

Subject-Independent Brain–Computer Interfaces Based on Deep Convolutional Neural Networks

O-Yeon Kwon^{ID}, Min-Ho Lee^{ID}, Cuntai Guan, *Fellow, IEEE*, and Seong-Whan Lee^{ID}, *Fellow, IEEE*

Abstract—For a brain–computer interface (BCI) system, a calibration procedure is required for each individual user before he/she can use the BCI. This procedure requires approximately 20–30 min to collect enough data to build a reliable decoder. It is, therefore, an interesting topic to build a calibration-free, or subject-independent, BCI. In this article, we construct a large motor imagery (MI)-based electroencephalography (EEG) database and propose a subject-independent framework based on deep convolutional neural networks (CNNs). The database is composed of 54 subjects performing the left- and right-hand MI on two different days, resulting in 21600 trials for the MI task. In our framework, we formulated the discriminative feature representation as a combination of the spectral–spatial input embedding the diversity of the EEG signals, as well as a feature representation learned from the CNN through a fusion technique that integrates a variety of discriminative brain signal patterns. To generate spectral–spatial inputs, we first consider the discriminative frequency bands in an information-theoretic observation model that measures the power of the features in two classes. From discriminative frequency bands, spectral–spatial inputs that include the unique characteristics of brain signal patterns are generated and then transformed into a covariance matrix as the input to the CNN. In the process of feature representations, spectral–spatial inputs are individually trained through the CNN and then combined by a concatenation fusion technique. In this article, we demonstrate that the classification accuracy of our subject-independent (or calibration-free) model outperforms that of subject-dependent models using various methods [common spatial pattern (CSP), common spatio-spectral pattern (CSSP), filter bank CSP (FBCSP), and Bayesian spatio-spectral filter optimization (BSSFO)].

Index Terms—Brain–computer interface (BCI), convolutional neural networks (CNNs), deep learning (DL), electroencephalography (EEG), motor imagery (MI), subject-independent.

Manuscript received July 20, 2018; revised January 6, 2019, May 6, 2019, and August 24, 2019; accepted October 3, 2019. Date of publication November 13, 2019; date of current version October 6, 2020. This work was supported in part by Institute for Information and Communications Technology Planning and Evaluation (IITP) Grant funded by the Korea Government (MSIT) (Development of Intelligent Pattern Recognition Softwares for Ambulatory Brain-Computer Interface) under Grant 2015-0-00185 and (Development of BCI based Brain and Cognitive Computing Technology for Recognizing User’s Intentions using Deep Learning) under Grant 2017-0-00451 and in part by the Samsung Research Funding Center of Samsung Electronics under Project SRFC-TC1603-02. (*Corresponding author: Seong-Whan Lee.*)

O.-Y. Kwon is with the Department of Brain and Cognitive Engineering, Korea University, Seoul 02841, South Korea (e-mail: data0311@gmail.com).

M.-H. Lee is with the Department of Computer Science, Nazarbayev University, Astana 010000, Kazakhstan (e-mail: minho.lee@nu.edu.kz).

C. Guan is with the School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798 (e-mail: ctguan@ntu.edu.sg).

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

Color versions of one or more of the figures in this article are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TNNLS.2019.2946869

I. INTRODUCTION

BRAIN–COMPUTER interface (BCI) is a system that enables a direct communication pathway between a human brain and external devices [1]. BCIs have shown great potential in a variety of clinical applications for communication, control, and rehabilitation [2]–[9]. Furthermore, recent BCI studies have attracted great attention as future technology for the next generation [10]–[15]. Over the course of numerous BCI studies [16]–[21], growing attention has been dedicated to the analysis of electroencephalography (EEG) signals, especially by movement imagination, called motor imagery (MI) [1], [2], [8], [12], [14], [16], [22]. This interest is due to the ability of MI to allow both healthy and disabled people to self-regulate brain signals without an external stimulus.

In BCI systems in general, a user’s brain signals can change over minutes, hours, or days due to the differences in the physiological and psychological characteristics of each individual at each time [22]. Moreover, the spatial origin, the amplitude change, and the variability of the brain signals exhibit subject-specific patterns [23]. If the brain signals are expected to change occasionally, then a decoding method to compensate for the changes in the brain signals is indispensable in the general use of BCIs. However, unfortunately, the process of calibration is an inconvenient and time-consuming task, which requires approximately 20–30 min for a reliable decoder to be built [24], [25]. In the current BCI research, the calibration process is one of the obstacles that prevents the practical use of BCI systems. In conclusion, this problem should be addressed to enable real-world applications of BCIs.

To this end, we construct a large-scale MI EEG database and propose a novel subject-independent framework based on spectral–spatial feature representation with deep convolutional neural networks (CNNs). The main contributions of this article can be summarized as follows.

- 1) We built a large MI-based EEG database with 54 healthy subjects over two sessions. This database is the largest MI BCI dataset thus far ever reported in the literature, and it provides a sufficient number of training samples for deep learning (DL) architectures (CNNs).
- 2) We proposed a new discriminative spectral–spatial input to represent a diversity of brain signal patterns across the subjects and sessions.
- 3) We represent the MI-induced event-related desynchronization (ERD) patterns through the CNN model

by applying a spatial fusion technique that combines spectral–spatial inputs from diverse frequency regions.

The remainder of this article is organized as follows. Section II discusses the related work. In Section III, we briefly introduce our large-scale database. Section IV elaborates the proposed framework. Details of the experimental results and analysis are discussed in Sections V and VI. Finally, Section VII concludes this article.

II. RELATED WORKS

Most conventional MI-based BCI systems were developed based on subject-dependent methods that require calibration time [25]–[28]. Common spatial patterns (CSPs) are one of the most widely used methods in MI-based BCI; CSPs maximize the differences in the variances for the binary classes [23], [25]. Based on CSP methods, advanced algorithms have been proposed, for example, common spatio-spectral pattern (CSSP) [29], filter bank CSP (FBCSP) [28], and Bayesian spatio-spectral filter optimization (BSSFO) [30]. The CSSP method is built on an extension of the CSP method with an embedded time delay [29]. The FBCSP method is based on composing a frequency range with nonoverlapping and static frequency bands (e.g., 4–8 Hz, 8–12 Hz, ...) [28]. The BSSFO method constructs a data-driven discriminative filter bank and bandwidth selection within a Bayesian framework [30]. However, these methods still require the calibration procedure to train the decoder.

In recent years, various strategies to reduce the calibration time in MI-based BCIs have been proposed. The zero-calibration method aims to immediately use the BCI system without the calibration process for a naive BCI user (e.g., without using training data from new users) and is also called the subject-independent method [31]–[35]. The existing zero-calibration methods are usually built on a transfer-based approach that uses preexisting data from other subjects. Ray *et al.* [31] established an MI database, extracted meaningful brain features based on the FBCSP method from a previously created database, and classified them using a linear support vector machine (SVM). Lotte *et al.* [32] used a public database that was composed of nine subjects and designed a user-independent model by considering multiresolution frequency decomposition to extract the EEG features, following the idea of the FBCSP method. Fazli *et al.* [33] proposed a zero-calibration ensemble method that extracted a temporal–spatial filter and built an ensemble model on the basis of linear discriminant analysis (LDA). In addition, studies that utilize minimum calibration data have been conducted [34]–[39]. However, these recent zero-training strategies in the literature have primarily concentrated on the linear properties of the given training samples and endeavored to extract EEG features based on conventional subject-dependent methods. At present, in the field of BCI, there is a great interest in investigating machine learning technology and applying it to the analysis of EEG decoding. One prominent example of such advances is the utilization of DL, which relates to other technology trends, such as big data and artificial intelligence.

DL has been shown to be a great success in computer vision, natural language processing, and many other fields [40]–[46].

As DL methods enabled us to effectively exploit distributed and hierarchical features through multiple layers of nonlinear information processing [40]–[44], it is expected to produce higher performance than traditional classification methods (e.g., SVM) [45]–[47]. To date, DL has been applied to MI-based BCIs and contributed to performance improvements [48]–[51], but they are all still limited to subject-dependent BCI.

