (±0.04)	0.67 (±0.01)	0.60 (±0.04)	0.77 (±0.08)	0.80 (±0.04)	0.76 (±0.07)	0.87 (±0.02)
P02	0.43 (±0.05)	0.67 (±0.07)	0.67 (±0.08)	0.57 (±0.09)	0.55 (±0.09)	0.72 (±0.06)	0.75 (±0.01)	0.60 (±0.03)	0.82 (±0.03)
P03	0.40 (±0.05)	0.65 (±0.03)	0.70 (±0.02)	0.65 (±0.01)	0.57 (±0.01)	0.80 (±0.03)	0.80 (±0.09)	0.75 (±0.02)	0.87 (±0.08)
P04	0.42 (±0.07)	0.57 (±0.05)	0.57 (±0.09)	0.60 (±0.07)	0.53 (±0.02)	0.72 (±0.05)	0.72 (±0.09)	0.70 (±0.04)	0.82 (±0.01)
P05	0.44 (±0.09)	0.62 (±0.01)	0.62 (±0.01)	0.65 (±0.08)	0.55 (±0.01)	0.75 (±0.04)	0.76 (±0.04)	0.75 (±0.09)	0.83 (±0.01)
P06	0.51 (±0.01)	0.60 (±0.06)	0.57 (±0.08)	0.52 (±0.08)	0.60 (±0.01)	0.70 (±0.07)	0.67 (±0.04)	0.70 (±0.01)	0.82 (±0.01)
P07	0.43 (±0.05)	0.55 (±0.02)	0.52 (±0.05)	0.42 (±0.08)	0.52 (±0.08)	0.65 (±0.02)	0.67 (±0.03)	0.65 (±0.09)	0.77 (±0.06)
P08	0.46 (±0.04)	0.77 (±0.06)	0.72 (±0.04)	0.62 (±0.03)	0.58 (±0.05)	0.82 (±0.01)	0.82 (±0.09)	0.65 (±0.09)	0.97 (±0.07)
P09	0.48 (±0.01)	0.60 (±0.06)	0.62 (±0.07)	0.55 (±0.02)	0.55 (±0.05)	0.65 (±0.01)	0.67 (±0.03)	0.70 (±0.05)	0.80 (±0.06)
P10	0.43 (±0.03)	0.57 (±0.07)	0.52 (±0.04)	0.52 (±0.01)	0.57 (±0.01)	0.67 (±0.02)	0.70 (±0.01)	0.65 (±0.01)	0.72 (±0.04)
Average	0.45 (±0.05)	0.62 (±0.04)	0.62 (±0.05)	0.58 (±0.05)	0.56 (±0.04)	0.72 (±0.04)	0.74 (±0.04)	0.69 (±0.05)	**0.83 (±0.04)**

decoding when using ELUs compared with other activation functions. The α parameters are the ELU hyper-parameter that controls the saturation value for negative inputs and $\alpha = 1$ was set for the proposed architecture. We adopted in our CNN block a momentum term of $m = 0.99$ with $\varepsilon = 1e-3$ for numerical stability. Then, an average pooling layer was introduced to reduce the number of trainable parameters in the transitions among DW-Conv2D, Sep-Conv2D, and the subsequent fully-connected layer in the classification block, the convolutional-to-dense connections. A pool size of $F_{p1} = (1, 3)$ or $F_{p2} = (1, 4)$ and pool stride of $S_{p1} = (1, 3)$ or $S_{p2} = (1, 4)$ were used.

In the classification block, two fully-connected layers flattened the features that were extracted through multiple layers. The softmax function and the cross-entropy function were applied as the activation and loss functions, respectively. An overall number of training epochs for learning was set to 400 epochs as depicted in Fig. 7. The proposed CVNet was trained according to each participant. This training approach is one of various model generation strategies leading to improvement in BCI performance due to differences in EEG signals between individuals [6]. The data dimension of both datasets was composed of three dimensions (channel×time×trial), and the experiment was conducted at a sampling rate of 250 Hz. Our models were implemented with an NVIDIA Titan×GPU 12 GB RAM and used PyTorch. The overall training procedure of CVNet is summarized in Algorithm 1.

Algorithm 1 Training Procedure of CVNet

Input: A set of training MI-EEG and ME-EEG samples
- $X = \{x_i\}_{i=1}^{D}$, $\{x_i\} \in \mathbb{R}^{C\times T}$ from ME training samples, where D is total number of trials with C channels and T sample points
- X_{tr}: a training set of ME-EEG data
- $Y = \{y_i\}_{i=1}^{D}$, $\{y_i\} \in \mathbb{R}^{C\times T}$ from MI training samples
- Y_{tr}: a training set of MI-EEG data
- $\Omega = \{O_i\}_{i=1}^{D}$: class labels from MI dataset, where $O_i \in \{0, 1, 2, 3\}$
- Ω_{tr}: a test set of MI-EEG data

Output: Trained CVNet model

Stage 1: Train channel-wise VAE
- Input X_{tr}, $X_N = \{(T, C_N)\}$, where N is the number of channels
- Find the latent vector z according to (2)
- Extract features X_N and reconstruct to $\widehat{X_N}$ using $Y_N = \{(T, C_N)\}$
- Generate loss value by calculating differences between $\widehat{X_N}$ and Y_N using (3) for generative model and regularization loss for probability distance

Stage 2: Train deep neural network
- $\widehat{X_N}$ and Y_N inputs for deep neural network
- Initialize the parameters of deep neural network to random values
- Generate loss value by calculating differences between deep neural network output and class label Ω_{tr}

