Deep-Learning-Based Automatic Selection of Fewest Channels for Brain–Machine Interfaces

Hyun-Seok Kim, Min-Hee Ahn, *Student Member, IEEE*, and Byoung-Kyong Min, *Member, IEEE*

***Abstract*—Due to the development of convenient brain–machine interfaces (BMIs), the automatic selection of a minimum channel (electrode) set has attracted increasing interest because the decrease in the number of channels increases the efficiency of BMIs. This study proposes a deep-learning-based technique to automatically search for the minimum number of channels applicable to general BMI paradigms using a compact convolutional neural network for electroencephalography (EEG)-based BMIs. For verification, three types of BMI paradigms are assessed: 1) the typical P300 auditory oddball; 2) the new top-down steady-state visually evoked potential; and 3) the endogenous motor imagery. We observe that the optimized minimal EEG-channel sets are automatically selected in all three cases. Their decoding accuracies using the minimal channels are statistically equivalent to (or even higher than) those based on all channels. The brain areas of the selected channel set are neurophysiologically interpretable for all of these cognitive task paradigms. This study shows that the minimal EEG channel set can be automatically selected, irrespective of the types of BMI paradigms or EEG input features using a deep-learning approach, which also contributes to their portability.**

***Index Terms*—Automation, brain–machine interface (BMI), cognitive system, deep learning, electroencephalography (EEG).**

Manuscript received 17 August 2020; revised 18 November 2020 and 3 January 2021; accepted 15 January 2021. Date of publication 26 February 2021; date of current version 18 August 2022. The work of Min-Hee Ahn was supported by the Basic Science Research Program under Grant 2019R1I1A1A01061545. The work of Byoung-Kyong Min was supported in part by the Basic Science Research Program under Grant 2018R1A2B6004084; in part by the Convergent Technology Research and Development Program for Human Augmentation under Grant 2020M3C1B8081319; in part by the Institute of Information & Communications Technology Planning & Evaluation (IITP) under Grant 2017-0-00451; and in part by the Information Technology Research Center (ITRC) Support Program under Grant IITP-2020-2016-0-00464, which are funded by the Korean government (MSICT) through the National Research Foundation of Korea. This article was recommended by Associate Editor H. A. Abbass. *(Hyun-Seok Kim and Min-Hee Ahn contributed equally to this work.) (Corresponding author: Byoung-Kyong Min.)*

Hyun-Seok Kim is with the Institute for Brain and Cognitive Engineering, Korea University, Seoul 02841, South Korea (e-mail: khs0330kr@korea.ac.kr).

Min-Hee Ahn is with the Institute for Brain and Cognitive Engineering, Korea University, Seoul 02841, South Korea, and also with the Laboratory of Brain and Cognitive Science for Convergence Medicine, College of Medicine, Hallym University, Anyang 14068, South Korea (e-mail: tigoum@korea.ac.kr).

Byoung-Kyong Min is with the Department of Brain and Cognitive Engineering, Korea University, Seoul 02841, South Korea (e-mail: min_bk@korea.ac.kr).

This article has supplementary material provided by the authors and color versions of one or more figures available at https://doi.org/10.1109/TCYB.2021.3052813.

Digital Object Identifier 10.1109/TCYB.2021.3052813

I. Introduction

BRAIN–MACHINE interfaces (BMIs) comprise a promising interfacing technology that can be used to directly control devices by human neural activity (i.e., thought) [1]–[3]. For example, a cognitive BMI can read human intentions from user electroencephalogram (EEG) signals reflecting higher order cognitive processes even when the physical behavior of the user is unknown [4]. EEG-based BMI is the most practical noninvasive BMI technique [5]; however, several challenges must be overcome to improve its performance. A key challenge involves the automatic minimization of the number of channels (i.e., electrodes placed on the scalp) required for BMI control. This is also crucial in terms of the portability of BMIs. Previous studies have attempted to propose channel-selection algorithms selectively for a specific BMI paradigm [6]–[8]. Furthermore, depending on the type of BMI task that induces characterized brain signals, the brain reacts differently among individuals as well as in a single person. Moreover, the number of optimal EEG channels is subject to change, depending on the characteristics of BMI paradigms and the individual variances in brain activity. This further complicates the challenge. In this regard, the development of a new potent technique to automatically select an optimally minimal EEG channel set is required.

Various algorithms for channel selection have been proposed to improve BMI performance [6], [9], [16]. For example, a common spatial pattern (CSP)-based EEG channel-selection algorithm was suggested in a few studies [6], [11], [12]. The algorithm accounts for the high value of CSP filter coefficients and selects the optimal channels based on filter weights. Other studies have employed support vector machine [13], linear discriminant analysis (LDA) [14], Riemannian distances between spatial covariance matrices [15], and signal-to-noise ratios [16] to mitigate the problem. These well-known machine-learning techniques have typically attempted to identify minimal channel sets showing the best decoding accuracies. However, most of these efforts aimed to select optimal channels specialized to each particular BMI paradigm [e.g., motor imagery (MI) or P300-based speller] [7]. Altogether, it is required to develop a deep-learning-based automatic channel-selecting model that is generally applicable to various repertoires of BMI cognitive paradigms.

To evaluate the contribution of each channel to the decoding accuracy, regardless of paradigm type, an intelligent machine-learning model, based on deep learning, is needed [17]–[20]. Deep learning is a specific machine-learning algorithm in which both features and classifiers are connectedly learned

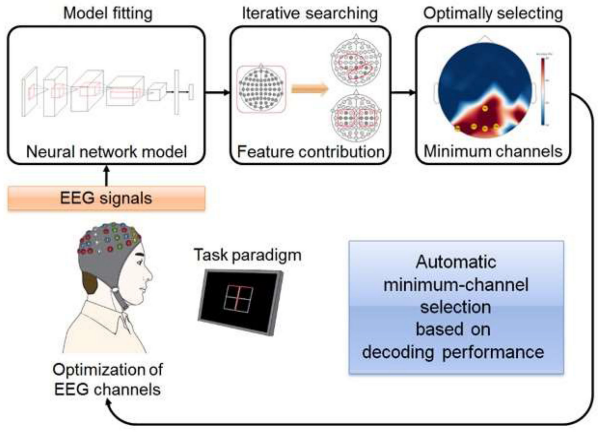

Fig. 1. Workflow of the proposed algorithm for minimally selecting channels for EEG-based BMIs. An automatic minimal-channel searching technique employs a closed-loop system with decoding performance feedback.

directly from the data. Deep-learning models are capable of detecting spatial structural features of a given dataset without the need to handcraft features [21]. Recent advances in deep learning have provided viable approaches extracting features automatically through deep layers of hidden units and have mitigated the drawbacks of conventional machine-learning limitations [21]. Deep learning has demonstrated the capabilities of detecting necessary features despite unwanted signal interference [22]. Hence, it has become a favorable technique for classifying biomedical signals, including EEG data, which typically contain many artifacts. An advantage of deep learning in EEG-channel selection is that the robustness and validity of a selected minimal channel set that corresponds to each BMI paradigm will be further consolidated, as big data are progressively collected for each BMI paradigm. As different EEG features are used during the training of a deep-learning approach, depending on the type of BMI paradigms, it provides an implication to identify significant EEG channels for channel reduction. Consequently, optimized channel sets for specified BMI tasks (designed for portability) can be determined. However, it remains unclear whether the deep-learning approach in EEG-channel selection can be generalizable to the various BMI tasks as well as to different types of EEG features.

Generally, the convolutional neural network (CNN) is a very popular deep-learning approach and is already used in BMI research [23]. A CNN (also known as, ConvNet) is a feed-forward network, in which information flows unidirectionally from the input via the hidden layers to the output [24]. Standard CNN architectures usually stack several convolution and nonlinearity (+ pooling) layers followed by other layers, typically fully connected, which serve as classification layers [25], [26]. Although CNNs have been generally employed for image recognition, this is not the case for electrophysiological signals [27]. Recently, however, there has been growing attention placed on the implementation of CNNs for EEG signal processing [28], [29].

Fundamentally, the necessity of big data in deep-learning approaches is due to the large number of parameters that have to be learned [30]. Therefore, general CNNs do not initially appear to be suitable for a relatively small number of EEG trials. However, a compact CNN specialized for the purpose (i.e., EEGNet) has recently been proposed to overcome the constraints of having access to relatively small numbers of EEG data [31]. EEGNet is optimized for a minimal number of learnable parameters for this reason, and it can be employed to a certain extent when modified techniques (e.g., data augmentation [32]–[36]) are to be considered. It performs well across all tested datasets without the need for data augmentation, making the model simpler to implement [31]. Furthermore, it is noteworthy that neurophysiologically interpretable features instead of artifacts and noise can be extracted from the EEGNet model [31]. This is essential to deciding the optimally minimal EEG channels consistent with their neural significance in each specific cognitive BMI paradigm. Taken together, we employ the compact CNN of EEGNet among a variety of deep-learning techniques.

We propose a compact CNN-based automatic minimal-channel searching technique that can be applied to different types of cognitive BMI paradigms (Fig. 1). As a closed-loop system enhances the ongoing BMI performance with direct real-time feedback [4], [37], the present minimal-channel selection technique employs a closed-loop system with decoding performance feedback. To verify the performance of this algorithm, three different types of BMI paradigms were applied: 1) the typical P300 dataset for auditory modality [38]; 2) the new six-class top-down steady-state visually evoked potential (SSVEP) dataset for visual modality [39]; and 3) the well-known bi-class MI task [40]. To automatically identify the minimally required channels for each BMI paradigm, we evaluate whether the decoding accuracy computed using the minimally selected channels is statistically equivalent to (or higher than) that computed when using all channels. Our approach also assesses the minimal channels that are most probably (greater than 95% confidence interval) selected across all participants. Additionally, based on the selected minimal set of EEG channels, we explore whether the proposed deep-learning approach can provide the neurophysiologically interpretable basis for each experimental modality.

II. Materials and Methods

For the practical use of BMIs, selecting optimal channels helps stabilize or improve classification performance, reduce setup time, and decrease computational complexity [41]. Generally, a channel-selection algorithm consists of four steps: 1) subset generation (i.e., a candidate for a minimum channel set); 2) subset evaluation; 3) stopping criterion; and 4) result validation [7]. Channel subset generation for evaluation is accomplished using a complete, sequential, or random search. Each candidate channel subset is then evaluated and compared with the previous best one until the given stopping criterion is satisfied.

We propose a deep-learning-based minimum-channel-selection method that is applicable to any BMI paradigm, irrespective of the number of channels and the number of decoding classes. Further technical details are provided in the following sections. Briefly, the problem statement of the minimum-channel selection problem is described in Section II-A.