There are two key issues to be addressed in utilizing DL approaches to develop a subject-independent BCI. First, there is no large-scale MI database with a large number of subjects and sessions available. This is because establishing a large database for EEG-based BCI requires a tremendous effort. Second, there were scarce studies on subject-independent BCIs based on DL, which requires strategies for extracting discriminative brain features from a large-scale database. Conventional methods worked well in the subject-dependent case but fared poorly for the subject-independent case, because brain signals from different subjects are highly variable, discriminative, and semantic [22], [24].

In this article, we propose a novel framework of spectral–spatial feature representation using CNNs from a large-scale MI EEG database. To the best of our knowledge, this is the first study to apply a method that has a sufficient number of EEG training samples and a feature extraction method that is fit to subject-independent BCI based on DL.

III. DATA ACQUISITION AND EXPERIMENTAL SETUP

A. Participants

The main purpose of our MI database was to capture large variations between the subjects and the sessions. In the experiment, 54 healthy subjects (aged 24–35: 29 male and 25 female) participated in experiments of two sessions. Among them, 38 subjects were naive BCI users, and the others had previous BCI experiment. The interval between the sessions was between one and two weeks. None of the subjects had a history of neurological, psychiatric, or any other pertinent disease.

At the beginning of the experiment, the subjects were seated comfortably in a chair with armrests 60 (± 5) cm from a 20-in LCD monitor (refresh rate: 60 Hz and resolution: 1600 \times 1200). During the experiment, participants were asked to relax and minimize their eye and muscle movements. Additionally, we provided a monetary incentive for the participants, approximately 80 dollars for a single experiment. This study was approved by the Institutional Review Board of Korea University (approval 1040548-KU-IRB-16-159-A-2), and all subjects agreed to participate in this experiment with written consent.

B. EEG Recording

In this experiment, the EEG data were recorded using 62 Ag/AgCl EEG electrodes and four EMG electrodes, with a sampling rate of 1000 Hz. BrainAmp (Brain Products, Munich, Germany) was used as the amplifier in the experiment. The EEG channels were nasion-referenced and grounded to electrode Fpz. The EMG channels were attached at the flexor and extensor of the left and right forearms. Each flexor digitorum

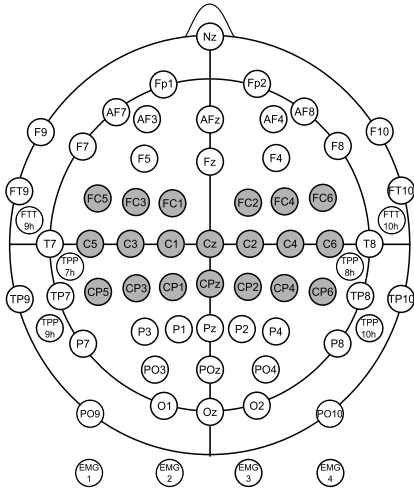


Fig. 1. EEG electrode placement of the international 10–20 system and the EMG placement (62 EEG and 4 EMG recording channels), which indicates the locations of the electrodes. The EEG electrodes shown in gray were used for the performance validation.

profundus muscle with the olecranon was used as an EMG electrode reference. The EEG/EMG electrode configuration and indexing numbers are shown in Fig. 1. The impedances of the EEG electrodes were maintained below 10 k Ω during the entire experiment and checked before the start of each phase. The EMG channels attached to collect the subject’s physical movements were excluded because they were out of scope in this article.

Basically, an antialiasing filter was applied before digitizing the EEG signals. Then, recorded EEG signals were downsampled to 100 Hz by the Nyquist theory for our interesting frequency ranges that are below 40 Hz. For the performance validation, 20 electrodes from the motor cortex region were selected (FC-5/3/1/2/4/6, C-5/3/1/z/2/4/6, and CP-5/3/1/z/2/4/6), as shown in gray circles in Fig. 1.

C. MI Paradigm

In the MI paradigm, all subjects were instructed to perform two-class MI tasks (movement imagination of the left or right hand; hand open and close) according to a visual cue (a left or right arrow) on the monitor. In the experiment, each trial started with a black fixation cross for a preparation of the MI task for 3 s at the center of the monitor. Afterward, the subject was instructed to perform the MI task for 4 s when the left or right arrow appeared on the monitor. After each MI task, the screen remained blank for 6 s (± 1.5). These entire procedures were based on the general BCI experimental setting [1].

The whole experiment is composed of an offline (training) and an online (testing) phase. As the EEG data were acquired in the offline phase, a projection matrix of CSP and a classification model based on CSP and LDA were obtained. Those CSP and LDA parameters were used for the analysis in the online phase. In the online phase, the real-time EEG data for 1.5 s were obtained and filtered with the same frequency band used in the offline phase. Afterward, the projection matrix of CSP from the offline phase was applied to the online EEG

data, and then, the logarithm variances were taken as features. Then, LDA parameters were applied to those features, and finally, the output was presented in the form of a left or right arrow to the user as real-time visual neurofeedback to acquire a high-quality EEG signal through user-adaptive training, which has been used in general BCI research [52]. We provided flexible break times for individual participants between the phases.

For the experimental environments, we used Psychophysics Toolbox Version 3 (PTB-3) and OpenBMI toolbox [53], which can run with MATLAB R2012a+ on Windows OS. The OpenBMI toolbox is a general-purpose software that we have developed since 2012, and it includes experimental codes as well as the modules for data analysis of BCI.¹

IV. SPECTRAL–SPATIAL FEATURE REPRESENTATION (SSFR)

To create a generalized feature representation from a large-scale MI dataset, we considered the following questions:

- 1) how to extract discriminant ERD patterns from continuous EEG data that include a diversity of brain signals in the frequency and spatial domains;
- 2) how to construct $m \times m$ input matrices \mathbf{C} from the EEG epoch \mathbf{X} (time samples \times channels \times trials) that would be suitable to be fed into the CNN architecture;
- 3) how to design a CNN framework that could represent the diversity of spectral–spatial features from the input matrices \mathbf{C} ;

A. Spectral–Spatial Input Generation

Let us denote $\mathbf{X}_n = \{\mathbf{x}_i\}_{i=1}^n \in \mathbb{R}^{T \times M}$ as a set of single-trial EEGs, and let $\mathbf{Y}_n = \{y_i\}_{i=1}^n$ be the matching class labels, where n is the number of EEG trials, T is the number of sample points, and M is the number of channels.

An individual’s EEG signal \mathbf{X}_n was bandpass-filtered with a fifth-order Butterworth filter based on the predefined frequency bands \mathbf{B} . We define \mathbf{B} as a set of predefined frequency bands, $\mathbf{B} = \{\mathbf{b}_k\}_{k=1}^K$ (e.g., [8, 30 Hz], [11, 20 Hz], and so on), where K is the number of frequency bands. The frequency bands included conventional frequency bands in reference to previous findings [23] and empirically extend the additional frequency bands; the details are shown in Fig. 2. Then, the filtered-EEG signals were segmented between 1000 and 3500 ms after stimulus onset (i.e., 2500 data samples \times 20 channels \times 200 trials) for an individual subject.

From a large-scale MI database, all single-trial EEGs $\mathbf{X}_{\text{source}} = \text{concat}\{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N\}$ were placed in sequence from the first subject ($N = 1$) to the last subject ($N = 54$), and the matching class labels $\mathbf{Y}_{\text{source}} = \text{concat}\{\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_N\}$ were sequentially concatenated, where N is the number of training subjects. Note that the target subject (the test subject) is always excluded from the training.

Given a certain frequency band \mathbf{b}_k and concatenated EEG signals $\mathbf{X}_{\text{source}}$, the filtered EEG signals \mathbf{E}_k were

¹<https://github.com/PatternRecognition/OpenBMI>

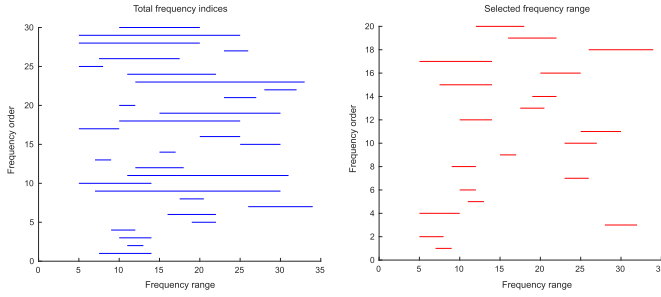


Fig. 2. Entire frequency information used in this article. Left: total number of frequency indices. The x -axis is the frequency range. The y -axis is the order of the frequency. Right: rearranged frequency indices according to the value of mutual information. The bottom ($N = 1$) of the y -axis has the highest mutual information value.

deterministically obtained. The bandpass filtering is applied as follows:

$$\mathbf{E}_k = \mathbf{b}_k \otimes \mathbf{X}_{\text{source}} \quad (1)$$

where \otimes is the bandpass filtering operation.