Stage 3: Fine-tune parameters
- Minimize loss values by tuning parameters of both Stage 1 and Stage 2

IV. EXPERIMENTAL RESULTS

A. Comparison of the Classification Performances Using the Proposed With Baseline Models

To evaluate the classification performance, we used the representative baseline models in the MI-BCI studies as below; CSP+LDA [48], DeepConvNet [47], ShallowConvNet [47], and EEGNet [49]. Table II displays the classification performance compared with that of conventional methods for Dataset I. In the combined EEG, there were significant differences between the comparison models and the proposed model. CVNet had the highest total average classification performance of 0.83 (±0.04), and the CSP+LDA method had the lowest accuracy of 0.56 (±0.04). The proposed method, for Dataset I, had a performance difference of 0.27 compared to CSP+LDA, and 0.09 compared to ShallowConvNet. For CVNet, the best classification performance, 0.97 (±0.07), was obtained for participant P08. In addition, there was a performance difference of approximately 0.25 compared with the result obtained for participant P10. We also compared the performances obtained using only the MI samples and the combined EEG samples. When comparing the averaged performance, differences of 0.11, 0.10, 0.12, and 0.11 were found for CSP+LDA, DeepConvNet, shallowConvNet, and EEGNet,

TABLE III
CLASSIFICATION PERFORMANCE OF PROPOSED AND CONVENTIONAL METHODS IN DATASET II ACROSS ALL PARTICIPANTS

Participants	Original EEG (only MI)				Combined EEG				
	CSP+LDA [48]	DeepConvNet [47]	ShallowConvNet [47]	EEGNet [49]	CSP+LDA [48]	DeepConvNet [47]	ShallowConvNet [47]	EEGNet [49]	CVNet
P01	0.22 (±0.05)	0.37 (±0.09)	0.37 (±0.06)	0.33 (±0.02)	0.52 (±0.03)	0.66 (±0.05)	0.67 (±0.01)	0.62 (±0.09)	0.70 (±0.01)
P02	0.30 (±0.02)	0.33 (±0.07)	0.35 (±0.03)	0.41 (±0.03)	0.50 (±0.09)	0.54 (±0.02)	0.58 (±0.02)	0.56 (±0.05)	0.63 (±0.08)
P03	0.28 (±0.07)	0.39 (±0.04)	0.43 (±0.08)	0.29 (±0.04)	0.42 (±0.04)	0.62 (±0.06)	0.65 (±0.08)	0.52 (±0.05)	0.72 (±0.09)
P04	0.27 (±0.01)	0.35 (±0.04)	0.37 (±0.05)	0.33 (±0.02)	0.44 (±0.01)	0.60 (±0.07)	0.52 (±0.02)	0.54 (±0.02)	0.69 (±0.07)
P05	0.30 (±0.06)	0.35 (±0.04)	0.35 (±0.03)	0.35 (±0.08)	0.43 (±0.09)	0.66 (±0.02)	0.54 (±0.09)	0.52 (±0.04)	0.73 (±0.01)
P06	0.29 (±0.01)	0.37 (±0.03)	0.27 (±0.09)	0.27 (±0.01)	0.44 (±0.09)	0.67 (±0.01)	0.69 (±0.07)	0.60 (±0.06)	0.72 (±0.02)
P07	0.28 (±0.03)	0.31 (±0.05)	0.33 (±0.08)	0.39 (±0.02)	0.47 (±0.04)	0.56 (±0.02)	0.54 (±0.04)	0.50 (±0.06)	0.66 (±0.03)
P08	0.27 (±0.06)	0.29 (±0.05)	0.35 (±0.05)	0.27 (±0.01)	0.43 (±0.01)	0.62 (±0.03)	0.67 (±0.05)	0.58 (±0.03)	0.74 (±0.06)
P09	0.27 (±0.07)	0.33 (±0.08)	0.29 (±0.06)	0.27 (±0.02)	0.42 (±0.02)	0.54 (±0.04)	0.54 (±0.02)	0.52 (±0.03)	0.63 (±0.01)
P10	0.27 (±0.01)	0.37 (±0.07)	0.41 (±0.05)	0.27 (±0.04)	0.46 (±0.04)	0.60 (±0.05)	0.62 (±0.04)	0.56 (±0.09)	0.68 (±0.07)
Average	0.27 (±0.04)	0.34 (±0.05)	0.35 (±0.06)	0.32 (±0.03)	0.45 (±0.04)	0.61 (±0.04)	0.60 (±0.04)	0.55 (±0.05)	**0.69 (±0.04)**

TABLE IV
STATISTICAL ANALYSIS OF CLASSIFICATION ACCURACY DIFFERENCES BETWEEN THE COMPARED METHODS USING A DIFFERENT NUMBER OF TRAINING SAMPLES FROM EACH PARTICIPANT FOR DATASET I AND DATASET II

Comparison groups	Dataset I				Dataset II			
	The number of training samples				The number of training samples			
	40	80	120	160	48	96	144	192
CVNet vs. CSP+LDA [48]	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$
CVNet vs. DeepConvNet [47]	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$
CVNet vs. ShallowConvNet [47]	$p < 0.001$	$p < 0.002$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$
CVNet vs. EEGNet [49]	$p < 0.001$	$p < 0.001$	$p < 0.003$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$

respectively. For the MI samples, there was a performance difference of approximately 0.17 between the CSP+LDA and ShallowConvNet methods. In addition, the CSP+LDA method showed a performance difference of 0.07 compared with EEGNet.