Section II-B presents the minimum-channel-selection compact CNN algorithm. Section II-C–F describes the evaluation procedures of the proposed algorithm using three EEG-based BMI datasets.

A. Problem Statement

The set of all channels is $\mathcal{C} = \{c_i \in \mathbb{R}^{N_{ch}} | i = 1, \ldots, N_{ch}\}$, where N_{ch} is the number of channels is equal to the cardinality, $|\mathcal{C}|$. Let $\boldsymbol{x}_i \in \mathbb{R}^T$ be a single-trial EEG signal recorded on the c_i channel, where T is the number of time samples. Let $\boldsymbol{X}_i = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{N_{ch}}]^T \in \mathbb{R}^{N_{ch} \times T}$ be the ith multichannel EEG trial. Dataset $\mathcal{D}$ contains the individual subject datasets, $\mathcal{D}_1, \ldots, \mathcal{D}_{N_{sbj}}$, where $\mathcal{D}_j$ is the dataset of the jth subject. Each dataset consists of N_{trial}-labeled EEG epochs, $\mathcal{D}_j = \{\boldsymbol{X}_i^j, y_i^j\}_{i=1}^{N_{\text{trial}}}$, where $y_i \in \{1, \ldots, N_{\text{class}}\}$ denotes the corresponding label. Each compact CNN model is fitted per subject j, $f^j : \mathbb{R}^{N_{ch} \times T} \to \mathbb{R}^{N_{\text{class}}}$, for feature extraction and classification. Let $\hat{f}_{\mathcal{C}}^j : \mathbb{R}^{|\mathcal{C}| \times T} \to \mathbb{R}^{N_{\text{class}}}$ be a trained compact CNN model of subject j for all channels, $\mathcal{C}$, with an input size of $|\mathcal{C}| \times T$. $\hat{f}_{\mathcal{S}}^j : \mathbb{R}^{|\mathcal{S}| \times T} \to \mathbb{R}^{N_{\text{class}}}$ is a trained compact CNN model of the jth subject for the subset of all channels denoted by $\mathcal{S}(\mathcal{S} \subseteq \mathcal{C})$, having an input size of $|\mathcal{S}| \times T$. Then, the decoding accuracy of the trained compact CNN model, $\hat{f}_{\mathcal{S}}^j$, of subject j for the subset of all channels, $\mathcal{S}$, is denoted by $A_{\mathcal{S}}^j = \text{Acc}(\hat{f}_{\mathcal{S}}^j, \mathcal{D}_j, \mathcal{S})$. Decoding accuracies of all subjects are denoted by $\boldsymbol{A}_{\mathcal{S}} = [A_{\mathcal{S}}^1, \ldots, A_{\mathcal{S}}^{N_{sbj}}]$. Architectural details of this compact CNN model, f, are described in the supplementary material in IEEExplore.

The objective of the proposed algorithm is to identify a minimum channel set across subjects denoted by $\widehat{\mathcal{S}}_{MC}$, based on any given dataset, $\mathcal{D}$, using the compact CNN model, f, that satisfies

$$\widehat{\mathcal{S}}_{MC} = \arg \min_{\mathcal{S}_{MC} \in \mathcal{C}} |\mathcal{S}_{MC}| \quad \text{s.t. } \text{eqvt}\left(\boldsymbol{A}_{\mathcal{S}_{MC}}, \boldsymbol{A}_{\mathcal{C}}\right) < \alpha \text{ or } \overline{A}_{\mathcal{S}_{MC}} > \overline{A}_{\mathcal{C}} \quad (1)$$

where $|\cdot|$ is the cardinality of the set, α is the threshold for statistical significance, $\overline{A}_{\mathcal{S}_{MC}}$ is the mean accuracy of $\boldsymbol{A}_{\mathcal{S}_{MC}}$, $\overline{A}_{\mathcal{C}}$ is the mean accuracy of $\boldsymbol{A}_{\mathcal{C}}$, and $\text{eqvt}(\boldsymbol{A}_{\mathcal{S}_{MC}}, \boldsymbol{A}_{\mathcal{C}})$ is the equivalency test comparing decoding accuracies that use the minimum channel set and those that use all channels. In this study, we define the minimum channel set as a subset of all channels producing decoding accuracies statistically equivalent to (or higher than) the one using all channels.

B. Minimum-Channel Selection Using Compact CNN Framework

As mentioned earlier, because the compact CNN approach is specifically optimized to a limited dataset size of EEG signals, and neurophysiologically interpretable features can be extracted from the compact CNN model [31], the compact CNN is employed for minimum-channel selection in the present study. To evaluate each channel's effect in decoding performance, the proposed channel-selection method is based on the sequential floating-forward selection algorithm [42], which starts with the most important channel based on its contribution to decoding performance. It gradually adds the other channels one-by-one while checking the improvement of the decoding accuracy updated by the just-added channel. The first training session using all channels is only used to measure the individual contribution of each channel. To individually evaluate each channel's contribution to decoding performance, the decoding accuracies of the trained model are computed using all but one channel set to zero. We hypothesize that the channel scoring higher decoding accuracy will be considered more essential in relation to the given paradigm than that of the lower decoding accuracy.

The compact CNN-based algorithm for minimum-channel selection consists of the following two steps: 1) minimum channel selection *within each subject* based on each channel's contributing score to decoding performance and 2) the minimum channel selection *across subjects* based on step 1 (see Algorithm 1 and Fig. S1 in the supplementary material). In the first step, the contributing score of each channel is individually computed by calculating the decoding accuracy of the trained compact CNN model with all but one channel set to zero.

First, we train the compact CNN model for all channels, $\mathcal{C}$, per each subject. For the process of the all-but-one-channel set to zero, a binary mask, $\boldsymbol{M}_i \in \mathbb{R}^{N_{ch} \times T}$, is applied to preserve only channel, c_i, as follows:

$$\boldsymbol{M}_i(n, m) = \begin{cases} 1, & \text{if } n = i \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

where n and m are the row and column indices of the binary mask, respectively. Using this binary mask, the input EEG signal, $\boldsymbol{X}_k$, is masked while preserving only one-channel, c_i, as follows:

$$\boldsymbol{X}_k^{\text{mask},i} = \boldsymbol{X}_k \odot \boldsymbol{M}_i. \quad (3)$$

The function $\odot$ denotes the elementwise multiplication. The decoding accuracy of $\hat{f}_{\mathcal{C}}$ for the masked-input EEG signal that preserves only one channel, c_i, is counted as the contributing score (SC) for each channel, c_i

$$SC_i = \text{Acc}(\widehat{f}_{\mathcal{C}}, \mathcal{D}^{c_i}, c_i) = \frac{1}{N_{\text{trial}}} \sum_{k=1}^{N_{\text{trial}}} \delta\left(\widehat{f}_{\mathcal{C}}\left(\boldsymbol{X}_k^{\text{mask},i}\right), y_k\right) \quad (4)$$

where $\mathcal{D}^{c_i}$ is the dataset that preserves a channel, c_i, using $\boldsymbol{M}_i$, and $\delta(x, y)$ is the delta function that is one if y is the predicted class of x and zero otherwise. It is assumed that a channel yielding higher decoding accuracy will be regarded as more important in terms of its contribution to decoding performance than one having lower decoding accuracy. Thus, a larger Z-score indicates a larger contribution to decoding performance. To determine the set of minimum channels within each subject, the Z-score transformation of contributing scores for channel c_i is used

$$Z_i = \frac{sc_i - \mu_{sc}}{\sigma_{sc}}, \quad i = 1, \ldots, N_{ch} \quad (5)$$

where $\mu_{sc} = (1/N_{ch}) \sum_{i=1}^{N_{ch}} sc_i$ and σ_{sc} are the mean and standard deviations of the contributing scores, respectively. Channels having a Z-score greater than the threshold, $\rho = 1.96$

(i.e., 95% confidence interval), are selected as the minimum channels for subject j, denoted as $\mathcal{S}^j_{\text{indivMC}}$

$$\mathcal{S}^j_{\text{indivMC}} = \{c_i : Z_i > 1.96\}. \tag{6}$$

This procedure is conducted for each subject. Subsequently, the minimum channel set per subject is obtained as $\mathcal{S}_{\text{indivMC}} = \{\mathcal{S}^1_{\text{indivMC}}, \ldots, \mathcal{S}^{N_{sbj}}_{\text{indivMC}}\}$, where $\mathcal{S}^j_{\text{indivMC}} \subseteq \mathcal{C}$. For stopping criteria to determine the minimum channels across subjects, the decoding accuracy of subject j for all channels is computed as follows:

$$A^j_{\mathcal{C}} = \text{Acc}\left(\widehat{f}^j_{\mathcal{C}}, \mathcal{D}_j, \mathcal{C}\right). \tag{7}$$

In the second step, the minimum channels across subjects are determined using iterative searching, based on whether it yields decoding accuracy statistically equivalent to (or higher than) that of all channels in the level of all subjects. Let n_i be the number of occurrences of channel c_i in the minimum-channel sets combined from all subjects, $\mathcal{S}_{\text{indivMC}}$. Let $\mathcal{C}^{\text{rank}}$ be reordered from the set of all channels, $\mathcal{C}$, based on their occurrence frequencies, n_i. Before the iterative-searching process for a minimum-channel set, $\widehat{\mathcal{S}}_{MC}$, that is generalized across all subjects, initial minimum-channel candidates are selected using the same Z-score criterion that is applied to determine the minimum channels within each subject. The initial minimum-channel candidates are considered as the most probable channels (noted as $\mathcal{S}_{mp}$ in the algorithm) when those channels are included in the most frequently (greater than 95% confidential interval) selected channels across all subjects. A candidate set of minimum channels, $\mathcal{S}_{MC}$, is evaluated until the following stopping criterion is satisfied. The stopping criterion is defined such that the decoding accuracies using the minimum channels, $\boldsymbol{A}_{\mathcal{S}_{MC}}$, are statistically equivalent to (or higher than) those using all channels $\boldsymbol{A}_{\mathcal{C}}$. To statistically examine the equivalence between these two decoding accuracies, two one-sided t-tests (i.e., equivalence tests) are used [43]. For this test, the null hypothesis states that differences between decoding accuracies from using all channels and the minimum-channel set are large enough to be considered worthwhile. If the null hypothesis is rejected, then the decoding accuracies between all channels and the minimum-channel set can be regarded as statistically equivalent. Even if the null hypothesis is accepted, there are two expected cases: the selected channels produce either lower or higher decoding accuracies than when using all channels. In the case of higher decoding accuracy being detected, the iteration procedure of channel selection is also stopped because even a small set of selected channels can yield better decoding performance than all channels. If the stopping criterion is not satisfied, the set of minimum-channel candidates, $\mathcal{S}_{MC}$, is sequentially updated with an additional channel, based on the systematic comparison between the decoding accuracy of current t and previous $t-1$ iterations, that is: 1) if the newly updated accuracy is higher than the result obtained by the previous iteration, the next-ranked channel, c^{rank}_{t+1}, is added to the set of minimum-channel candidates, $\mathcal{S}_{MC}$ and 2) if the newly updated accuracy is lower than the result obtained by the previous iteration, the channel, c^{rank}_t, that was recently added in the current iteration is rejected, and the next-ranked channel, c_{t+1}, is added to the set of minimum-channel candidates, $\mathcal{S}_{MC}$. Although, in this study, we employed a sequentially adding-channel approach until the next-ranked channel in performance priority did not contribute to an increase in current accuracy, other methodological approaches are also applicable. Thus, the best approach leading to the highest accuracy should be further investigated in future studies.