After the bandpass filtering based on predefined frequency bands, a standard CSP algorithm [25]–[28] is utilized, in which $\mathbf{W}_k \in \mathbb{R}^{U \times U}$ is analytically obtained from \mathbf{E}_k by calculating a generalized eigenvector problem, where U is the number of spatial filters to be obtained from the CSP algorithm.

Before measuring the discriminative power between the classes, a feature vector for each frequency band \mathbf{b}_k is obtained through a matrix multiplication between the filtered EEG signals \mathbf{E}_k and the spatial filters \mathbf{W}_k . The variance of the spatially filtered signals is calculated, and then, the logarithm is computed for the feature

$$\mathbf{V}_k = \log(\text{var}(\mathbf{W}_k^T \cdot \mathbf{E}_k)). \quad (2)$$

Since it is uncertain which frequency bands can compose discriminative brain features, we encode these uncertainties using mutual information over the frequency bands \mathbf{b}_k and features \mathbf{V}_k . Mutual information is a measure of the mutual dependence between two random variables and the uncertainty of the random variables

$$I(\mathbf{V}_k; \mathbf{Y}_{\text{source}}) = H(\mathbf{V}_k) - H(\mathbf{V}_k | \mathbf{Y}_{\text{source}}) \quad (3)$$

where $H(\cdot)$ is the entropy and $H(\cdot | \cdot)$ is the conditional entropy.

Since the value of mutual information is obtained for each frequency band, frequency bands \mathbf{b}_k are rearranged in the descending order (\mathbf{b}_p) from largest to smallest in terms of the value of the mutual information in Fig. 2. This method is used to select informative subsets of the original features [54], and on the basis of our belief, it would help us to discover the frequency band that could contribute to composing discriminative ERD among large-scale MI data. On the basis of the rearranged frequency bands \mathbf{b}_p in Fig. 2, the spectral EEG signals \mathbf{E}_k and the spatial filters \mathbf{W}_k were also rearranged to \mathbf{E}_p and \mathbf{W}_p according to the new frequency order \mathbf{b}_p .

From the spectrally and spatially optimized EEG epoch (\mathbf{W}_p and \mathbf{E}_p), the covariance matrix \mathbf{C}_p is calculated

$$\mathbf{C}_p = \text{cov}(\mathbf{W}_p^T \cdot \mathbf{E}_p) \quad (4)$$

where the input set $\mathbf{C} = \{\mathbf{C}_1, \mathbf{C}_2, \dots, \mathbf{C}_P\} \in \mathbb{R}^{U \times U \times P}$, U is the number of spatial filters, and P is the number of frequency indices.

B. Feature Representation Using CNN

In this article, we utilize the CNN approach to learn representation from a set of spectral–spatial inputs (\mathbf{C}). The CNN model consists of several computational blocks [40]–[43]; here, the proposed framework is composed of three convolutional layers, a concatenation layer for fusion, and a fully connected layer. Given the CNN input \mathbf{C}_p for each frequency band \mathbf{b}_p , the outcome of the CNN layer at the L th layer can be obtained as $f_L(\mathbf{C}_p)$ (see [40]–[43]). To design the CNN, the characteristics of each spectral–spatial input (\mathbf{C}_p) for each frequency band (\mathbf{b}_p) should be taken into consideration. In the CNN layers, the spectral–spatial inputs are individually passed through the CNN model for each frequency band. Each outcome of the convolutional layer is considered as an input into the subsequent computational block. From the diverse frequency bands (\mathbf{b}_p), each convolutional feature is extracted by the CNN, as shown in Fig. 3

$$\mathbf{G}_p^L = f_L(\mathbf{C}_p) \quad (5)$$

where \mathbf{G}_p^L is the outcome of \mathbf{C}_p at the L th convolutional layer, $f(\cdot)$ denotes the CNN model, and $f_L(\cdot)$ composes L convolutional layers of $f(\cdot)$, $f_L(\cdot) = f(\cdot; \Pi^1, \Pi^2, \dots, \Pi^L)$, where Π is the weight matrix of each layer.

The whole CNN framework is trained to reduce the loss function by stochastic gradient descent [40]–[43]. Here, the cross-entropy loss function is used, and feature maps are processed by a rectified linear (ReLU) function.

C. Feature Fusion and Classification

In this section, a feature fusion process is presented by integrating discriminant spectral–spatial inputs. Our approach starts by utilizing multiple frequency bands \mathbf{b}_p , as shown in Fig. 2. Each frequency segment in \mathbf{b}_p represents the diversity of ERD patterns that contain the inherent information. Our main motivation is to integrate all the segments (\mathbf{b}_p) that contain the discriminant ERD patterns in the brain signals.

For the integration of multiple convolutional features, we encode the convolutional features and transform them into a linear vector representation as the outcome feature for a concatenation fusion layer, where the high-dimensional convolutional feature vectors are transformed into a low-dimensional vector for the fusion process.

Let us denote \mathbf{C}_p as the features on the L th layer, which can be represented as $\mathbf{G}_p^L = \mathbf{S} = \{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_P\} \in \mathbb{R}^D$ of the set \mathbf{P} \mathbf{D} -dimensional convolutional features. Thus, a \mathbf{D} -dimensional vector was obtained in each frequency band (\mathbf{b}_p), and all the individual \mathbf{D} -dimensional vectors were concatenated as follows:

$$\mathbf{S}^{\text{concat}}_{(p \times d)} = a_d, \dots, \mathbf{S}^{\text{concat}}_{(p \times d) - (p-1)} = z_d \quad (6)$$

where $\mathbf{S}^{\text{concat}} \in \mathbb{R}^{(P \times D)}$. Here, P is the number of inputs at the L th layer, \mathbf{D} is the dimension of the first fully connected layer of the CNN, and $a = \mathbf{s}_1$, $z = \mathbf{s}_p$.

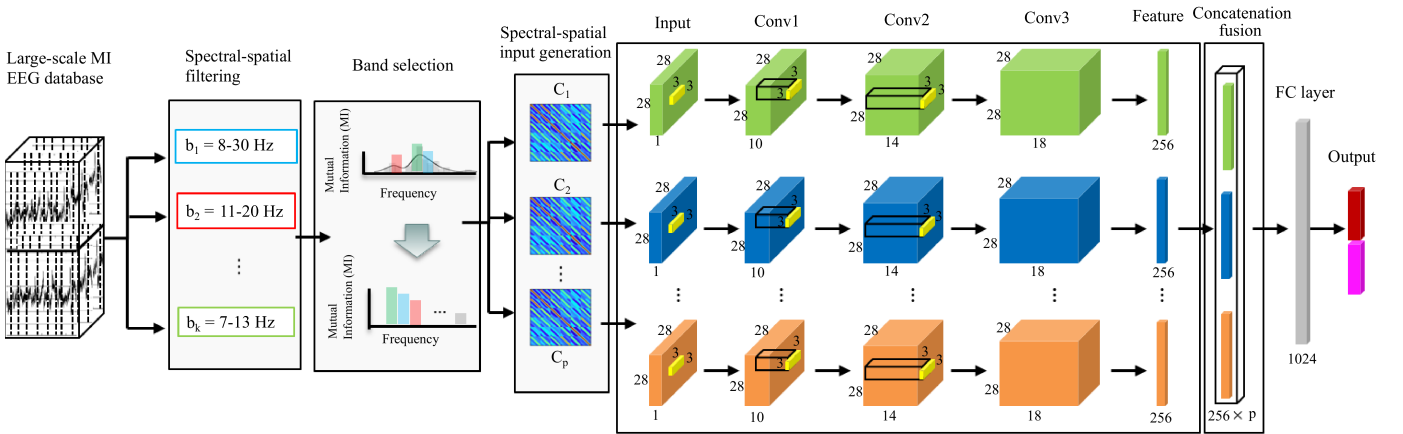


Fig. 3. Proposed subject-independent framework based on deep CNNs.