Table III displays the performance comparison results for Dataset II. The models adopted for comparison were identical to those adopted in the case of Dataset I. The proposed method shows the highest grand-averaged classification performance, at 0.69 (±0.04), and the lowest result is obtained for CSP+LDA, at 0.45 (±0.04). In this experiment, the model proposed herein exhibits a performance difference of at least 0.08 and at most 0.24 compared with DeepConvNet and CSP+LDA, respectively. In the proposed model, the best classification accuracy, 0.74 (±0.06), was achieved for participant P08. In addition, there was a performance difference of approximately 0.11 compared with the results obtained for participants P02 and P08. Compared with the existing methods, the maximum performance difference for the proposed method was observed to be 0.30 (participant P03) and 0.31 (participant P08). Further, we compared the performances obtained for the MI and combined EEG samples. When comparing the averaged performance, differences of 0.18, 0.27, 0.25, and 0.23 were found for CSP+LDA, DeepConvNet, ShallowConvNet, and EEGNet, respectively. For DeepConvNet, a performance difference of approximately 0.33 was obtained for participant P08. This result shows that CVNet demonstrated the highest classification performance for various upper extremity movements when compared to the conventional models.

When the MI samples were employed, CVNet exhibited a significant difference compared with other methods (Table IV). Before the statistical analysis, the normality and homoskedasticity tests were conducted owing to small population samples. We applied the Shapiro–Wilk test, which is normally applied to both our dataset (Dataset I) and the public dataset (Dataset II) as a normality validation, separately. We confirmed that the null hypothesis (H0) satisfies the normality. Thereafter, the assumption of homoskedasticity was also satisfied using Levene's test for each comparative group (e.g., CVNet vs. CSP+LDA in Table IV). Hence, the paired *t*-test was used to investigate the statistical significance of the difference in accuracy between CVNet and other comparative groups according to the number of training samples in both datasets. The results showed that CVNet exhibited significantly higher performance than the other methods with respect to the number of training samples in all cases. In particular, a remarkable statistical significance was observed for the case of the smallest number of training samples (40 and 48). Furthermore, CVNet performed significantly better than conventional methods on both datasets.

B. Classification Performances According to the Number of Training Samples

The overall classification accuracy with regard to the number of training samples in Dataset I is shown in Fig. 4. In the case of combined EEG samples, we trained the model with 160, 120, 80, and 40 training samples (80% of the total number of samples), and the number of test samples was 40, 30, 20, and 10 respectively (20% of the total number of

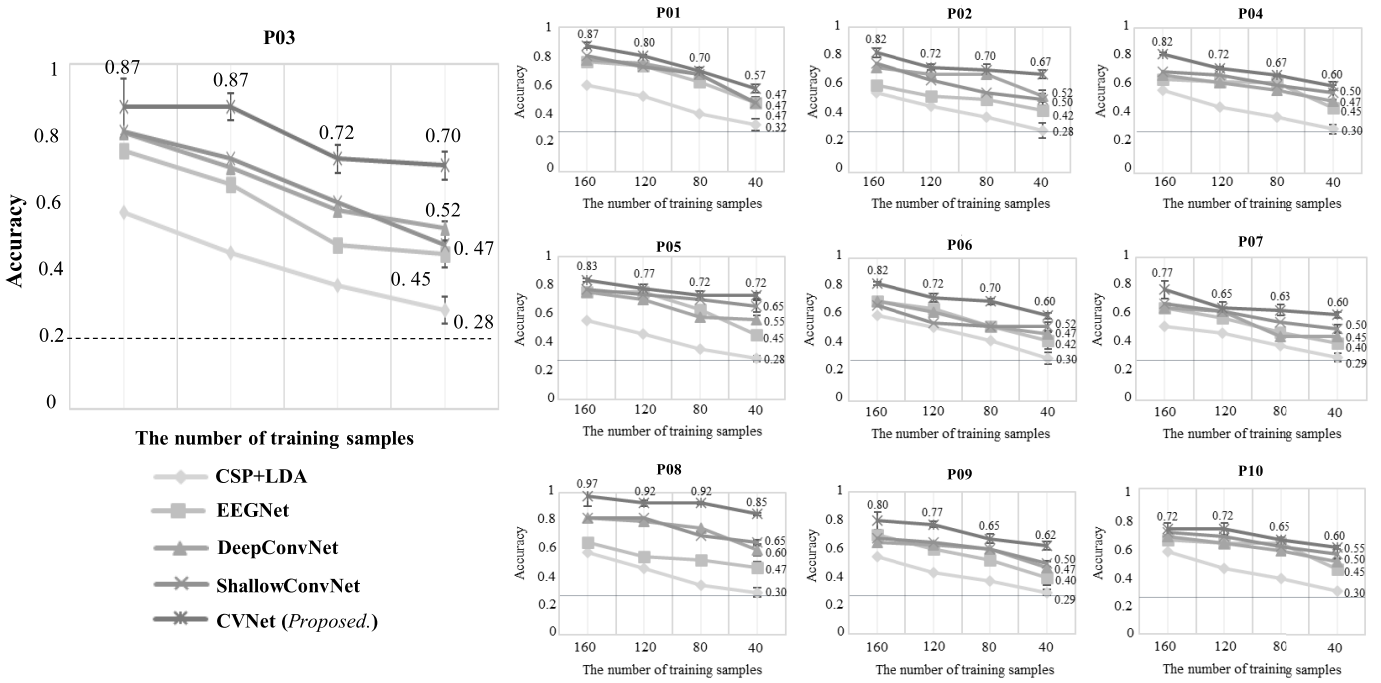

Fig. 4. Classification performances according to the number of training samples for all participants for Dataset I. Representative participant P03 is shown at the top. The black dashed line indicates the chance-level accuracy for Dataset I.