C. Three EEG Datasets for Evaluating the Proposed Approach

To test whether our algorithm can be applied to different BMI paradigms, three types of BMI paradigms are analyzed. Study 1 applies a typical P300-based BMI (auditory task), and Study 2 applies a new top-down SSVEP-based BMI (visual task). These studies reanalyze the datasets of previous studies [38], [39]. Study 3 assesses an MI dataset from other researchers [40].

In Study 1, the EEG data were recorded from 14 healthy individuals (seven females; age 28.2 (mean) $\pm$ 6.8 (standard deviation) years). In Study 2, the EEG data were recorded from 20 healthy individuals (ten females; age 25.7 $\pm$ 4.6 years). In Study 3, the EEG data of 20 healthy individuals (ten females; age 24.3 $\pm$ 3.7 years) were randomly selected from the pool of 52 participants collected [40]. Studies 1 and 2 were conducted in accordance with the ethical guidelines established by the Institutional Review Board (IRB) of Hallym University College of Medicine (IRB No. 2016-I013 for Study 1), Korea University (IRB No. KU-IRB-13–43-A-2 for Study 2), and the Declaration of Helsinki (World Medical Association, 2013). Study 3 was approved by the IRB of Gwangju Institute of Science and Technology (IRB No. 20130527-HR-02).

1) Study 1 (Auditory P300): To test a typical P300-based BMI paradigm, we use an auditory oddball dataset [38]. During EEG acquisition, the participants performed an auditory oddball task [see Fig. 2(a)]. In the oddball paradigm, two types of stimuli were presented at random. One stimulus occurred less frequently than the others (i.e., the oddball). The participant was required to discriminate the rare stimulus (target) from the frequent ones (standard) by noting the occurrence of the target, typically by pressing a button [44]–[46]. Because the P300 component of the event-related potential (ERP) reflects fundamental cognitive processes [47]–[50] and is often obtained using the oddball paradigm [51], the oddball task is employed in this study.

The oddball task consisted of 80 target stimuli and 320 standard stimuli, which were presented at random. The frequency of the target stimulus was 8 kHz and that of the standard stimulus was 500 Hz. The length of each auditory stimulus was 200 ms, with 10 ms for each rising and falling phase. Each auditory stimulus was presented for 200 ms with a variable interstimulus interval (ISI) ranging randomly between 1300 and 1700 ms, centered at 1500 ms.

2) Study 2 (Steady-State Visually Evoked Potential): To evaluate the other type of BMI paradigm, a newly proposed visual top-down SSVEP-based BMI dataset [39] was used. In the top-down SSVEP paradigm, a 6 $\times$ 6-cm grid-shaped

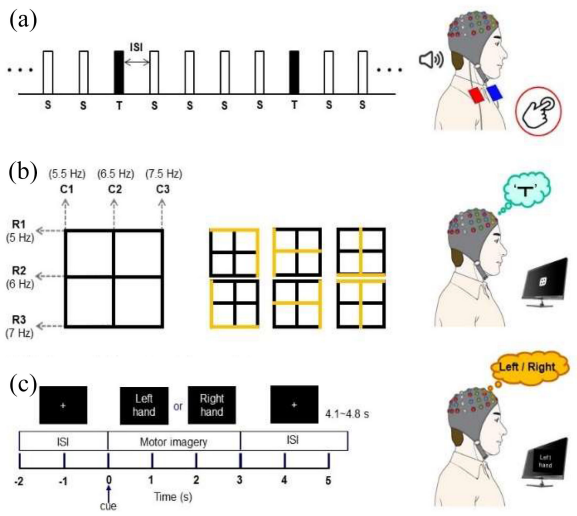

Fig. 2. Experimental designs of Studies 1, 2, and 3. (a) Study 1: A task-flow diagram with a series of auditory stimuli comprising both target and standard sounds. White bars and the letter "S" indicate standard stimuli, and black bars and the letter "T" indicate target stimuli. During the oddball task, the participants were instructed to perform a sound-discrimination task by pressing a button. (b) Study 2: Top-down SSVEP paradigm with an SSVEP-inducing grid-shaped line array consisting of three rows (R1–R3) and three columns (C1–C3) of the individually flickering lines. The middle panel depicts the examples of attended flickering-line composites (in yellow). (c) Study 3: At the beginning of each trial, the monitor showed a black screen with a fixation cross for 2 s; the participant was then ready to perform the MI experiment. One of two instructions ("left hand" or "right hand") appeared randomly on the screen for 3 s, and participants were instructed to imagine four actual finger movements.

flickering line array was designed to induce a mentally generated letter within the restricted visual angle of the participant to evoke a corresponding SSVEP. To generate individual SSVEPs based on every flickering line, each row and column had an individual flickering frequency ranging between 5 and 7.5 Hz [see Fig. 2(b)]. These frequencies are shown to be effective in inducing SSVEPs in humans [52], [53].

The underlying concept behind decoding the thoughts of a participant using SSVEPs induced by our grid-shaped line array is as follows. When a participant pays attention to a subset of flickering lines, whose combination represents the shape of a letter or symbol, we expect that the frequencies corresponding to those lines will be detected as dominant SSVEP features. This experiment comprised four blocks with a short break in between; each block included 60 trials. In each block, each of six predefined Korean letters (ㄱ, ㄴ, ㅗ, ㅜ, ㅏ, and ㅓ) was cued ten times for each participant for perception by a 1-s auditory cue presented 500 ms before the onset of a flickering grid-shaped line array. The participants were instructed to simultaneously attend to the two lines that jointly comprised a letter. Intertrial intervals ranged from 1000 to 1500 ms, centered at 1250 ms. After a 1-s auditory cue (an analog instruction sound), which pronounced the Korean letter to which the participant was required to attend; after a subsequent 500-ms buffer period, the grid-shaped line array was presented for 5 s. The experimental condition had six classes of stimuli, and each class comprised 40 trials. Consequently, each experimental condition had 240 trials.

3) Study 3 (Motor Imagery): In Study 3, we analyzed an EEG dataset of MI of left and right hands, as studied by other researchers [40]. A BMI experiment for MI was conducted with 52 participants (19 females; age, 24.8 $\pm$ 3.9 years), but the current study randomly selected 20 participants (ten females; age, 24.3 $\pm$ 3.7 years) among them. Before beginning the MI experiment, participants were asked to conduct real hand movements. In the beginning of each trial, the monitor showed a black screen with a fixation cross for 2 s. The participant was then ready to perform hand movements [after the black screen gives a cue to the participant; Fig. 2(c)]. One of two instructions ("left hand" or "right hand") appeared randomly on the screen for 3 s, and participants were requested to move the appropriate hand depending on the instruction given. After the movement, when the blank screen reappeared, the participant was given a break ranging randomly from 4.1 to 4.8 s. These processes were repeated 20 times for one class (one run). If additional details about the experimental protocol of Study 3 are necessary, please refer to [40].

The MI experiment was conducted using the same paradigm as the real hand-movement experiment. Participants were instructed to imagine four actual finger movements: 1) touching each index; 2) middle; 3) ring; and 4) little finger to the thumb within 3 s. Each participant practiced these actual finger movements, then performed the MI experiment. Five or six runs were performed. After each run, the classification accuracy was computed over one run and given to the participant as feedback to increase motivation. In the present study, we used only MI data for channel selection.

D. EEG Acquisition and Preprocessing

Regarding Studies 1 and 2, EEG signals were measured using a BrainAmp DC amplifier (Brain Products, Germany) with 32 Ag/AgCl electrodes in an actiCAP (Brain Products, Germany) in accordance with the international 10–10 system. An electrode was placed on the tip of the nose as a reference, and a ground electrode was placed at electrode AFz. Electrode impedances were maintained below 5 kΩ prior to data acquisition. The EEG was recorded at 1000 Hz for Study 1 and 500 Hz for Study 2. The eye-movement activity was monitored using an electrooculogram (EOG) electrode placed suborbitally on the left side, and vertical and horizontal electro-ocular activity was computed by using two pairs of electrodes placed vertically and horizontally with respect to both eyes (i.e., Fp1 and EOG for the vertical EOG, F7, and F8 for the horizontal EOG). The EOG activity was corrected offline using independent component analysis [54]. The EEG was segmented from 500-ms prestimulus to 1000-ms post-stimulus for Study 1 and from 500-ms prestimulus to 5000-ms post-stimulus for Study 2. EEG epochs having amplitudes higher than $+100$ μV or lower than -100 μV and with gradients higher than 50 μV/ms were automatically excluded. Thus, if the EEG amplitude per unit time (ms) exceeded ± 50 μV in every moving time window, the epoch matching this criterion was automatically rejected. In Study 2, EEG data were downsampled to 125 Hz.

In Study 3, EEG data were recorded at 512 Hz using a Biosemi ActiveTwo system (BioSemi B.V., The Netherlands)

with 64 Ag/AgCl electrodes based on the international 10–10 system. To make the number of channels comparable to Studies 1 and 2, the corresponding 30-out-of-64 channels were used for further analyses. For preprocessing, Butterworth filtering at the fourth order was used for high-pass and band-pass filtering. If a band-passed (8–30 Hz) trial had an amplitude greater than ± 100 μV within 0.5–2.5 ms, the trial was rejected. The frequency band involved somatosensory rhythm [55], [56], and the time window was determined using an algorithm for selecting a discriminative time interval [56]. In addition, this frequency range was employed to minimize any possible high-frequency frontal and temporal muscle artifacts [57], [58]. All trials for each subject were preprocessed with high-pass filtering and a common average reference. Then, it was filtered both spectrally (8–30 Hz) and temporally (0.5–2.5 s after stimulus onset) and downsampled to 128 Hz.