The concatenated fusion layer \mathbf{S}^{concat} , which is the output of the algorithm, was connected with a fully connected layer, as shown in Fig. 3. Finally, the value of the two outputs that indicate left- or right-hand MI was obtained from the soft-max classifier.

D. Parameter Setting

In the training session, 53 subjects (i.e., $N = 1, \dots, 53$) over two sessions (a total number of 21 200 trial samples) were used to train the CNN. In the 30 predefined frequency bands (i.e., $K = 1, \dots, 30$) in Table I, 20 rearranged frequency bands (i.e., $P = 1, \dots, 20$) were selected in this article. Therefore, 424 000 samples (20 frequency bands \times 53 subjects \times 200 trials \times 2 sessions) were utilized as inputs for the CNN architecture.

Specifically, all spatial filters of the CSP ($U = 10$) were used in this article. Based on (4), all EEG epochs were transformed into covariance matrices ($M \times M$; here, $M = 20$), and zero padding was applied ($\hat{M} \times \hat{M}$; here, $\hat{M} = 28$). Thus, the $\hat{M} \times \hat{M}$ matrix was used for the input of the CNN architecture. The dimension of the first fully connected layer neurons' activations of the CNN \mathbf{D} was 256, and the number of layers L was 4. The learning rate of η was 0.00001, and a 50% drop-out rate was applied in the training phase. A filter size of 3×3 was used. For the batch size, we used 100 samples for each mini-batch. Adam optimizer was used for the optimization algorithm. All learning parameters were determined by our iterated experiments. The entire experiment was conducted on a desktop computer with a dual-core i7-7700 4.2-GHz processor and 16-GB memory. TensorFlow was basically used to implement the algorithms (e.g., concatenate); see Algorithm 1.

E. Performance Evaluation

For the performance comparison, we evaluated the decoding accuracy of the proposed method by comparing the previous subject-independent approaches [31], [32] as well as subject-dependent approaches [23], [28]–[30].

Specifically, a fused model [31], an MR FBCSP (multiresolution FBCSPs) method [32], and a pooled CSP method [32]

were implemented as subject-independent approaches. For the fused model, the EEG signals were decomposed using six frequency bands, as in [31]. The MR FBCSP method also used decomposition with three sub-decompositions of different resolutions, which included 20 different features [32]. Additionally, we evaluated the basic ensemble method (Pooled CSP [32]), which concatenates all the given training samples and obtains the CSP filters from the concatenated samples.

Additionally, CSP [23], CSSP [29], FBCSP [28], and BSSFO [30] were evaluated as subject-specific methods for comparison. For the analysis of the CSP algorithm, a widely used frequency band (8–30 Hz) was utilized to obtain the spatial patterns [23]. The CSSP method is implemented utilizing time delay embedding from 1 to 15 [29]. The FBCSP method is decomposed using nine frequency bands, as in [28]. In BSSFO, 30 particles were used for the analysis and were iterated 10 times to achieve robust classification results [30]. For all the subject-specific methods, bandpass filtering was applied with a fifth-order Butterworth filter at the frequency band. With respect to spatial filtering, the two highest and two lowest spatial patterns were used for the analysis.

For the classifier (except for in the proposed method), we used linear classifiers (LDA) that are specified by the discriminant functions for a binary classification problem. This is because LDA not only achieved the highest accuracy in the subject-independent environment previously [32] but also is widely utilized in the subject-dependent environment [37].

Fig. 4 shows an example of how we utilized the training and testing samples according to the method of analysis. For the subject-dependent validation, the offline data from session2 (the black box) was used to derive the classifier parameters, and the online data from session2 (the gray box) was used for the performance validation using pretrained parameters. For the subject-independent validation, we utilized a leave-one-subject-out cross-validation (LOSO-CV) procedure [31], [32]. For example, assume that subject1 is the target subject (the test subject). Except for the target subject, all the training samples (the black boxes) were used to train the classifier, and then, the online data from session2 for the target subject (the gray box) were used for the performance evaluation. The same test data were used in both environments.

Algorithm 1 Spectral–Spatial Feature Representation Algorithm**Input:** A set of training samples from a large-scale MI database.

- $\mathbf{x} \in \mathbb{R}^{T \times M}$: EEG data with T sample points and M channels.
- $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^n$: an individual's single trial in an EEG epoch, where n is the total number of trials.
- $\mathbf{Y} = \{y_i\}_{i=1}^n$: class labels, where $y_i \in \{+1, -1\}$.
- K = number of spectral filter bands.
- P = number of selected spectral filter bands.
- u = half the number of spatial patterns in a spatial pattern learning algorithm.

Output: Spectral-spatial convolutional features \mathbf{S} .

- $\mathbf{S} \in \mathbb{R}^{1 \times (D \times P)}$: D and P are the dimension of the first fully-connected layer and the number of inputs, respectively.

Procedures: $\mathbf{X}_{concat} = \text{concat}\{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N\}$ ▷ Concatenate all training samples: N is the number of training subjects $\mathbf{Y}_{concat} = \text{concat}\{\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_N\}$ **for** $k = 1$ **to** K **do**| $\mathbf{E}_k = \mathbf{b}_k \otimes \mathbf{X}_{concat}$ ▷ \mathbf{b}_k is the predefined bandpass filter, $\mathbf{E}_k = \{\mathbf{e}_k^i\}_{i=1}^{(N \times n)}$ | Solve $\mathbf{W}_k^T (\Sigma^{(+)} + \Sigma^{(-)}) \mathbf{W}_k = I$

▷ Perform a CSP algorithm

| $\hat{\mathbf{W}}_k$ = the first u and the last u column vectors in \mathbf{W}_k .| **for** $i = 1$ **to** $(N \times n)$ **do**| | $\mathbf{v}_k^i = \log[\text{var}(\hat{\mathbf{W}}_k^T \cdot \mathbf{e}_k^i)]$ ▷ Refer to Section III-Eq.2, $\mathbf{V}_k = \{\mathbf{v}_k^i\}_{i=1}^{(N \times n)}$ | **end**| $I(\mathbf{V}_k; \mathbf{Y}_{concat}) = H(\mathbf{V}_k) - H(\mathbf{V}_k | \mathbf{Y}_{concat})$

▷ Refer to Section III-Eq.3

endRearrange the frequency bands \mathbf{b}_p from the largest MI value to smallest MI value.**for** $p = 1$ **to** P **do**| Rearrange the spectral filters and spatial filters ($\mathbf{E}_p, \mathbf{W}_p$).| **for** $i = 1$ **to** $(N \times n)$ **do**| | $\mathbf{c}_p^i = \text{cov}(\hat{\mathbf{W}}_p^T \cdot \mathbf{e}_p^i)$ ▷ Spectral-spatial input generation, $\mathbf{C}_p = \{\mathbf{c}_p^i\}_{i=1}^{(N \times n)}$ | **end**| $\mathbf{G}_p = f(\mathbf{C}_p)$

▷ Refer to Section III-Eq.5

end $\mathbf{S} = \text{concat}\{\mathbf{G}_1, \mathbf{G}_2, \dots, \mathbf{G}_P\}$

▷ Feature fusion process: refer to Section III-Eq.6

TABLE I
FREQUENCY BANDS' INFORMATION

All 30 frequency bands	[[7.5 ;14] [11 ;13] [10 ;14] [9 ;12] [19 ;22] [16 ;22] [26 ;34] [17.5 ;20.5] [7 ;30] [5 ;14] [11 ;31] [12 ;18] [7 ;9] [15 ;17] [25 ;30] [20 ;25] [5 ;10] [10 ;25] [15 ;30] [10 ;12] [23 ;27] [28 ;32] [12 ;33] [11 ;22] [5 ;8] [7.5 ;17.5] [23 ;26] [5 ;20] [5 ;25] [10 ;20]]
------------------------	---

V. RESULTS

Table II shows the averaged decoding accuracies across all subjects for individual methods. The decoding accuracies for the subject-independent methods are 65.65% (± 16.11), 67.37% (± 16.01), 68.59% (± 15.28), and 74.15% (± 15.83) for the pooled CSP, fused model, MR-FBCSP, and the proposed method, respectively. In the multiple comparison test of the subject-independent methods and the proposed method, the result of an ANOVA test with Bonferroni was [$F_{(3,212)} = 2.9184, p = 0.0351$]. The results of a paired t -test between the subject-independent methods and the proposed method were $p < 0.001$. In case of the subject-dependent methods, the decoding accuracies are 68.57% (± 17.57), 69.68% (± 18.53), 70.59% (± 18.56), and 71.02% (± 18.83) for CSP,

CSSP, FBCSP, and BSSFO, respectively. In the multiple comparison test of the subject-dependent methods and the proposed method, the result of the ANOVA test with Bonferroni was [$F_{(4,265)} = 0.7373, p = 0.5672$]. The results of a paired t -test between the subject-dependent methods and the proposed method were $p < 0.05$.