samples). The training sample and test sample were randomly selected and constructed so that there was no overlap in the entire data set. Even when the number of training samples decreased, we found that the proposed method had the highest accuracy compared to the other methods. When 10 samples per class were used, the classification performance decreased to a minimum of 0.25 (participant P01) and a maximum of 0.55 (participant P08). For participant P01, the performance difference was 0.30 when comparing the cases of 160 and 40 training samples. In addition, there was a minimum difference of 0.11 for P05.

Fig. 5 shows the average MI classification performances obtained by CSP+LDA, EEGNet, DeepConvNet, ShallowConvNet, and CVNet, respectively, for Dataset II. We trained the models with 192, 144, 96, and 48 training samples, respectively (80% of the total number of samples). In addition, the corresponding number of test samples was 48, 36, 24, and 12 (20% of the total number of samples). The proposed method is designed to operate and learn channel-wise VAE and CNN at the same time. Similarly, when the number of training samples decreased, both the channel-wise VAE and CNN were trained with the reduced training set. The results show that the proposed method outperformed the other methods even if only 48∼144 samples from all participants were available for training. For the CSP+LDA method, when the number of training samples was 48, it was 0.24 below the chance level (0.25). On the other hand, CVNet showed the highest accuracy compared to the other methods, with a result of 0.45. There was a performance difference of approximately 0.24 when comparing the cases of 48 and 192 training samples for the proposed model. In the case of 48 training samples, the proposed model had a performance difference of at least

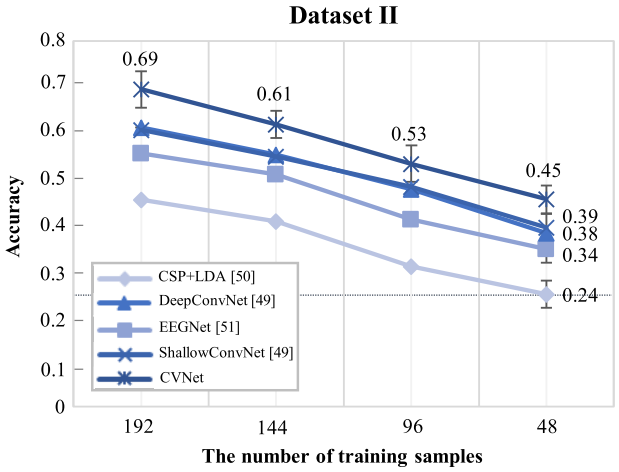

Fig. 5. Averaged MI classification accuracies obtained by CSP+LDA, EEGNet, DeepConvNet, ShallowConvNet, and CVNet with varying numbers of training samples for Dataset II. The black dashed line indicates the chance-level accuracy.

0.06 and at most 0.21 compared with ShallowConvNet and CSP+LDA. This result confirms the potential of the CVNet method for application MI classification even if the number of training samples is reduced.

C. Performance Measurement Using Confusion Matrices

Fig. 6 shows the confusion matrix for each class with the proposed model for Dataset I and Dataset II. Each confusion matrix represents the true label in the column and the predicted label in the rows. In Dataset I, the averaged accuracy has a classification accuracy of 0.83. The true positive value is

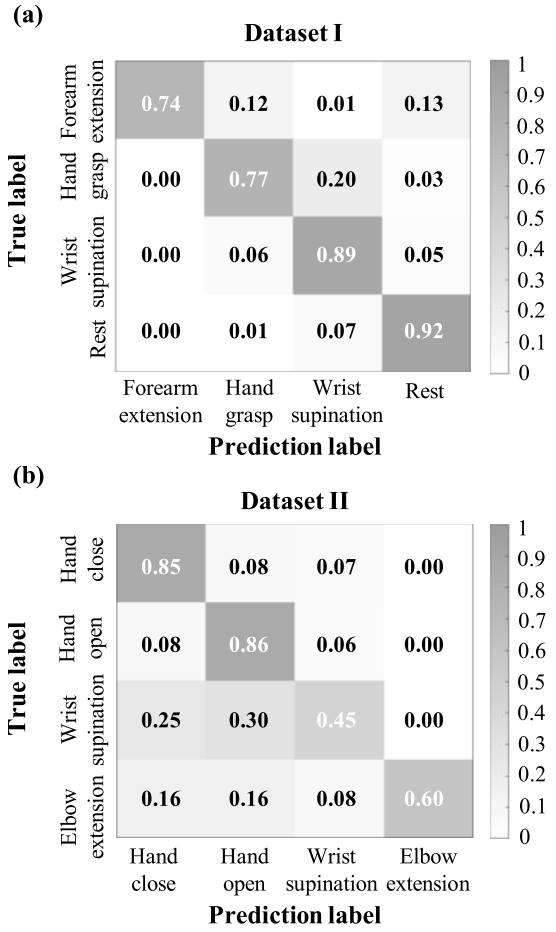

Fig. 6. Confusion matrices of each class across subject according to Dataset I and Dataset II.