For each dataset of Studies 1, 2, and 3, each epoch was a matrix X where the number of rows was the number of samples, and the number of columns was the number of channels, and each epoch was associated to a class. In Study 1, 80 epochs were used for the target condition, and for the standard condition, the same number of 80 epochs was randomly selected out of the 320 epochs. Therefore, 80 trials per condition were used, and each trial's dimension was 30 channels $\times$ 200 time points. In Study 2, 40 epochs were used for every six conditions, and each trial's dimension was 30 channels $\times$ 625 time points. In Study 3, 100 epochs were used for each of two conditions (i.e., left- and right-hand MI), and each trial's dimension was 30 channels $\times$ 256 time points.

To reveal whether the channel-selected brain regions were neurophysiologically consistent with the most dominantly activated brain regions of each BMI paradigm, responsive brain topographical patterns were additionally computed and compared with the corresponding channel-selected brain regions.

In Study 1, the scalp distribution of the P300 peak was used because P300 is an ERP component in the time domain. For Studies 2 and 3, topographical activation patterns for the neural bases of EEG features [59] were used to efficiently exhibit their spectral EEG features in the frequency domain. We estimated the activation pattern involving all scalp channels by computing the correlation between the continuous trialwise classifier output and the time-course of EEG features from all channels. Assuming that the task-relevant and task-irrelevant (i.e., uninformative) signals are uncorrelated, the activation pattern is given by the correlation between the classifier output and the time course of individual features [59].

E. Training and Evaluation

To train the compact CNN, the *He*-normal initializer [60] was used as the weight initialization method. An Adam optimizer [61] was employed with a learning rate $\alpha = 0.001$, and decay parameters $\beta 1 = 0.9$ and $\beta 2 = 0.999$ were used to minimize the categorical cross-entropy loss function. The batch size and training iterations (epochs) were set to 32 and 1000, respectively. An early stopping method was used to train the model while monitoring the loss on the validation dataset. The best model was saved by waiting for 200 epochs of no improvements to pass before the training was stopped. We split the dataset into training, validation, and test subsets at portions of 5:1:2, respectively [62]. The model was trained using the training set, and the trained model was evaluated using the validation set to save the best one.

During this process, the training and validation sets were used to determine the trained model of the minimal-channel selection, and the remaining trials were then tested using this trained model. When splitting each fold, trials were equally distributed across all classes. This decoding procedure was performed for each participant separately. The decoding accuracy was computed based on the signals of all channels. Finally, to compare this result, the decoding accuracy using the minimally selected channels was also computed. To statistically examine whether the decoding accuracies obtained by using the minimally selected channels were significantly equivalent to those obtained using all channels, equivalence tests were performed [43]. The equivalence test yielded statistical significance when $p < 0.05$, given equivalence bounds of -2.5 and 2.5 (on a raw scale). Additionally, Pearson's correlation coefficients were computed to assess the statistical relationship between the decoding accuracies of all and minimal channels.

All analyses were performed using the Python programming language (Python Software Foundation, https://www.python.org), and Keras [63] with a TensorFlow backend [64] was used for CNN model implementation.

F. Comparisons With Other Algorithms

It is noteworthy that most of the previously proposed channel selection algorithms [6]–[8] selected an appropriate subset of channels yielding the best decoding performance. The present study, by contrast, aims to reduce the number of EEG channels to the lowest possible number while maintaining decoding performance that is statistically equivalent to (or higher than) using all channels. In this respect, it is difficult to directly compare the present study with the previous ones. However, given the statistically equivalent decoding accuracies, the number of selected channels and neurophysiological accounts of the selected channel locations can be compared across different channel-selection algorithms. Thus, two other channel-selection algorithms (i.e., LDA [14] and ConvNet [29]) are introduced to compare performances with the present study. LDA [14] and ConvNet [29] were employed to select channels because their performance enables feature training to decode task-related information with higher efficiency. In the ConvNet approach, five standard convolutional layers are used. Using the same dataset, these two algorithms for channel selection are compared with the proposed compact CNN approach. Additionally, the standard set of eight channels studied for the P300-based BMI Study 1: Fz, Cz, P3, Pz, P4, PO7, PO8, and Oz [65]; the SSVEP-based BMI Study 2: PO3, POz, PO4, PO7, PO8, O1, Oz, and O2 [22]; and the MI-based BMI Study 3: C3, C4, C5, C6, FC3, FC4, CP3, and CP4 [66] were parallelly analyzed for performance comparison with our method. However, because the datasets used in Studies 1 and

TABLE I
COMPARISON OF DECODING ACCURACIES (%) AND NUMBERS OF SELECTED CHANNELS OBTAINED FOR DIFFERENT CHANNEL-SELECTION ALGORITHMS IN THE P300-ODDBALL (AUDITORY) PARADIGM (STUDY 1)

ID	AC (30-ch)	MC & MP (2-ch)	LDA (19-ch)	CN (5-ch)	6-ch
1	75.0	87.5	87.5	62.5	87.5
2	57.1	71.4	71.4	57.1	16.7
3	97.1	97.1	94.1	94.1	94.1
4	87.5	77.5	85.0	77.5	85.0
5	66.7	50.0	50.0	66.7	83.3
6	88.9	75.0	61.1	77.8	77.8
7	90.0	95.0	92.5	72.5	80.0
8	69.2	76.9	84.6	76.9	92.9
9	53.8	69.2	61.5	61.5	65.4
10	90.5	81.0	71.4	71.4	81.8
11	88.5	84.6	88.5	88.5	80.8
12	92.5	82.5	92.5	82.5	87.5
13	78.9	73.7	60.5	73.7	81.6
14	78.4	54.1	73.0	56.8	66.7
Mean	79.6	76.8	76.7	72.8	77.2

*ID: subject ID; AC: all channels; MC: minimum channels of the compact CNN (i.e., Cz and FC1); MP: most probable channels of the compact CNN (i.e., Cz and FC1); LDA: linear discriminant analysis (i.e., Fz, F3, F4, F7, F8, FC1, FC2, FC5, FC6, Cz, C3, C4, CP1, CP6, P3, P7, T7, T8, and Oz); CN: ConvNet (i.e., Fz, FC1, Pz, P3, and Oz); and 6-ch: the standard six channels (i.e., Fz, Cz, Pz, P3, P4, and Oz). The number of selected channels is noted in parentheses.

2 were based on a 32-channel configuration, only six channels among the standard eight were available for Study 1 (i.e., Fz, Cz, Pz, P3, P4, and Oz) and only three were available for Study 2 (i.e., O1, Oz, and O2). Accordingly, those available channels were inevitably selected for this comparison. To statistically compare the decoding accuracies across these different algorithms, paired t-tests were employed.

III. RESULTS

A. Study 1

The two centro-frontal channels (Cz and FC1) were automatically selected for the auditory P300-based paradigm using a compact CNN model. As shown in Fig. 3 and Table I, the decoding accuracies obtained using compact CNN-based minimum channels (76.8%) were computed using the model parameters statistically equivalent to (or higher than) those obtained using all channels (79.6%). The accuracy ratio of the minimum channels to all channels was 96.5%. Additionally, the accuracies computed by using both all channels and minimally selected channels were highly correlated [$r(14) = 0.581$, $p < 0.05$]. From Table I, it is evident that the central channels (e.g., Cz and FC1) were also selected as the most probable channels across all participants. Because these two channels were identical to the minimally selected channels, this comparison is not presented in Table I to avoid redundancy. As shown in Table I, the decoding accuracies of LDA and ConvNet algorithms were computed using the model parameters for minimally selected channels, yielding statistically equivalent (or better) performance compared with using all channels (LDA: 76.7%; ConvNet: 72.8%). Additionally, their numbers of selected channels were adversely more numerous than that of the compact CNN (compact CNN: 2; LDA: 19; ConvNet: 5). Furthermore, their selected channels did not match with the neurophysiologically interpretable region. This

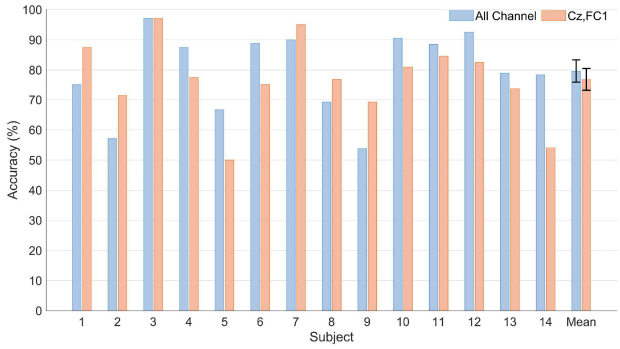

Fig. 3. Comparison between the decoding accuracies obtained using all channels and those obtained using the compact CNN-based minimal channels (Cz and FC1) for 14 participants (Study 1). Error bars represent standard deviations.

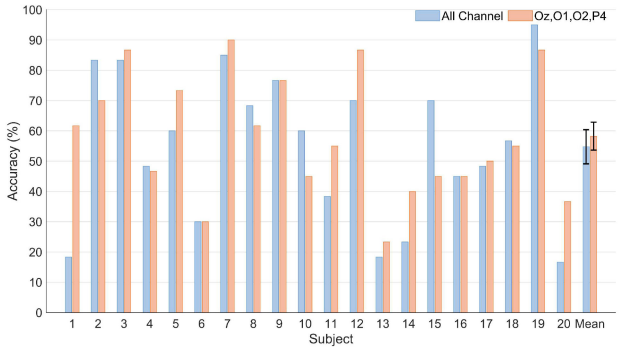

Fig. 4. Comparison between the decoding accuracies obtained using all channels and those obtained using a compact CNN-based minimal channels (Oz, O1, O2, and P4) across 20 participants (Study 2). Error bars represent standard deviations.

means that the locations of selected channels were not focused on the vertex of the brain, which is the dominantly activated region during the performance of auditory P300 paradigm (LDA: Fz, F3, F4, F7, F8, FC1, FC2, FC5, FC6, Cz, C3, C4, CP1, CP6, P3, P7, T7, T8, and Oz; ConvNet: Fz, FC1, Pz, P3, and Oz). Although the typical six channels yielded significantly similar decoding accuracies to the compact CNN [$t(13) = -0.077$, ns; compact CNN: 76.8%; 6-ch: 77.2%], the compact CNN was advantageous in terms of the fewer number of minimally selected channels (compact CNN: 2; 6-ch: 6).