Tables III and IV show the performance changes of the proposed method according to the kernel size and the number of feature maps, as shown in Fig. 3. The decoding accuracies with the change in kernel size are 73.98% (± 15.57), 74.15% (± 15.83), and 73.72% (± 16.06) for 2×2 , 3×3 , and 5×5 kernels, respectively. The decoding accuracies according to the number of feature maps are 73.51% (± 16.33), 73.90% (± 16.65), and 74.15% (± 15.83) when the first feature map was 5, 8, and 10, respectively. Note that other parameters are based on Section IV-D.

Fig. 5 shows the scatter plots that demonstrate the performance comparison for individual subjects. The x -axis is the classification performance of the competing methods, and the y -axis is the performance of the proposed method. The top row depicts the subject-dependent approaches in the order of CSP, FBCSP, and BSSFO, and the bottom row displays the

TABLE II
PERFORMANCE COMPARISON OF SUBJECT-DEPENDENT AND SUBJECT-INDEPENDENT METHODS

Approach	Method	Mean (SD)	Median	Range (Max-Min)
Subject-dependent	CSP [23]	68.57 (17.57)	64.50	58.00 (100.00-42.00)
	CSSP [29]	69.68 (18.53)	63.00	58.00 (100.00-42.00)
	FBCSP [28]	70.59 (18.56)	64.00	55.00 (100.00-45.00)
	BSSFO [30]	71.02 (18.83)	63.50	52.00 (100.00-48.00)
	Proposed method (dependent)	71.32 (15.88)	66.45	53.10 (99.00-45.90)
Subject-independent	Pooled CSP [32]	65.65 (16.11)	58.00	55.00 (100.00-45.00)
	Fused model [31]	67.37 (16.01)	62.50	57.00 (98.00-41.00)
	MR FBCSP [32]	68.59 (15.28)	63.00	49.00 (97.00-48.00)
	Proposed method (independent)	74.15 (15.83)	75.00	60.00 (100.00-40.00)

TABLE III
CLASSIFICATION PERFORMANCE ACCORDING TO THE KERNEL SIZE

Kernel size	Mean (SD)	Median	Range (Max-Min)
2×2	73.98 (15.57)	74.50	53.00 (100.00-47.00)
3×3	74.15 (15.83)	75.00	58.00 (100.00-42.00)
5×5	73.72 (16.06)	76.00	59.00 (100.00-41.00)

TABLE IV
CLASSIFICATION PERFORMANCE ACCORDING TO THE NUMBER OF FEATURE MAPS

Feature maps	Mean (SD)	Median	Range (Max-Min)
5, 10, 15	73.51 (16.33)	76.00	57.00 (99.00-42.00)
8, 16, 20	73.90 (16.65)	74.00	57.00 (99.00-42.00)
10, 14, 18	74.15 (15.83)	75.00	58.00 (100.00-42.00)

subject-independent approaches in the order of pooled-CSP, fused model, and MR FBCSP. The percentages of subjects who showed higher performance in the proposed method than in the other methods are 74.1% (40 of 54, $p \leq 0.001$), 79.6% (43 of 54, $p \leq 0.001$), 77.8% (42 of 54, $p \leq 0.001$), 64.8% (35 of 54, $p \leq 0.001$), 57.4% (31 of 54, $p \leq 0.001$), and 51.9% (28 of 54, $p \leq 0.001$) for pooled-CSP, fused model, MR FBCSP, CSP, FBCSP, and BSSFO, respectively. A paired t -test is used for the statistical analysis. Additionally, the percentages of subjects who achieved 80% classification performance in each method are 22.2% (12 of 54), 29.6% (16 of 54), 25.9% (14 of 54), 33.3% (18 of 54), 37.0% (20 of 54), 37.0% (20 of 54), and 44.0% (24 of 54) for pooled-CSP, fused model, MR FBCSP, CSP, FBCSP, BSSFO, and the proposed method, respectively.

Figs. 6 and 7 show the performance changes by the number of subjects and the frequency bands. Fig. 6 shows the performance changes according to the number of subjects. To avoid the computational cost, the certain test subject (participant $N = 39$) and the number of frequency bands (i.e., \mathbf{b}_p , $p = 1, \dots, 20$) were determined. The x -axis indicates that the number of subjects (N) increased in steps of one from $N = 5$ to $N = 53$. In each increasing step, a certain number of subjects for the training were randomly selected from among all subjects, and the classification performance was validated with the test subject. Note that the composition of a certain number of subjects was randomly changed over five iterations, and the classification performance was obtained from the average of five iteration results. For example, assume that we have ten subjects (i.e., $N = 1, \dots, 10$). Ten subjects were randomly selected among all subjects for each iteration.

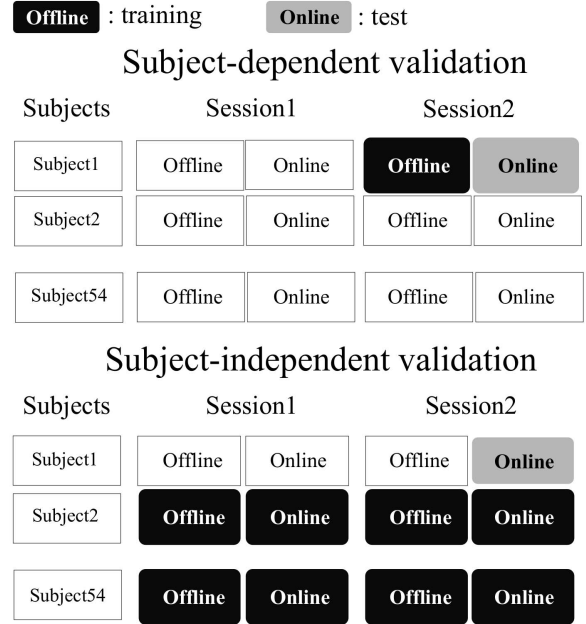


Fig. 4. Example of data validation according to the method of analysis. Black box: training samples. Gray box: testing samples.

At each iteration, the total number of samples (i.e., 10 subjects \times 200 trials \times 2 sessions \times 20 frequency bands) was used to train the proposed CNN model, and it was then applied to the test subject. In Fig. 6, the box plot presents the mean value and the standard deviation of all the results. The accuracies are 77.80% (± 4.66), 80.40% (± 5.41), 84.60% (± 0.55), 80.60% (± 6.51), 83.20% (± 3.63), and 84.0% (± 1.30) when 5, 10, 20, 30, 40, and 53 subjects were used, respectively.

Fig. 7 shows the performance variations with a different number of frequency bands. The classification performance was validated by increasing the number of frequency bands (\mathbf{b}_p) from a single optimal frequency band (i.e., \mathbf{b}_1) to all frequency bands (i.e., \mathbf{b}_p , $p = 1, \dots, 30$). The classification performance was obtained by LOSO-CV for all subjects ($N = 1, \dots, 54$). For example, assume that we selected one subject for the test data ($N = 1$) and the remaining subjects for the training data ($N = 2, \dots, 54$) and ten frequency indices (i.e., \mathbf{b}_p , $p = 1, \dots, 10$) were used to train the CNN model (i.e., 53 subjects \times 200 trials \times 2 sessions \times 10 frequency bands); then, the trained CNN model was applied to the test data. This procedure was repeated for all 54 subjects, and the averaged value of the 54 iteration results