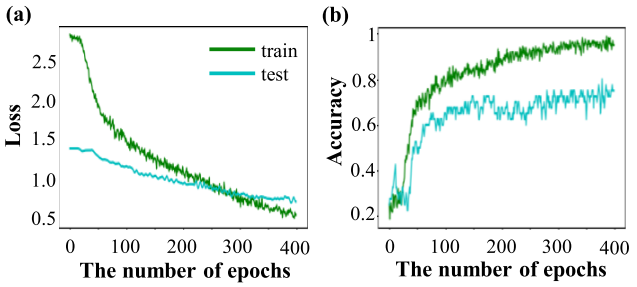

Fig. 7. Convergence curve of the training error and classification accuracy. (a) Training loss. The errors were calculated to extract the optimized parameters during network training. (b) Classification performance according to the number of epochs.

0.74 for the 'Forearm extension' class and 0.92 for the 'Rest' class. We also observed the main confusion between the 'Hand grasp' and 'Wrist supination' classes. However, for the 'Wrist supination' class, the true positive value was 0.89, which was higher than that of the other classes. The results show that there is a difference in performance for each class, but it is confirmed that there is no significant difference from the overall performance. Further, for Dataset II, the average classification accuracy was 0.69. As shown in the confusion matrix, the true positive of the 'Hand open' class had the highest value among all tasks (0.86). However, the proposed model confused the 'Wrist supination' class with the 'Hand open' class. The true-positive value for 'Wrist supination' class was 0.45. Thus, these results demonstrate that it is possible to decode various upper limb movements corresponding to the same limb, which were included in Dataset I and Dataset II.

D. Convergence Curve of Training Error

We experimented with reducing the number of training samples for practical and real-time BCI applications. This allows CVNet to see the potential for decoding the user's intent, even when the number of training samples is small. To confirm the convergence process of CVNet, the error change curve and accuracy according to the epoch are shown in Fig. 7. The proposed model learned the information regarding the participants' upper limb movements from the brain signal data when training the CVNet model with 160 samples in Dataset I. Most of the participants in this study showed convergence within 300–400 epochs.

V. Discussion

Recent research in the field of BCI has suggested various classification methods using advanced machine learning methods. However, only a few studies have classified the user's intention to perform upper limb movement tasks corresponding to only a single limb. In this study, we proposed CVNet, a deep learning framework based on the VAE method based on inter-task transfer learning to decode intuitive user intentions. In the channel-wise VAE method located in the front, it is designed to analyze each channel by inserting each channel as an input rather than inserting all channels at once. Therefore, it is a useful structure for extracting spatial information about each channel. The proposed method could contribute to the novel BCI advances with the following strength. First, it can compensate for the spatial features of the MI samples. Second, the classification performance is maintained to some extent even when the number of training samples decreases. As a result, it provides the possibility of contributing to practical BCI-based controlling external devices in real-time with less calibration time. In addition, to the best of our knowledge, this is the first attempt to reconstruct ME samples using the VAE method for MI classification.

In addition, Fig. 8 shows the representation of spatial distributions per task, and compares the neurophysiological patterns between original EEG data and reconstructed EEG data. Fig. 8(a) depicted the distribution in terms of the anatomical model using brain signal analysis toolbox: Brainstorm (https://neuroimage.usc.edu/brainstorm/) for Dataset I. To demonstrate the neurophysiological rationale, the spatial distributions for ME tasks and MI tasks are shown. This supported the proper adaptation of the concept of inter-task transfer learning in this study. The spatial distribution corresponding to supination was projected onto a rendered brain to visualize the spatial distribution in the cortex. The spatial source-based feature maps of 'Forearm extension', 'Hand grasp', and 'Wrist supination', which are the MI tasks for