B. Study 2

The four occipito-parietal channels (Oz, O1, O2, and P4) were automatically selected for the top-down SSVEP-BMI paradigm using a compact CNN model. Fig. 4 and Table II demonstrate the decoding accuracies of all 20 participants. The average accuracy of the minimum channels was 58.3%, computed using the model parameters significantly equivalent to (or higher than) those of all channels (54.7%). The ratio of their decoding accuracies (minimum channels versus all channels) was 106.6%. Additionally, the accuracies computed using all channels and minimally selected channels were highly correlated [$r(20) = 0.794$, $p < 0.001$]. From Table II, it is evident that the occipital channels, Oz and O1, were selected as the most probable channels across all participants. As shown

TABLE II
COMPARISON OF DECODING ACCURACIES (%) AND NUMBERS OF SELECTED CHANNELS OBTAINED FOR DIFFERENT CHANNEL-SELECTION ALGORITHMS IN THE TOP-DOWN SSVEP (VISUAL) PARADIGM (STUDY 2)

ID	AC (30-ch)	MC (4-ch)	MP (2-ch)	LDA (2-ch)	CN (4-ch)	3-ch
1	18.3	61.7	21.7	18.3	15.0	63.3
2	83.3	70.0	46.7	31.7	68.3	73.3
3	83.3	86.7	58.3	33.3	85.0	66.7
4	48.3	46.7	33.3	26.7	26.7	50.0
5	60.0	73.3	53.3	21.7	51.7	73.3
6	30.0	30.0	28.3	15.0	18.3	36.7
7	85.0	90.0	68.3	48.3	81.7	96.7
8	68.3	61.7	60.0	13.3	71.7	86.7
9	76.7	76.7	46.7	35.0	83.3	76.7
10	60.0	45.0	45.0	45.0	20.0	36.7
11	38.3	55.0	33.3	28.3	48.3	66.7
12	70.0	86.7	65.0	35.0	20.0	80.0
13	18.3	23.3	21.7	25.0	10.0	26.7
14	23.3	40.0	35.0	18.3	16.7	46.7
15	70.0	45.0	23.3	11.7	18.3	33.3
16	45.0	45.0	26.7	20.0	15.0	40.0
17	48.3	50.0	36.7	30.0	65.0	46.7
18	56.7	55.0	51.7	31.7	18.3	50.0
19	95.0	86.7	75.0	13.3	71.7	83.3
20	16.7	36.7	23.3	21.7	16.7	36.7
Mean	54.7	58.3	42.7	26.2	41.1	58.5

*ID: subject ID; AC: all channels; MC: minimum channels of the compact CNN (i.e., Oz, O1, O2, and P4); MP: most probable channels of the compact CNN (i.e., Oz and O1); LDA: linear discriminant analysis (i.e., Oz and O2); CN: ConvNet (i.e., Oz, O1, O2, and P3); and 3-ch: the standard 3 channels (i.e., Oz, O1, and O2). The number of selected channels is noted in parentheses.

TABLE III
COMPARISON OF DECODING ACCURACIES (%) AND NUMBERS OF SELECTED CHANNELS OBTAINED FOR DIFFERENT CHANNEL-SELECTION ALGORITHMS IN THE MI PARADIGM (STUDY 3)

ID	AC (30-ch)	MC (4-ch)	MP (2-ch)	LDA (3-ch)	CN (6-ch)	8-ch
1	85.7	85.7	67.4	63.3	69.4	77.6
2	62.0	88.0	68.0	76.0	78.0	84.0
3	52.5	67.5	55.0	65.0	57.5	60.0
4	68.0	76.0	72.0	52.0	58.0	66.0
5	76.0	50.0	46.0	44.0	60.0	50.0
6	66.7	56.7	53.3	60.0	66.7	71.7
7	52.0	46.0	48.0	66.0	36.0	54.0
8	66.0	64.0	48.0	62.0	58.0	56.0
9	64.0	56.0	58.0	64.0	54.0	78.0
10	52.0	74.0	60.0	72.0	54.0	62.0
11	94.0	94.0	74.0	86.0	96.0	94.0
12	98.0	96.0	70.0	94.0	94.0	98.0
13	76.0	78.0	78.0	78.0	90.0	86.0
14	94.0	86.0	88.0	84.0	88.0	94.0
15	42.6	63.8	59.6	57.4	46.8	59.6
16	76.1	67.4	52.2	78.3	52.2	63.0
17	68.0	66.0	66.0	72.0	52.0	50.0
18	46.9	63.3	53.1	51.0	55.1	65.3
19	64.0	64.0	46.0	52.0	70.0	72.0
20	72.0	74.0	76.0	66.0	76.0	86.0
Mean	68.8	70.8	61.9	67.1	65.6	71.4

*ID: subject ID; AC: all channels; MC: minimum channels of the compact CNN (i.e., C3, C4, FC5, and PO7); MP: most probable channels of the compact CNN (i.e., C3 and FC5); LDA: linear discriminant analysis (i.e., C3, C4, and P3); CN: ConvNet (i.e., F3, FC2, C3, CP2, CP6, and P4); and 8-ch: the standard eight channels (i.e., C3, C4, C5, C6, FC3, FC4, CP3, and CP4). The number of selected channels is noted in parentheses.

in Table II, the decoding accuracies of LDA and ConvNet algorithms were computed using the model parameters for minimally selected channels, yielding statistically equivalent (or better) performance, compared with using all channels (LDA: 26.2%; ConvNet: 41.1%). Although the compact CNN and LDA methods shared the two occipital channels (i.e., Oz and O2) as the minimally selected ones, the decoding accuracy of LDA was significantly lower than that of the compact CNN [$t(19) = 7.555$, $p < 0.001$; compact CNN: 58.3%; LDA: 26.2%]. Regarding the ConvNet, although the number of selected channels using ConvNet was the same as that of the compact CNN (compact CNN: 4; ConvNet: 4), the decoding accuracy of ConvNet was significantly lower than that of the compact CNN ($t(19) = 3.959$, $p < 0.005$; compact CNN: 58.3%; ConvNet: 41.1%). The typical three channels yielded significantly similar decoding accuracies to the compact CNN [$t(19) = -0.120$, ns; compact CNN: 58.3%; 3-ch: 58.5%]

C. Study 3

The four centro-lateral channels (i.e., C3, C4, FC5, and PO7) were automatically selected for the MI paradigm using a compact CNN model. As shown in Fig. 5 and Table III, the decoding accuracies obtained using compact CNN-based minimum channels (70.8%) were computed using the model parameters for minimally selected channels, which yielded statistically equivalent (or better) performance compared to using all channels (68.8%). The accuracy ratio of the minimum channels to all channels was 102.9%. Additionally, these accuracies were highly correlated ($r(20) = 0.633$, $p < 0.005$). From Table III, it is evident that the two centro-frontal channels (i.e., C3 and FC5) were selected as the most probable channels across all participants. As shown in Table III, the decoding accuracies of LDA and ConvNet algorithms were computed using the model parameters for minimally selected channels, which yielded statistically equivalent (or better) performance compared to using all channels (LDA: 67.1%; ConvNet: 65.6%). Although the number of minimally selected channels by LDA was one less than that of the compact CNN (compact CNN: 4; LDA: 3), the decoding accuracy of LDA was not significantly different from that of the compact CNN ($t(19) = 1.559$, ns). Regarding ConvNet, its number of selected channels (i.e., six) was higher than that of the compact CNN (i.e., 4), and the selected channels of ConvNet were not located around the

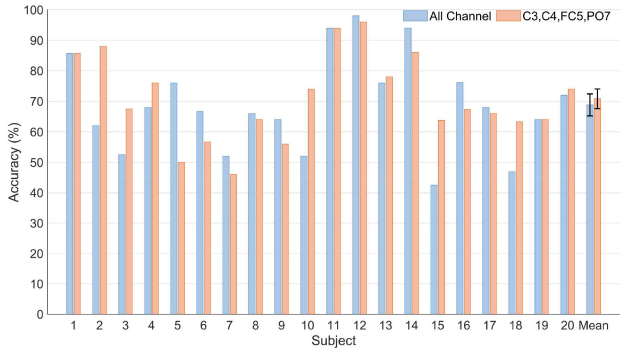

Fig. 5. Comparison between the decoding accuracies obtained using all channels and those obtained using a compact CNN-based minimal channels (C3, C4, FC5, and PO7) across 20 participants (Study 3). Error bars represent standard deviations.

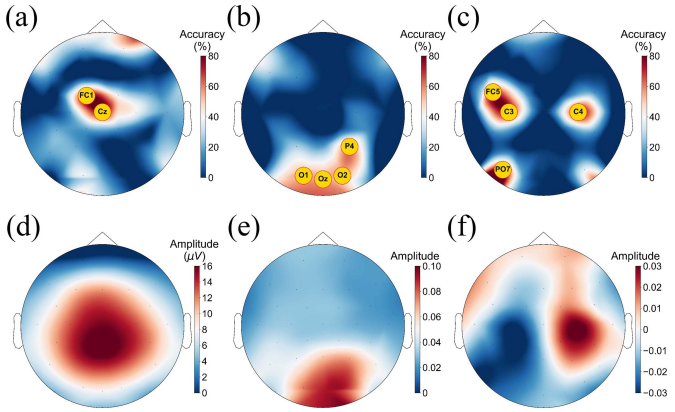

Fig. 6. Topographical distributions of the grand-averaged decoding accuracy of (a) Study 1, (b) Study 2, and (c) Study 3 after the final search for a set of minimal channels. The minimally selected channels for each modality (auditory for Study 1, visual for Study 2, and motor for Study 3) are highlighted. Their corresponding neurophysiologically responsive topographical patterns are also displayed in parallel: (d) for Study 1, (e) for Study 2, and (f) for Study 3. Note the consistently overlapped brain regions across these two approaches: minimal-channel selection and task-relevant activation pattern, in all three paradigms. The topographical view is from the vertex, with the nasion at the top right of the image. Color-scaled bars represent the grand-averaged decoding accuracies of each channel for (a), (b), and (c); P300 amplitudes for (d); and amplitudes of activation-pattern maps for (e) and (f).

neurophysiologically interpretable regions (i.e., C3 and C4 for MI). Although the typical eight channels yielded significantly decoding accuracies similar to those of the compact CNN [$t(19) = -0.245$, ns; compact CNN: 70.8%; 8-ch: 71.4%], the compact CNN was advantageous in terms of having fewer minimally selected channels (compact CNN: 4; 8-ch: 8).

As shown in Fig. 6, Studies 1, 2, and 3 exhibited the most prominent regions of the brain (i.e., centro-frontal regions for the auditory paradigm, occipital regions for the visual paradigm, and lateralized central regions for the MI paradigm) for automatic selection of minimal EEG channels. As shown in Fig. 6, the locations of neurophysiological activation overlapped considerably with those of the compact CNN-based channel-selections in each experimental paradigm. These consistent topographical overlappings imply that the proposed deep-learning approach can provide neurophysiologically explainable results.