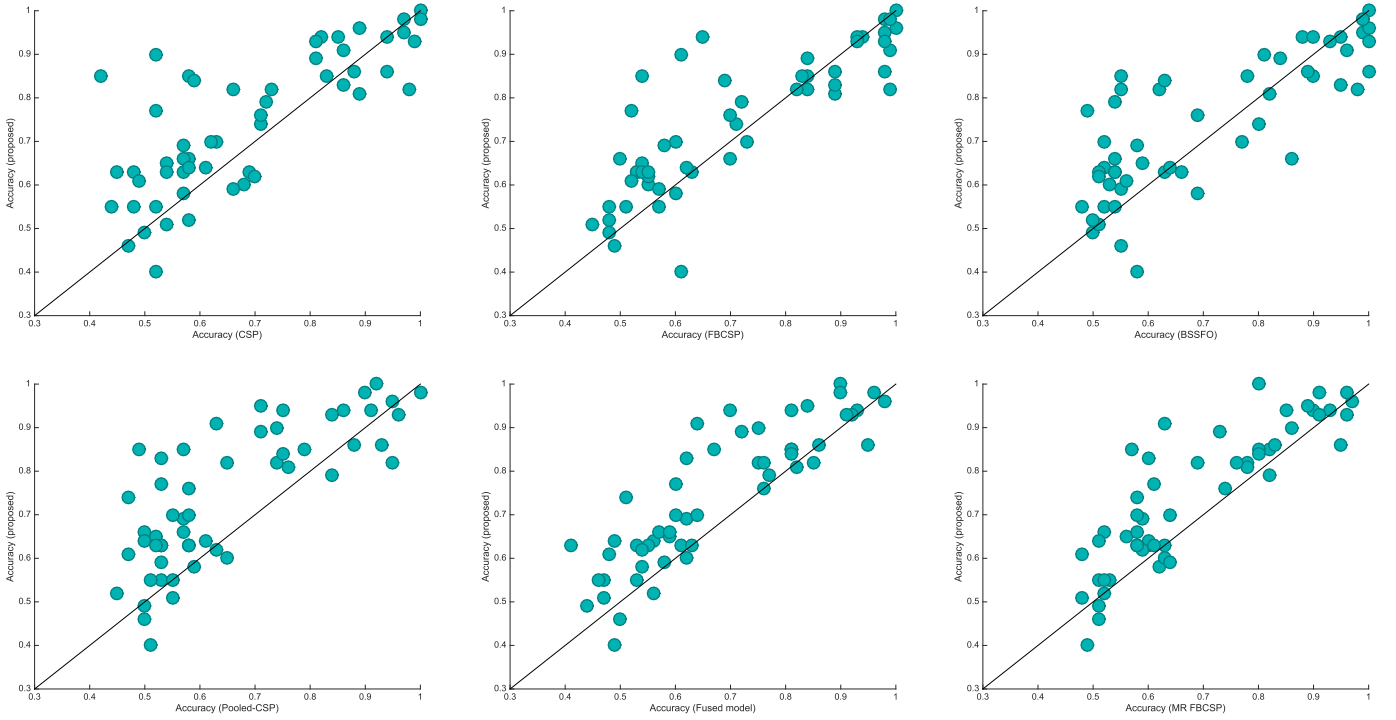


Fig. 5. Scatter plot of individual classification performance in our MI database. Top row: subject-dependent methods, CSP, FBCSP, and BSSFO. Bottom row: subject-independent methods, pooled CSP, fused model, and MR FBCSP. The horizontal axis is the decoding accuracies of the other methods, and the vertical axis is the decoding accuracies of the proposed method.

was presented. The decoding accuracies shown in Fig. 7 are 52.84% (± 6.70), 68.00% (± 14.74), 71.68% (± 15.27), 73.56% (± 16.04), 74.15% (± 15.83), 74.26% (± 15.95), and 73.59% (± 16.15) for 1, 5, 10, 15, 20, 25, and 30 bands, respectively.

VI. DISCUSSION

A. Statistical Significance Tests of Experimental Results

For statistical significance tests of the experimental results on our dataset, the null hypothesis in this article is that the proposed method produces the same mean accuracy as the competing methods. We computed an ANOVA test with Bonferroni for a multiple comparison test. In addition, we computed the p -values using a paired t -test to assess whether the differences in the classification accuracies between two methods are at a significant level. In an ANOVA test of subject-independent methods, the proposed method significantly outperforms the competing methods [$F_{(3,212)} = 2.9184$, $p = 0.0351$]. In a paired t -test, the proposed method also significantly outperforms other methods, pooled CSP [$t_{(53)} = -5.8110$, $p < 0.001$], fused model [$t_{(53)} = -6.1649$, $p < 0.001$], and MR FBCSP [$t_{(53)} = -5.0649$, $p < 0.001$]. In the multiple comparison test of the subject-dependent methods and the proposed method, the result of the ANOVA test showed no significant difference [$F_{(4,265)} = 0.7373$, $p = 0.5672$]. In a paired t -test, the differences in the classification accuracies between the proposed method and subject-dependent methods are at a significant level, CSP [$t_{(53)} = -3.5823$, $p < 0.001$], CSSP [$t_{(53)} = -2.8723$, $p = 0.0058$], FBCSP [$t_{(53)} = -2.5877$, $p = 0.0124$], and BSSFO [$t_{(53)} = -2.0430$, $p = 0.046$].

B. Effect of the Number of Subjects and Frequency Bands

Due to the lack of a large public MI database, the previous studies using DL could work only in subject-dependent environments [48]–[51]. With the use of our large-scale MI database, it is interesting to investigate how many data are truly needed to develop an acceptable subject-independent model with a DL network (i.e., what is the number of training samples required to train the CNN?).

Our MI database includes a large number of subjects over multiple sessions. Therefore, the database includes a variety of brain signals, which represent inherent features for individual subjects and EEG variability over different sessions.

Fig. 6 shows the performance changes according to the number of subjects. Fig. 6 also shows that the classification performance varies with the composition of the subjects used to train the CNN. For the training of the CNN, a particular number of subjects were randomly chosen among all subjects in each increasing step. The arbitrary selection process was repeated five times, and the averaged performance was calculated based on the results of five iterations, as described in Section V. In Fig. 6, it can be seen that there were enormous performance variations over each iteration. For example, when five subjects were used, the results of the first and second iterations showed significant differences in the classification performance. This could be interpreted as the presence of significant differences in the data between different subjects, and therefore, a small number of training samples could either help or harm the actual training. This has been confirmed in the real world by other researchers who considered negative