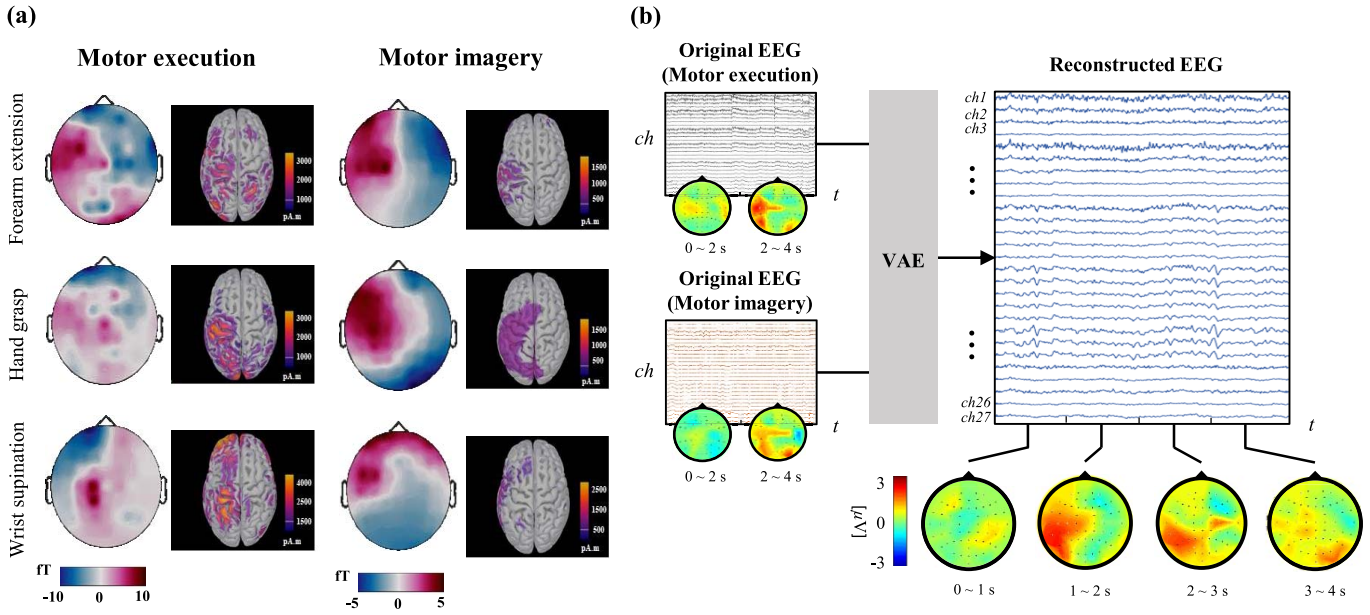

Fig. 8. Neurophysiological analysis using spatial distribution. (a) Representation of spatial distributions per each motor execution task and motor imagery task. The overall brain activation changes are shown nearby the motor cortex in the brain. (b) Comparison of signal patterns before and after VAE. Reconstructed EEG means the output of VAE and showed similar patterns with original EEG signals.

representative participant P03, are shown using 60 electrodes. The difference between ME and MI from all participants showed similar spatial patterns, we presented the representative patterns in participant P03 as Fig. 8(a). The bright color of the spatial patterns indicates activated regions and the dark color indicates inactivated regions. This means that cortical activation was observed in both the ME and MI sessions. The center of each ME and MI task distribution is located at or near the motor cortex area. EEG electrodes in the MI session showed smaller changes compared to EEG electrodes in the ME session. Compared to each of the MI tasks, the ME tasks showed significantly activated brain regions. In the hand grasp task, the left hemisphere was activated in both the ME and MI sessions. In Fig. 8(b), we compared the brain signal patterns between original EEG and reconstructed EEG through the VAE for representative participant P03 in the 'Hand grasp' tasks, representatively. The output of VAE, the reconstructed EEG, generated similar patterns between both original EEG (ME and MI) signals. The training targets (ME-EEG and MI-EEG) reconstructed the amplitude value according to the channel within a time. Especially, when observing the spatial patterns using a scalp topographic map only with the reconstructed EEG value when there was a movement of the right upper extremity, it was confirmed that the left hemisphere was activated during the 1∼3 s. It has been proved that is generated similar to the existing spatial pattern so that the proposed CVNet was well trained with effects of data augmentation. However, in order for the EEG reconstruction of the proposed model to fully evaluate its usefulness in a variety of various datasets, the advanced VAE in the training stage seems to be needed through additional quantitative comparisons between the reconstructed EEG and the original EEG. We are currently planning to solve the limitations of this part through combination with the advanced GAN model.

Recently, generative model-based approaches have been applied in the EEG-based BCI domain. Bi *et al.* [50] proposed a semi-supervised algorithm. They used ICA and a Kalman smoother to improve the SNR. Subsequently, VAE was used to provide robust features of EEG signals. Dai *et al.* [51] decoded MI using a CNN architecture with a VAE for classification. In this network, the classification of the extracted CNN features is performed via a deep network VAE. In our method, the VAE is used to reconstruct ME-EEG signals channel by channel. These signals are similar to MI-based EEG signals of characteristics and they also reduce noise arising from movements and the scale of the amplitude. These deep learning methodologies in the BCI domain have contributed to the EEG decoding of various paradigms. In this study, we found that deep learning techniques can be used to decode the MI of various intuitive tasks corresponding to a single arm. Also, we reconstructed the ME samples to MI samples using the channel-wise VAE method. Using this method, spatial information can be obtained and brain signals can be refined before EEG data enters the deep learning architecture.

To confirm the importance of the training data samples, we identified the difference in performance between the MI samples and the combined EEG samples. These results show that in the case of using only the MI paradigm, the classification performance is still degraded because participants imagine moving their muscles and quickly feel tired. Therefore, in this study, the ME paradigm was also employed during training, as shown in Fig. 1, to improve the MI decoding performance.

In signal processing, the technology in the field of speech has substantially advanced, so even if a user's speech/voice

signals are calibrated once, they can be used in a real-time application. Recently, BCI advances have been achieved, but the performance of the developed techniques is still degraded and technical limitations exist. To solve this problem, we conducted a performance test by reducing the number of training samples and proposed a model that can contribute to overcoming existing limitations. However, the performance still tends to be insufficient with regard to scaling many classes. To solve this problem, we plan to modify the model with a real-time closed-loop deep learning architecture with an adaptive structure. In addition, in fact, we selected the EEG channels differentially for each dataset in the pre-processing step. In Dataset I, which was collected in our laboratory environment, we chose the 27 channels that are the motor region part was clearly activated. On the other hand, similarly, Dataset II has tried to maintain the same analysis in this study because the previous studies have been performed analysis using the entire channels. Therefore, the proposed method remains several limitation issues such as the generalizability for a variety of datasets.