A brief demonstration of our automatic selection of the compact CNN-mediated minimal channels for EEG-based BMIs is presented as a video clip in the supplementary material (see Supplementary Video Clip 1).

IV. Discussion

In this study, a deep-learning approach for minimum-channel selection was evaluated using three EEG datasets: 1) auditory [38]; 2) visual [39]; and 3) motor-imagery [40] BMI datasets, each with their dataset sizes and feature characteristics. By applying a compact CNN model [31], we identified an optimally minimized channel set for all of these different EEG-based BMI paradigms with neurophysiologically consistent rationales (Fig. 6). Accordingly, we validated that the compact CNN in the EEG channel selection was generally applicable across different BMI paradigms in the presence of limited data and could extract neurophysiologically interpretable features. Generally, because ocular artifacts (e.g., eye movements and blinks) are regarded as the major interference in EEG-based BMI studies, artifact compensation methods, such as independent component analysis or adaptive filters, are frequently introduced to isolate pure cerebral EEG signals [67]. However, because the compact CNN of the EEGNet model is principally designed to extract neurophysiologically significant features (interpretable with a given task) as opposed to artifact or noise features [31], the proposed channel selection method is robust against task-irrelevant artifact contamination. Furthermore, the decoding accuracies obtained using the minimum channels were statistically equivalent to (or even higher than) those obtained using all channels. This provided us with a feasible concept for the automatic selection of optimized minimal EEG channels using a compact CNN model, thereby aiding in the development of a potent portable BMI technology.

In the auditory EEG oddball paradigm, the channels Cz and FC1 were selected as the minimum channels as well as the most probable channels across all the participants. Because the dipole sources of auditory processing generate the most prominent activity around the vertex of the brain (i.e., the location of the Cz channel) [68], we consistently observed an automatic selection of centro-frontal channels in the auditory EEG paradigm. These channels were minimally selected based on the compact CNN model, which yielded decoding accuracies statistically equivalent to (or higher than) those that used all channels. This implies that only the minimally selected channels could sufficiently represent the decoding performance generated by all channels. Furthermore, this minimal set of channels was also consistent with the task-relevant neurophysiological activities. Because the oddball task required participants to respond actively based on a cognitive decision related to the presented stimulus type during the performance of the task, the results obtained from this task were interpreted principally as auditory top-down effects.

Similarly, in the visual top-down SSVEP paradigm, the two occipital channels (i.e., Oz and O1) were selected as the most probable channels across all participants. These occipital channels were a concise set of representative channels that best reflected the characteristics of the visual top-down SSVEP-BMI paradigm. This finding is consistent with neurophysiological rationales because the occipital regions of these channels are around the primary visual cortex, which is considered an essential brain region for processing the top-down SSVEP-BMI task. Moreover, the compact CNN-based occipito-parietal channels (i.e., Oz, O1, O2, and P4), centered at the most probable two occipital channels, were minimally selected based on the model parameters, which yielded decoding accuracies statistically equivalent to (or higher than) those from using all channels. This indicates that the minimally selected occipito-parietal channels were sufficient for computing the decoding accuracy of all channels.

This study, which used the compact CNN model, also yielded a distinctly enhanced decoding accuracy (54.7% using all channels and 58.3% using the four minimally selected occipito-parietal channels) when compared with the previous

study, which used regularized LDA with shrinkage (35.8% using all channels and 42.5% using the three occipital channels) [39]. This implies that the deep-learning approach improved the decoding performance when compared with the linear machine-learning technique. This obvious enhancement in decoding performance is another advantage of the compact CNN classifier over the linear classifier of the SSVEP dataset using minimal channel subsets. Furthermore, this study consistently provided neurophysiologically interpretable accounts (see Fig. 6), although one of the issues still debated in deep-learning studies is whether or not it is explainable. In this respect, the current deep-learning-based decoding approach may contribute to a feasibility study that could shed light on the potential explanations underpinning artificial intelligence.

Regarding the MI paradigm, the compact CNN-based centro-lateral channels (i.e., C3, C4, FC5, and PO7) were minimally chosen based on model parameters, which yielded decoding accuracies statistically equivalent to (or higher than) those of using all channels. This observation is also consistent with the neurophysiological findings of motor imagery [69]. During an MI task, hand positions of the motor homunculus in both hemispheres are located around the C3 and C4 positions within the primary motor cortex, respectively, for right/left-hand motor imagery [70]. For example, distinct MI patterns emerge for both symmetric bilateral C3 and C4 positions using EEG phase synchronization combined with spectral activity [71]. All these observations reflect that the proposed channel selection method yielded a paradigm-independent robust performance with consistent neural significance.

In principle, the performance evaluation of the neural network with all but one channel set to zero would provide lower performance than using all channels. However, the minimal channels yielded decoding performances statistically equivalent to (or even higher than) those of all channels in the present study. This is understandable from the viewpoint of the characteristics of EEG signals. This also means that it is common in EEG studies that a few selected channels around the task-relevant brain regions (rather than all channels) can achieve the best decoding performance. This is because the most activated brain regions are principally based on the neurophysiologically regional characteristics of specific cognitive tasks. Therefore, the EEG signals (or features) from task-irrelevant channels (located around the less or nonactivated brain regions) might not be informative. Thus, they could disrupt optimal performance [72], [73]. Accordingly, only a small number of task-relevant channels can often yield significantly higher performance than all channels that include task-irrelevant ones [39]. These special characteristics of EEG signals in the convolutional filters provide an advantageous rationale for the feasibility of deep-learning-based minimal-channel selection. Simultaneously, due to this task-dependent regional specificity, the proposed channel-selection technique would not require relatively large amounts of data, as required for a deep-learning approach.

Compared with other channel-selection algorithms (i.e., LDA [14] and ConvNet [29]), the present study showed the highest decoding accuracies using a reasonably lower number of selected channels while maintaining decoding performance statistically equivalent to (or even higher than) that of all channels. It also provided neurophysiologically interpretable accounts for selected channels. For example, compared with LDA and ConvNet, the number of minimally selected channels from the compact CNN in Study 1 was lowest, and its decoding accuracies were highest. The minimal channels of LDA adversely included 17 more channels compared to those of the compact CNN, and the selected channels by LDA were sparsely distributed over the brain (eccentric from the vertex, Cz), which was not neurophysiologically interpretable for auditory P300 paradigm. Regarding Study 2, although the numbers of minimally selected channels were comparable across different channel-selection algorithms, the compact CNN yielded significantly higher decoding accuracies, compared with LDA and ConvNet. Furthermore, the number of selected channels by ConvNet (i.e., six) is higher than that of the compact CNN (i.e., four) in Study 3. The selected channels of ConvNet were not located around the neurophysiologically interpretable regions (i.e., C3 and C4) for MI. The number of minimally selected channels by ConvNet was disadvantageously more than that of the compact CNN in both Studies 1 and 3. Notably, the number of minimally selected channels by LDA profoundly varied, depending on the type of BMI paradigms (e.g., 19 for Study 1 and three for Study 3). Taken together, these observations reflect that the compact CNN is more robust and stable across different BMI paradigms.

Our observation of these task-relevant and neurophysiologically significant channels as a minimum-channel set for decoding users' intention is advantageous for portable EEG-based BMIs. For example, installation and computing time can be decreased by reducing the number of EEG channels for BMIs. This is vital for the real-time application of the BMI technology. A reduced number of channels enables BMI users to be more comfortable and may substantially reduce the financial cost of setup from the viewpoint of their practical applications. For example, a type of EEG–BMI could be efficiently and conveniently controlled by a headband-shaped EEG acquisition device (or embedded in a head-mounted device), which would have only selected channels for BMI control. In future studies, we will attempt to apply this approach to other types of BMIs to validate its generality and expandability. In several domain applications, CNNs have been very successful, because they can learn the most task-relevant features. However, their performance principally depends on their architectures and learnable hyper-parameters. Thus, the parameter optimization for channel selection in the present study is newly suggested to the EEG-based BMI field. Nonetheless, we will continuously attempt to further improve the efficiency of the minimum-channel selection in future studies. For example, because fast sparse CNNs for enhanced efficiency and higher accuracy have recently been proposed with increased accuracies and a reduced number of parameters [74], the number of learnable parameters in this study could be further reduced by replacing the standard convolutional layer with depthwise and separable convolutional layers, resulting in higher accuracies. From a practical viewpoint, the proposed method could also be improved using transfer-learning methods [75], [76]. Using transfer learning,

a pretrained model with selected minimum channels could be directly applied to new subjects without a calibration phase, which is typically time consuming for EEG-based BMIs. Additionally, as the graph theory perspective might open new insight into classical classification-based approaches, the probabilistic graphical models in future studies could enhance the performance of channel selection because these types of models can learn the relationship between the activation areas and BMI-tasks. For example, a common Bayesian network was proposed to discriminate multiclass MI EEG signals [77]. Altogether, the deep-learning-based channel selection would improve processing speed and accuracy, which benefits the feasibility and portability of EEG-based BMI technology.

The present study has some implicational constraints that are worth mentioning. First, based on our current observations under a limited number of given tasks, we could not address the causal relationships between minimal and optimal channels. Furthermore, checking whether or not the minimum number of channels was actually optimal across a variety of experimental paradigms or tasks was out of the scope of the present study. This presents a potential topic for our future studies. However, it would be advantageous if the minimum number of channels was to yield optimal decoding performance at least in a certain type of experimental paradigm or task. Applying a minimized number of EEG channels would enable EEG-based BMIs to be compact and more portable, which is surely advantageous to their practical use. In the present study, using three representative cognitive tasks, we provided a preliminary investigation of their relationships. Thus, apart from the consistent observations across different experimental paradigms, the relationship between minimal and optimal channels in the presence of limited data should be carefully considered when interpreting our results based on a deep-learning-based decoder.

There are also crucial limitations in the deep-learning-based decoder, which constrains the ways we can collect sufficient datasets. Although the number of EEG data is usually insufficient for a deep-learning approach and thus impedes EEG-based deep learning, there are growing efforts to overcome this constraint by utilizing several data-augmentation methods, such as generative adversarial networks [32], [33]; empirical mode decomposition [34], [35]; or random transformations (e.g., jittering, rotation, scaling, or frequency warping) [36]. However, in the present study, the compact-CNN approach of EEGNet was employed, which could be optimized to a relatively small dataset of EEG signals without the need for vast data augmentation [31]. Furthermore, because the most activated brain regions are principally based on the neurophysiologically regional characteristics of specific cognitive tasks (i.e., task-dependent regional specificity), the proposed method for channel selection would not require large amounts of data to reach the optimal decision of minimal channels. Moreover, because the present study employed only three layers of a neural network, which provide a shallow network rather than the highly complicated architecture (or higher data dimensionality) typically observed in the deep-learning studies of CNNs, such as GoogleNet [78], ReNet [79], Xception [80], and EfficientNet [81], the present study did not use big data, which is typically required for deep networks. In summary, the dataset size would not be highly problematic in this study. Despite these exemptions, the limited dataset size should be carefully considered when interpreting our results.