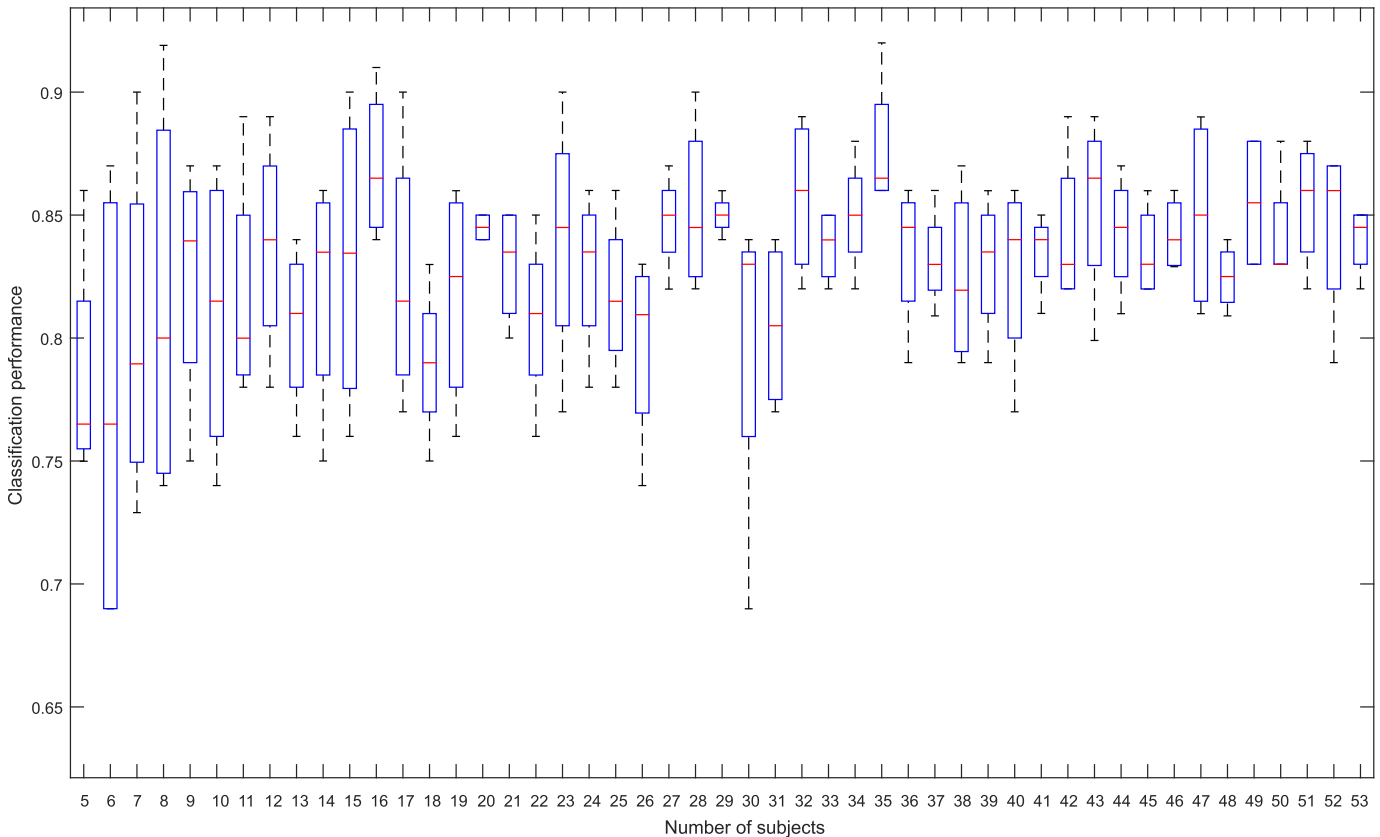


Fig. 6. Classification performance changes with the number of subjects. One subject was chosen for the test data, and 20 frequency indices were used for this experiment. The composition of subjects for training was increased in steps of one from $n=5$ to $n=53$ and was randomly changed for each iteration. The box plot shows the classification performance over five iterations.

transfer (an indication that knowledge transfer has a negative impact on target learning) [55].

The larger the number of training samples that are utilized, the better the classification performance that can be obtained, as shown in Fig. 6. From this perspective, we can explain the differences in the classification performance between the subject-dependent and subject-independent environment, as shown in Table II. The results in Table II show that while the CNN works equally well as other conventional approaches in the subject-dependent condition, it works much better than the conventional approaches in the subject-independent condition. Fig. 6 clearly shows that more training samples contribute to better classification performance. With both the large dataset and DL, our proposed method outperformed the subject-dependent methods.

In addition, the investigation of user-specific frequency bands was primarily considered in the subject-dependent BCI study to improve the decoding accuracy. One prominent strategy [30] is to stochastically find multiple frequency ranges that are optimal for individual subjects. However, from the perspective of subject-independent BCI, it is difficult to find or define specific frequency bands due to the unavailability of training data from the target subjects [22]. Therefore, we included standard frequency bands (e.g., mu rhythm and beta rhythm) that have been widely used in MI studies [23] and additional frequency bands that are empirically determined

and overlap with each band, as shown in Fig. 2. The reason why we deployed a large number of overlapping frequency bands is that there is no single frequency range that is the most discriminative for all subjects. Therefore, we decided to use a filter bank that consisted of a broad variety of frequency bands, to provide a greater diversity of EEG patterns from all training samples. As the number of frequency bands increased in Fig. 7, the classification performance also improved, which is in line with our previous belief that each frequency segment in \mathbf{b}_p represents the diversity of EEG patterns that contained the inherent information. Furthermore, as shown in Fig. 7, the performance improvement slows down after 20 frequency indices ($\mathbf{b}_p, p > 20$). We determined that 20 frequency indices are the proper size of the filter bank ($\mathbf{b}_p, p = 20, \dots, 30$) to show performance superior to that of a single optimal frequency band (\mathbf{b}_1). Furthermore, an increase in the filter bank leads to an increase in the number of training samples, which is beneficial to the CNN training. As an example, a strategy of changing the configuration of the filter bank to increase the number of training samples is similar to that of data augmentation [22].

To summarize, a large-scale MI database is essential for the enhancement of classification performance, and a proper configuration of the filter bank is also an essential factor in profoundly influencing the decoding accuracy of the subject-independent BCI based on DL.

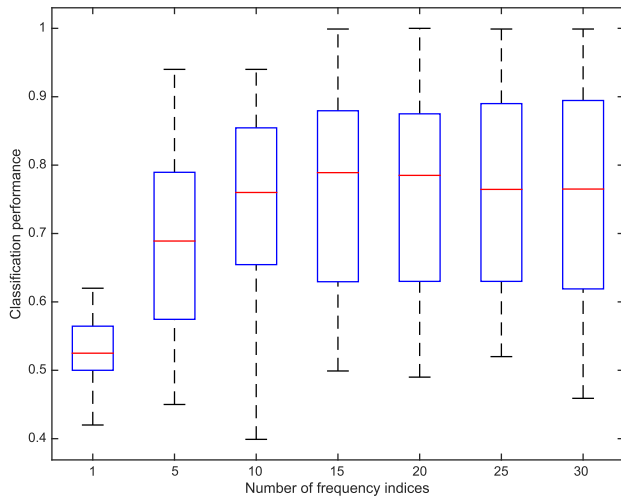


Fig. 7. Classification performance variations according to the number of frequency indices. The mean performance for all subjects was obtained by LOSO-CV.

C. Consideration of the Proposed Input and Features

In this article, one of the main objectives is to construct generalizable brain features. In fact, the signal variability, the amplitude change, the temporal onset variations, and the spatial origin of the EEG segments are different from subject to subject. If we use simple features, such as the logarithm and variance of the EEG signals, it is difficult to extract significant and generalizable features that can include the information for different subjects.

As it is known, CSP maximizes the variance differences between binary class signals. The principle of CSP is to perform a linear spatial transformation to project the high-dimensional EEG data into a low-dimensional spatial subspace with such a projection matrix, where each row represents a spatial filter that consists of the weights for each channel. Specifically, the first and last columns of the spatial filters are the most important spatial filters, which results in the largest variance of one task and the smallest variance of the other. Overall, CSP-based methods were used to form a subset of the most discriminant spatial filters (i.e., one or more spatial filters).

With the feature representation and CNN architecture presented in this article, we believe that the learning process should be based on the most important spatial filters as well as the relations between each column of spatial filters that would be helpful to construct generalizable brain features from a large database. Therefore, we started with the idea of combining CSP from various frequency bands which, in turn, could capture relations between the columns of spatial filters. To support our idea, there were many studies in the past that were relevant to our idea and that utilize the relationships between the features, such as the connectivity or correlation between other feature dimensions, to improve the classification performance or the predictions in research on traditional neuroscience, MI, and other subjects [56]–[59].

In addition, to understand how much the given input is related to the MI signals, let us study the first and last columns

of spatial filters that are the most significant information in CSP, which demonstrates the largest variance of one class and the smallest variance of the other class. From these insights, if we apply CSP to the EEG signals and examine the covariance matrix of the given input, one side of the covariance matrix would have relatively high values compared with the other side due to the largest and smallest spatial filters. For example, the first spatial filter was meant to maximize the left class of MI, and therefore, the left diagonal components in the left MI input would have comparatively high values. To understand this aspect in more detail, let us investigate the comprehension about what features are learned from the networks and how those features are related to MI signals.

The inputs (covariance matrices) are shown at the top part of Fig. 8, and it can be seen that one side of the diagonal components in the covariance matrix has relatively high values (red) due to the discriminating effect of the spatial filters. This can be interpreted as follows: the given input is a representation of the MI signals (left or right motor imagination). To see what features have been primarily learned and to understand how CNNs work due to the large number of interacting and nonlinear operations, we visualized the activation of each convolutional network layer in Fig. 8 of the CNN [60]. In Fig. 8, the relative value of the activation in each convolutional layer is displayed in color. A larger amount of red tells us that there is more activation in each convolutional network layer. Through more active regions in the activation map, it is possible to see what features the network learned.

From each activation map, it can be seen that the proposed method effectively exploited the distributed and hierarchical features through multiple layers. On the first activation mapped in the first layer, the network focused on the simplistic features, such as straight edges and simple colors in the given input. As the layers become deeper, the learning process is based on the values according to the MI class in the covariance input. On the last activation map of the last layer, the network focused not only on the most important spatial filters (for the left class, more red appears on the upper left corner, while for the right class, more red appears on the lower right corner) but also on correlations (more red dots appear on the lower left and upper right part) between different feature dimensions. From this observation, we can conclude that the most important spatial filters are significant for the classification in addition to the relations between the other feature dimensions that are critical for the classification.