Currently, the MI-BCI places a high cognitive load on the participant due to the long calibration time. In general, collecting the EEG datasets and training the deep learning model takes approximately 2∼3 hours. Long training times are a significant part of the challenges facing the BCI domain [52]. Thus, several studies have attempted to use only a small amount of data or used a dataset from another participant using the transfer learning method. To solve these BCI problems, we investigated a deep learning method that applies inter-task transfer learning [53], [54] using only a small number of training samples. Transfer learning methods are being developed as one of the approaches for reducing the training time for practical BCI systems [54]. A task-to-task BCI system could reduce the number of required training samples, thereby reducing participant fatigue. The ultimate goal of this study is to develop a practical BCI system by increasing the classification performance for intuitive MI tasks. Therefore, we adopted an algorithm based on the method of transfer learning between tasks [53]–[55] and plan to apply advanced algorithms based on state-of-the-art methods.

VI. Conclusion and Future Works

In this work, we decoded various upper limb movements based on brain signals via the proposed VAE deep learning framework. The proposed method was able to classify intuitive MI tasks on both our dataset (forearm extension, hand grasp, wrist supination, and rest) and a public dataset (elbow extension, wrist supination, hand open, and hand close). The CVNet method was designed to reconstruct the ME samples in a channel by channel considering the MI samples. The experimental performances showed that CVNet could reconstruct the EEG signals for the ME samples. In addition, the decoding performance was significantly improved compared to that of other methods. Therefore, the CVNet model can support the development of BCIs which may enable control based on imagined movement. Furthermore, we demonstrated the feasibility of using ME samples to improve the MI classification performance. This possibility has the potential to improve classification performance and reduce calibration time in the intuitive BCI paradigm.

In future work, we will reduce the portion of MI session samples so that the calibration time is reduced in a real-time BCI system. We will also modify the CVNet framework to enable real-time robotic arm control by improving the MI classification performance. We believe that the proposed model will contribute to the development of intuitive BCI system technology and will support the daily lives of healthy people and of people with movement disorders.

Acknowledgment

The authors thank Prof. C. Guan for the useful discussion of the experiment, and J.-H. Cho and B.-H. Kwon for their help with dataset construction.

References

[1] J. R. Millan and J. Mourino, "Asynchronous BCI and local neural classifiers: An overview of the adaptive brain interface project," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 2, pp. 159–161, Jun. 2003.

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

[3] S. R. Soekadar *et al.*, "Hybrid EEG/EOG-based brain/neural hand exoskeleton restores fully independent daily living activities after quadriplegia," *Sci. Robot.*, vol. 1, no. 1, pp. 1–9, Dec. 2016.

[4] 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.

[5] C. I. Penaloza and S. Nishio, "BMI control of a third arm for multitasking," *Sci. Robot.*, vol. 3, no. 20, pp. 1–7, Jul. 2018.

[6] 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.

[7] R. Mane, T. Chouhan, and C. Guan, "BCI for stroke rehabilitation: Motor and beyond," *J. Neural Eng.*, vol. 17, no. 4, Aug. 2020, Art. no. 041001.

[8] A. Ravi, N. H. Beni, J. Manuel, and N. Jiang, "Comparing user-dependent and user-independent training of CNN for SSVEP BCI," *J. Neural Eng.*, vol. 17, no. 2, Apr. 2020, Art. no. 026028.

[9] R. Chavarriaga and J. D. R. Millán, "Learning from EEG error-related potentials in noninvasive brain–computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 18, no. 4, pp. 381–388, Aug. 2010.

[10] T. Liu, G. Huang, N. Jiang, L. Yao, and Z. Zhang, "Reduce brain computer interface inefficiency by combining sensory motor rhythm and movement-related cortical potential features," *J. Neural Eng.*, vol. 17, no. 3, Jun. 2020, Art. no. 035003.

[11] 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.

[12] S. Aliakbaryhosseinabadi, E. N. Kamavuako, N. Jiang, D. Farina, and N. Mrachacz-Kersting, "Classification of EEG signals to identify variations in attention during motor task execution," *J. Neurosci. Methods*, vol. 284, pp. 27–34, Jun. 2017.

[13] 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.

[14] L. Yao, N. Mrachacz-Kersting, X. Sheng, X. Zhu, D. Farina, and N. Jiang, "A multi-class BCI based on somatosensory imagery," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 8, pp. 1508–1515, Aug. 2018.

[15] 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, Aug. 2017, Art. no. e0182578.

[16] 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," *Front. Neurosci.*, vol. 12, p. 130, Mar. 2018.

[17] K. J. Miller, G. Schalk, E. E. Fetz, M. den Nijs, J. G. Ojemann, and R. P. N. Rao, "Cortical activity during motor execution, motor imagery, and imagery-based online feedback," *Proc. Nat. Acad. Sci. USA*, vol. 107, pp. 4430–4435, Mar. 2010.

[18] M. Tariq, L. Uhlenberg, P. Trivailo, K. S. Munir, and M. Simic, "Mu-beta rhythm ERD/ERS quantification for foot motor execution and imagery tasks in BCI applications," in *Proc. 8th IEEE Int. Conf. Cognit. Infocommun. (CogInfoCom)*, Sep. 2017, pp. 000091–000096.

[19] 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.

[20] A. M. Azab, L. Mihaylova, K. K. Ang, and M. Arvaneh, "Weighted transfer learning for improving motor imagery-based brain–computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 7, pp. 1352–1359, Jul. 2019.

[21] 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.

[22] V. K. Benzy, A. P. Vinod, R. Subasree, S. Alladi, and K. Raghavendra, "Motor imagery hand movement direction decoding using brain computer interface to aid stroke recovery and rehabilitation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 12, pp. 3051–3062, Dec. 2020.