V. Conclusion

In this study, we investigated an automatic selection algorithm for three representative types (auditory versus visual versus motor) of BMI paradigms. Irrespective of the types of task paradigms and input EEG features, an optimized minimal EEG channel set can be selected using a deep-learning model customized to the small dataset of EEG signals. This compact CNN method produces considerably reliable decoding accuracies using minimally selected channel subsets, which are statistically equivalent to (or even higher than) those using all channels. Furthermore, the channel-selection approach proposed by a compact CNN classifier yields significantly higher decoding accuracy than that by a linear classifier typically used for the SSVEP dataset. The proposed deep-learning method of EEG channel selection innovatively confirms its capability of extracting and harnessing interpretable EEG features generally corresponding to known neurophysiological observations (Fig. 6), providing compelling evidence favorable for the current hot issue of explainable artificial intelligence. Such neurophysiologically corresponding results of channel selection are consistently observed, irrespective of the type of experimental modalities (P300: auditory; SSVEP: visual; and MI: motor modality). By comparison, previous channel-selection studies have typically dealt with single modalities [7], [8], [82]. Moreover, the proposed approach does not rely on specific input EEG features, so it can be based on either the time domain (i.e., ERP or evoked potential), frequency domain (e.g., SSVEP flickering frequency), or time-frequency domain (e.g., event-related (de)synchronization in MI). Therefore, the proposed technique is generalized more than previously proposed channel-selection methods. It would be advantageous if the algorithm were designed to take the raw data without prior knowledge of the important features, determine the minimum channels, and achieve comparable classification accuracies. However, due to the current limitations in EEG-based BMI technology, this could be the focus of one of our future studies on the fewest channel selection in BMIs. As big data, typically advantageous for a deep-learning technique, are progressively collected for each cognitive task, the generality of the proposed deep-learning approach to the various repertoires of BMI tasks would be reinforced and further stabilized. In short, these notions reflect several potential advantages of the present deep-learning-based approach over existing channel-selection methods for the EEG-based BMIs.

The proposed deep-learning-based automatic channel-selection system could become a potential future technology for portable EEG-based BMIs. The fewer the number of channels used, the smaller the amount of data will be required. Consequently, the processing speed and portability of the EEG-based BMIs would be improved. This novel approach could, in principle, use an adaptive-learning approach on various BMI paradigms: a technique that may ultimately become

useful for future channel-selection techniques applicable to state-of-the-art BMI software platforms, such as Web-based BMIs [83]. This study demonstrates that a deep-learning technique on EEG signals can provide promising and potent task-relevant EEG features that enable practical and ubiquitous channel-selection applications for the BMI technology.

ACKNOWLEDGMENT

The authors are grateful to Seong-Eun Kim for his valuable comments on this study and to Ji-Wan Kim for his assistance during EEG acquisition.

REFERENCES

[1] D. Tan and A. Nijholt,"Brain–computer interfaces and human-computer interaction," in *Brain-Computer Interfaces*. London, U.K.: Springer, 2010, pp. 3–19.
[2] B. Allison, B. Graimann, and G. Pfurtscheller, *Brain–Computer Interfaces: Revolutionizing Human–Computer Interaction*. Heidelberg, Germany: Springer, 2010.
[3] 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.
[4] B.-K. Min, R. Chavarriaga, and J. D. R. Millan, "Harnessing prefrontal cognitive signals for brain–machine interfaces," *Trends Biotechnol.*, vol. 35, no. 7, pp. 585–597, Jul. 2017, doi: 10.1016/j.tibtech.2017.03.008.
[5] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review," *Sensors*, vol. 12, no. 2, pp. 1211–1279, 2012.
[6] M. Arvaneh, C. Guan, K. K. Ang, and C. Quek, "Optimizing the channel selection and classification accuracy in EEG-based BCI," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 6, pp. 1865–1873, Jun. 2011.
[7] T. Alotaiby, F. E. A. El-Samie, S. A. Alshebeili, and I. Ahmad,"A review of channel selection algorithms for EEG signal processing," *EURASIP J. Adv. Signal Process.*, vol. 2015, no. 1, p. 66, 2015.
[8] M. Schröder *et al.*, "Robust EEG channel selection across subjects for brain-computer interfaces," *EURASIP J. Appl. Signal Process.*, vol. 2005, pp. 3103–3112, Nov. 2005.
[9] F. Qi *et al.*,"Spatiotemporal-filtering-based channel selection for single-trial EEG classification," *IEEE Trans. Cybern.*, vol. 51, no. 2, pp. 558–567, Feb. 2021.
[10] B. Chakraborty, L. Ghosh, and A. Konar, "Optimal selection of EEG electrodes using interval type-2 fuzzy-logic-based semiseparating signaling game," *IEEE Trans. Cybern.*, early access, Feb. 20, 2020, doi: 10.1109/TCYB.2020.2968625.
[11] Y. Wang, S. Gao, and X. Gao, "Common spatial pattern method for channel selelction in motor imagery based brain-computer interface," in *Proc. IEEE 27th Annu. Conf. Eng. Med. Biol.*, Shanghai, China, 2006, pp. 5392–5395.
[12] J. Meng, G. Liu, G. Huang, and X. Zhu, "Automated selecting subset of channels based on CSP in motor imagery brain-computer interface system," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Guilin, China, 2009, pp. 2290–2294.
[13] T. N. Lal *et al.*, "Support vector channel selection in BCI," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1003–1010, Jun. 2004.
[14] K. A. Colwell, D. B. Ryan, C. S. Throckmorton, E. W. Sellers, and L. M. Collins, "Channel selection methods for the P300 speller," *J. Neurosci. Meth.*, vol. 232, pp. 6–15, Jul. 2014.
[15] A. Barachant and S. Bonnet, "Channel selection procedure using Riemannian distance for BCI applications," in *Proc. 5th Int. IEEE/EMBS Conf. Neural Eng.*, Cancun, Mexico, 2011, pp. 348–351.
[16] H. Cecotti *et al.*, "A robust sensor-selection method for P300 brain–computer interfaces," *J. Neural Eng.*, vol. 8, no. 1, 2011, Art. no. 016001.
[17] L. Q. Thang and C. Temiyasathit, "Investigation of regularization theory for four-class classification in brain–computer interface," in *Future Data and Security Engineering*. Cham, Switzerland: Springer, 2014, pp. 275–285.
[18] J. Farquhar and N. J. Hill, "Interactions between pre-processing and classification methods for event-related-potential classification: Best-practice guidelines for brain–computer interfacing," *Neuroinformatics*, vol. 11, no. 2, pp. 175–192, 2013.
[19] I. Sturm, S. Lapuschkin, W. Samek, and K.-R. Müller, "Interpretable deep neural networks for single-trial EEG classification," *J. Neurosci. Meth.*, vol. 274, pp. 141–145, Dec. 2016.
[20] Z. Mao, "Deep learning for rapid serial visual presentation event from electroencephalography signal," Ph.D. dissertation, Dept. Elect. Comput. Eng., Univ. Texas San Antonio, San Antonio, TX, USA, 2016.
[21] D. Mzurikwao *et al.*, "A channel selection approach based on convolutional neural network for multi-channel EEG motor imagery decoding," in *Proc. IEEE 2nd Int. Conf. Artif. Intell. Knowl. Eng. (AIKE)*, Sardinia, Italy, 2019, pp. 195–202.
[22] N.-S. Kwak, K.-R. Müller, and S.-W. Lee, "A convolutional neural network for steady state visual evoked potential classification under ambulatory environment," *PLoS ONE*, vol. 12, no. 2, 2017, Art. no. e0172578.
[23] F. Lotte *et al.*, "A review of classification algorithms for EEG-based brain–computer interfaces: A 10 year update," *J. Neural Eng.*, vol. 15, no. 3, 2018, Art. no. 031005.
[24] Y. LeCun *et al.*, "Backpropagation applied to handwritten zip code recognition," *Neural Comput.*, vol. 1, no. 4, pp. 541–551, Dec. 1989.
[25] Y.-L. Boureau, F. Bach, Y. LeCun, and J. Ponce, "Learning mid-level features for recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, San Francisco, CA, USA, 2010, pp. 2559–2566.
[26] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. 14th Int. Conf. Artif. Intell. Stat.*, 2011, pp. 315–323.
[27] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals," *Comput. Biol. Med.*, vol. 100, pp. 270–278, Sep. 2018.
[28] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert, "Deep learning-based electroencephalography analysis: A systematic review," *J. Neural Eng.*, vol. 16, no. 5, 2019, Art. no. 051001.
[29] R. T. Schirrmeister *et al.*, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Hum. Brain Map.*, vol. 38, no. 11, pp. 5391–5420, 2017.
[30] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
[31] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, 2018, Art. no. 056013.
[32] S. Haradal, H. Hayashi, and S. Uchida, "Biosignal data augmentation based on generative adversarial networks," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, 2018, pp. 368–371.
[33] G. Ramponi, P. Protopapas, M. Brambilla, and R. Janssen, "T-CGAN: Conditional generative adversarial network for data augmentation in noisy time series with irregular sampling," 2018. [Online]. Available: arXiv:1811.08295.
[34] 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.
[35] J. Dinarès-Ferran, R. Ortner, C. Guger, and J. Solé-Casals, "A new method to generate artificial frames using the empirical mode decomposition for an EEG-based motor imagery BCI," *Front. Neurosci.*, vol. 12, p. 308, May 2018.
[36] D. Freer and G.-Z. Yang, "Data augmentation for self-paced motor imagery classification with C-LSTM," *J. Neural Eng.*, vol. 17, no. 1, 2020, Art. no. 016041.
[37] A. Rakshit, A. Konar, and A. K. Nagar, "A hybrid brain-computer interface for closed-loop position control of a robot arm," *IEEE/CAA J. Autom. Sinica*, vol. 7, no. 5, pp. 1344–1360, Sep. 2020.
[38] S. K. Hong, S. Park, M.-H. Ahn, and B.-K. Min, "Top-down and bottom-up neurodynamic evidence in patients with tinnitus," *Hear. Res.*, vol. 342, pp. 86–100, Dec. 2016, doi: 10.1016/j.heares.2016.10.002.
[39] B.-K. Min, S. Dähne, M.-H. Ahn, Y.-K. Noh, and K.-R. Müller, "Decoding of top-down cognitive processing for SSVEP-controlled BMI," *Sci. Rep.*, vol. 6, Nov. 2016, Art. no. 36267.
[40] H. Cho, M. Ahn, S. Ahn, M. Kwon, and S. C. Jun, "EEG datasets for motor imagery brain–computer interface," *GigaScience*, vol. 6, no. 7, pp. 1–8, 2017.
[41] 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, 2020.
[42] P. Pudil, J. Novovičová, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognit. Lett.*, vol. 15, no. 11, pp. 1119–1125, 1994.
[43] D. Lakens, "Equivalence tests: A practical primer for *t* tests, correlations, and meta-analyses," *Soc. Psychol. Pers. Sci.*, vol. 8, no. 4, pp. 355–362, May 2017, doi: 10.1177/1948550617697177.
[44] C. C. Duncan-Johnson and E. Donchin, "On quantifying surprise: The variation of event-related potentials with subjective probability," *Psychophysiology*, vol. 14, no. 5, pp. 456–467, 1977.
[45] J. Polich, "Habituation of P300 from auditory stimuli," *Psychobiology*, vol. 17, no. 1, pp. 19–28, 1989.

[46] R. Verleger and P. Berg, "The waltzing oddball," *Psychophysiology*, vol. 28, no. 4, pp. 468–477, 1991.
[47] J. Polich, "Cognitive brain potentials," *Current Directions Psychol. Sci.*, vol. 2, no. 6, pp. 175–179, 1993.
[48] E. Donchin and M. G. H. Coles, "Is the P300 component a manifestation of context updating?" *Behav. Brain Sci.*, vol. 11, no. 3, pp. 357–427, 1988.
[49] R. Johnson, "The amplitude of the P300 component of the event-related potential: Review and synthesis," in *Advances in Psychophysiology*, vol. 3. Greenwich, CT, USA: JAI Press, 1988, pp. 69–137.
[50] T. W. Picton, "The P300 wave of the human event-related potential," *J. Clin. Neurophysiol. Official Publ. Amer. Electroencephalograph. Soc.*, vol. 9, no. 4, pp. 456–479, Oct. 1992. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/1464675
[51] J. Katayama and J. Polich, "Auditory and visual P300 topography from a 3 stimulus paradigm," *Clin. Neurophysiol. Official J. Int. Federation Clin. Neurophysiol.*, vol. 110, no. 3, pp. 463–468, Mar. 1999. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/10363770
[52] X. Gao, D. Xu, M. Cheng, and S. Gao, "A BCI-based environmental controller for the motion-disabled," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 2, pp. 137–140, Jun. 2003, doi: 10.1109/TNSRE.2003.814449.
[53] M. A. Pastor, J. Artieda, J. Arbizu, M. Valencia, and J. C. Masdeu, "Human cerebral activation during steady-state visual-evoked responses," *J. Neurosci. Official J. Soc. Neurosci.*, vol. 23, no. 37, pp. 11621–11627, Dec. 2003. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/14684864
[54] S. Makeig, T. P. Jung, A. J. Bell, D. Ghahremani, and T. J. Sejnowski, "Blind separation of auditory event-related brain responses into independent components," *Proc. Nat. Acad. Sci. USA*, vol. 94, no. 20, pp. 10979–10984, Sep. 1997. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/9380745
[55] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 4, pp. 441–446, Dec. 2000.
[56] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Muller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 41–56, 2008.
[57] G. H. Klem, "Artifacts," in *Current Practice of Clinical Electroencephalography*, J. S. Ebersole and T. A. Pedley, Eds. Philadelphia, PA, USA: Lippincott Williams Wilkins, 2003, pp. 271–287.
[58] I. I. Goncharova, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "EMG contamination of EEG: Spectral and topographical characteristics," *Clin. Neurophysiol.*, vol. 114, no. 9, pp. 1580–1593, 2003.
[59] S. Haufe *et al.*, "On the interpretation of weight vectors of linear models in multivariate neuroimaging," *Neuroimage*, vol. 87, pp. 96–110, Feb. 2014.
[60] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1026–1034.
[61] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014. [Online]. Available: arXiv:1412.6980.
[62] S. Lemm, B. Blankertz, T. Dickhaus, and K.-R. Müller, "Introduction to machine learning for brain imaging," *NeuroImage*, vol. 56, no. 2, pp. 387–399, 2011.
[63] F. Chollet. (2019). *Keras: The Python Deep Learning Library. 2015*. [Online]. Available: https://Keras.io
[64] M. Abadi *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," 2016. [Online]. Available: arXiv:1603.04467.
[65] D. J. Krusienski *et al.*, "A comparison of classification techniques for the P300 speller," *J. Neural Eng.*, vol. 3, no. 4, pp. 299–305, 2006.
[66] B. Xu *et al.*, "Motor imagery based continuous teleoperation robot control with tactile feedback," *Electronics*, vol. 9, no. 1, p. 174, 2020.
[67] C. S. Kim, J. Sun, D. Liu, Q. Wang, and S. G. Paek, "Removal of ocular artifacts using ICA and adaptive filter for motor imagery-based BCI," *IEEE/CAA J. Autom. Sinica*, early access, Jan. 25, 2017, doi: 10.1109/JAS.2017.7510370.
[68] R. Näätänen, W. Teder, K. Alho, and J. Lavikainen, "Auditory attention and selective input modulation: A topographical ERP study," *Neuroreport*, vol. 3, no. 6, pp. 493–496, Jun. 1992. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/1391755
[69] D. J. McFarland, L. A. Miner, T. M. Vaughan, and J. R. Wolpaw, "Mu and beta rhythm topographies during motor imagery and actual movements," *Brain Topogr.*, vol. 12, no. 3, pp. 177–186, 2000.
[70] G. D. Schott, "Penfield's homunculus: A note on cerebral cartography," *J. Neurol. Neurosurg. Psychiatry*, vol. 56, no. 4, pp. 329–333, 1993.
[71] C. Liu, Y. Fu, J. Yang, X. Xiong, H. Sun, and Z. Yu, "Discrimination of motor imagery patterns by electroencephalogram phase synchronization combined with frequency band energy," *IEEE/CAA J. Autom. Sinica*, vol. 4, no. 3, pp. 551–557, Jul. 2017.
[72] A. Ghaemi, E. Rashedi, A. M. Pourrahimi, M. Kamandar, and F. Rahdari, "Automatic channel selection in EEG signals for classification of left or right hand movement in Brain Computer Interfaces using improved binary gravitation search algorithm," *Biomed. Signal Process. Control*, vol. 33, pp. 109–118, Mar. 2017.
[73] Y. Yang, I. Bloch, S. Chevallier, and J. Wiart, "Subject-specific channel selection using time information for motor imagery brain–computer interfaces," *Cogn. Comput.*, vol. 8, no. 3, pp. 505–518, 2016.
[74] E. Elsen, M. Dukhan, T. Gale, and K. Simonyan, "Fast sparse ConvNets," 2019. [Online]. Available: arXiv:1911.09723.
[75] L. Xie, Z. Deng, P. Xu, K.-S. Choi, and S. Wang, "Generalized hidden-mapping transductive transfer learning for recognition of epileptic electroencephalogram signals," *IEEE Trans. Cybern.*, vol. 49, no. 6, pp. 2200–2214, Jun. 2019.
[76] J. Li, S. Qiu, Y.-Y. Shen, C.-L. Liu, and H. He, "Multisource transfer learning for cross-subject EEG emotion recognition," *IEEE Trans. Cybern.*, vol. 50, no. 7, pp. 3281–3293, Jul. 2020.
[77] L. He, D. Hu, M. Wan, Y. Wen, K. M. Von Deneen, and M. Zhou, "Common Bayesian network for classification of EEG-based multiclass motor imagery BCI," *IEEE Trans. Syst. Man, Cybern., Syst.*, vol. 46, no. 6, pp. 843–854, Jun. 2016.
[78] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, 2015, pp. 1–9.
[79] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Las Vegas, NV, USA, 2016, pp. 770–778.
[80] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Honolulu, HI, USA, 2017, pp. 1251–1258.
[81] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," 2019. [Online]. Available: arXiv:1905.11946.
[82] M. Arvaneh, C. Guan, K. K. Ang, and C. Quek, "Robust EEG channel selection across sessions in brain-computer interface involving stroke patients," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Brisbane, QLD, Australia, 2012, pp. 1–6.
[83] P. Stegman, C. S. Crawford, M. Andujar, A. Nijholt, and J. E. Gilbert, "Brain–computer interface software: A review and discussion," *IEEE Trans. Human-Mach. Syst.*, vol. 50, no. 2, pp. 101–115, Apr. 2020.

Hyun-Seok Kim received the B.S. degree in electronics and electrical engineering from Pusan National University, Busan, South Korea, in 2008, and the Ph.D. degree in bioengineering from Seoul National University, Seoul, South Korea, in 2018.

Since 2018, he has been working as a Postdoctoral Research Fellow with the Institute for Brain and Cognitive Engineering, Korea University, Seoul. His research interests include machine learning, computer-aided brain-disease diagnosis, biosignal processing, and brain–computer interfaces.

Min-Hee Ahn (Student Member, IEEE) received the master's degree in computer engineering from Pusan National University, Busan, South Korea, in 1998, and the Ph.D. degree in brain and cognitive engineering from Korea University, Seoul, South Korea, in 2018, with a thesis on EEG source-level brain–machine interfacing technology.

Since 2019, he has been working as a Research Professor with the Laboratory of Brain and Cognitive Science for Convergence Medicine, College of Medicine, Hallym University, Anyang, South Korea. His research interests include EEG analysis of tinnitus patients, deep learning modeling of dizziness, and tCS-based neuromodulation for the treatment of tinnitus.

Byoung-Kyong Min (Member, IEEE) received the M.S. degree in neurobiology and physiology from Northwestern University, Evanston, IL, USA, in 1998, and the Ph.D. degree in biological psychology from Magdeburg University, Magdeburg, Germany, in 2007.

He currently works as an Associate Professor with the Department of Brain and Cognitive Engineering, Korea University, Seoul, South Korea. He has combined ultrasound sonication (or transcranial current stimulation) with an EEG-based BMI to accomplish a noninvasive human brain-to-brain interface. His research interests include the spectral analysis of brain electrical activity (EEG/MEG) and cognitive BMIs.