[23] J.-S. Bang, M.-H. Lee, S. Fazli, C. Guan, and S.-W. Lee, "Spatio-spectral feature representation for motor imagery classification using convolutional neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jan. 15, 2021, doi: 10.1109/TNNLS.2020.3048385.

[24] 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.

[25] X. Wu, B. Zhou, Z. Lv, and C. Zhang, "To explore the potentials of independent component analysis in brain–computer interface of motor imagery," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 3, pp. 775–787, Mar. 2020.

[26] D. Borra, S. Fantozzi, and E. Magosso, "Interpretable and lightweight convolutional neural network for EEG decoding: Application to movement execution and imagination," *Neural Netw.*, vol. 129, pp. 55–74, Sep. 2020.

[27] 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.

[28] Y. Huang, D. Erdogmus, M. Pavel, S. Mathan, and K. E. Hild, "A framework for rapid visual image search using single-trial brain evoked responses," *Neurocomputing*, vol. 74, no. 12, pp. 2041–2051, Jun. 2011.

[29] J. Jin *et al.*, "The study of generic model set for reducing calibration time in P300-based brain–computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 1, pp. 3–12, Jan. 2020.

[30] W. Wei, S. Qiu, X. Ma, D. Li, B. Wang, and H. He, "Reducing calibration efforts in RSVP tasks with multi-source adversarial domain adaptation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 11, pp. 2344–2355, Nov. 2020.

[31] H. Lu, H.-W. Eng, C. Guan, K. N. Plataniotis, and A. N. Venetsanopoulos, "Regularized common spatial pattern with aggregation for EEG classification in small-sample setting," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 12, pp. 2936–2946, Dec. 2010.

[32] W. Tu and S. Sun, "A subject transfer framework for EEG classification," *Neurocomputing*, vol. 82, pp. 109–116, Apr. 2012.

[33] Y. Liang and Y. Ma, "Calibrating EEG features in motor imagery classification tasks with a small amount of current data using multisource fusion transfer learning," *Biomed. Signal Process. Control*, vol. 62, Sep. 2020, Art. no. 102101.

[34] T. Jia, K. Liu, C. Qian, C. Li, and L. Ji, "Denoising algorithm for event-related desynchronization-based motor intention recognition in robot-assisted stroke rehabilitation training with brain-machine interaction," *J. Neurosci. Methods*, vol. 346, Dec. 2020, Art. no. 108909.

[35] K. Gregor Hartmann, R. Tibor Schirrmeister, and T. Ball, "EEG-GAN: Generative adversarial networks for electroencephalograhic (EEG) brain signals," 2018, *arXiv:1806.01875*.

[36] Y. Jiao, Y. Deng, Y. Luo, and B.-L. Lu, "Driver sleepiness detection from EEG and EOG signals using GAN and LSTM networks," *Neurocomputing*, vol. 408, pp. 100–111, Sep. 2020.

[37] Y. Luo, L.-Z. Zhu, Z.-Y. Wan, and B.-L. Lu, "Data augmentation for enhancing EEG-based emotion recognition with deep generative models," *J. Neural Eng.*, vol. 17, no. 5, Oct. 2020, Art. no. 056021.

[38] B.-H. Lee, J.-H. Jeong, and S.-W. Lee, "SessionNet: Feature similarity-based weighted ensemble learning for motor imagery classification," *IEEE Access*, vol. 8, pp. 134524–134535, 2020.

[39] 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.

[40] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Comput.*, vol. 7, no. 6, pp. 1129–1159, Nov. 1995.

[41] A. Kachenoura, L. Albera, L. Senhadji, and P. Comon, "ICA: A potential tool for BCI systems," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 57–68, Jan. 2008.

[42] M. Z. Baig, N. Aslam, and H. P. H. Shum, "Filtering techniques for channel selection in motor imagery EEG applications: A survey," *Artif. Intell. Rev.*, vol. 53, no. 2, pp. 1207–1232, Feb. 2020.

[43] V. Romei, J. Gross, and G. Thut, "Sounds reset rhythms of visual cortex and corresponding human visual perception," *Current Biol.*, vol. 22, no. 9, pp. 807–813, May 2012.

[44] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," 2013, *arXiv:1312.6114*.

[45] C. Doersch, "Tutorial on variational autoencoders," 2016, *arXiv:1606.05908*.

[46] P. Baldi, "Autoencoders, unsupervised learning, and deep architectures," in *Proc. ICML Conf. Unsupervised Transf. Learn.*, 2012, pp. 37–49.

[47] 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, 2017.

[48] 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. IEEE Symp. Comput. Intell., Cognit. Algorithms, Mind, Brain (SSCI-CCMB)*, Apr. 2013, pp. 146–151.

[49] 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.

[50] L. Bi, J. Zhang, and J. Lian, "EEG-based adaptive driver-vehicle interface using variational autoencoder and PI-TSVM," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 10, pp. 2025–2033, Oct. 2019.

[51] M. Dai, D. Zheng, R. Na, S. Wang, and S. Zhang, "EEG classification of motor imagery using a novel deep learning framework," *Sensors*, vol. 19, no. 3, p. 551, 2019.

[52] 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, pp. 1–15, Oct. 2020.

[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] 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.

[55] X. Zhao, J. Zhao, W. Cai, and S. Wu, "Transferring common spatial filters with semi-supervised learning for zero-training motor imagery brain–computer interface," *IEEE Access*, vol. 7, pp. 58120–58130, 2019.