This result could shed light on why the proposed method outperformed other CSP-based methods that use few spatial filters and linear characteristics to classify brain signals. When only a few spatial filters of CSP are used, the correlations between different feature dimensions in the inputs are lost, which leads to degraded classification performance. In our strategy, as we consider all input information as well as the correlations between all the dimensions, our approach can outperform other methods.

To summarize, the proposed feature shows a high level of performance through learning the most significant spatial filters as well as the relationships between different feature dimensions that are related to MI. This study is the first attempt

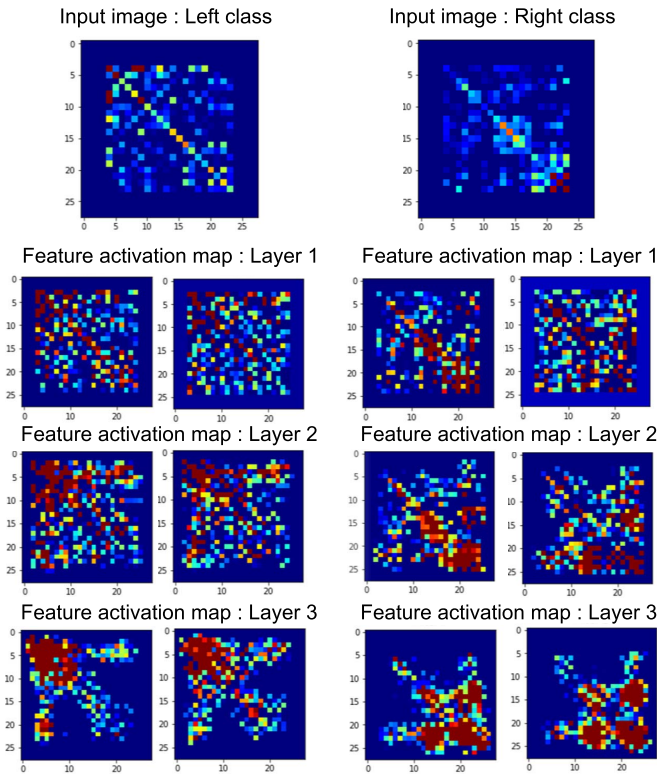


Fig. 8. Visualization of the proposed input and the 28×28 activations on each convolutional layer. For the interpretability of the visualization, two filters in each convolutional layer were shown. The activations in each convolutional layer show the corresponding response of the given input. For example, in the right class, as the lower right has the class-specific discriminative regions according to the most important spatial filters (showing red colors), the lower right in the activation map primarily shows a set of significant regions. Furthermore, the lower left and the upper right tend to show discriminative patterns from the relations between other feature dimensions. In nearly every case, we have found that someone can guess what class a neuron represents by observing a set of optimized and preferred images.

to exploit generalizable brain features in the CNN pipeline for subject-independent BCI.

D. Consideration of the Proposed Framework

In this article, the proposed framework is composed of three convolutional layers, a concatenation layer for fusion, and a fully connected layer. The input of our framework is the covariance matrix, which has a dimension of 28×28 . Since the input dimension is relatively small, we hypothesize that significant features could be extracted through a few convolutional layers rather than many convolutional layers. Based on these considerations, we used three layers in the CNN. Similarly, we think that similar features could be extracted regardless of whether the number of feature maps in the convolutional layer is large or small. Based on these considerations, we decided to start with ten feature maps. Because of the small input size, we also used a small kernel size. The above-mentioned considerations might also be beneficial to overcome a possible overfitting problem due to the relatively small number of free parameters in the proposed model.

To determine the convergence process of the proposed framework, we investigated the evolution of the classification

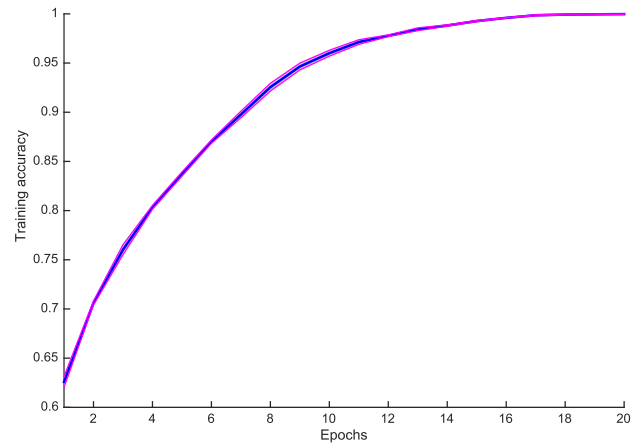


Fig. 9. Effect of the number of epochs on the classification performance in the training phase. Blue line: averaged accuracy over five iterations. Magenta region: standard deviation of the mean accuracy.

results during the training phase. The training process converged within approximately 20 epochs for all subjects. Thus, we chose 20 epochs for early stopping, as shown in Fig. 9. The total number of free parameters is 72 264 076.

Our approach starts by utilizing multiple frequency bands. We believe that those frequency bands have discriminant brain features. Thus, we trained all frequency bands individually to obtain frequency-wise features through the convolutional layers. Then, we used the concatenation fusion technique and integrated all frequency-wise features, which transformed high-dimensional convolutional features into a vector with low dimensionality for parameter and dimension reduction [61], [62].

In the training environment (i.e., 53 subjects and 20 frequency bands), the time duration was 12 min for each individual subject. The testing session of the proposed method requires a duration of approximately 0.15 ms to classify a single trial. With the subject-independent approach, the BCI system is pretrained before being employed by a new subject. Therefore, the testing time is of more concern than the training time in the BCI system. The test time duration of other methods in the training (testing) phase was 48.18 s (0.051 s), 853.418 s (0.036 s), and 1049.6679 s (0.022 s) for pooled CSP, fused model, and MR FBCSP, respectively.

E. Consideration of the Performance Results

From the individual classification performance shown in Fig. 5, it can be seen that the proposed method would sacrifice a certain percentage of the excellent subjects' performance while significantly improving some moderate subjects' performance. We suggest that the reason is that the proposed method could construct generalizable features. Generally, those who showed a low classification performance should require a number of training samples to extract discriminant brain features. However, the experimental process for moderate users in a subject-dependent environment would end before acquiring many samples that can help them construct their own brain features. In this article, we provide a large number of training samples from a variety of subjects and various

frequencies' information. This could be helpful in constructing generalizable features for moderate performers and improving their classification performance.

However, excellent subjects might have their own inherent brain features and thus might not need many training samples. In other words, excellent subjects might already have discriminative features. This means that generalizable brain features could deteriorate the classification performance of excellent subjects. In our opinion, this aspect is not a major concern, as we know that it is very difficult to improve the performance for moderate and poor subjects while it is much easier to achieve high performance for excellent subjects. For example, one can use the resting state of the brain signal to predict whether this subject is an excellent BCI user [63]. Once we know that a subject is an excellent BCI user, we can apply user-specific or adaptation methods with very few trials to refine the classifier, to achieve good classification performance [37], [39]. Furthermore, even for moderate BCI users, we can also apply adaptation methods based on a pretrained model. If the initial value of the classifier parameters is stable and generalized, adaptation methods will be effective for those who are naive BCI users or experienced to create their own decoder for the long-term use of BCIs.

F. Consideration of Current Issues in Motor Imagery

In current MI-based BCI, the intrasubject and intersubject variability of brain signal is one of the fundamental issues [23]–[27]. In a recent interesting study, they showed that a large number of BCI users have performance variations over both subjects and sessions [64]. In fact, many researchers have also raised similar questions in that there would be a high level of subject-wise signal variability on EEG signals when the participants perform the MI tasks [22]. Here, we believed that the variability of brain signals in individual subjects is inevitable; therefore, we should find out solutions to prevent performance degradation by the signal variability. To do this, we examined this issue as to how subject-wise signal variability affects performance degradation and how we can obtain better classification performance despite these signal variability.

In Fig. 6, as more subjects were used for training, we could see that performance degradation by subject-wise signal variability was reduced (high mean accuracy and low variation). In Fig. 7, utilizing multiple frequency indices allows the model to achieve a robust result while preventing performance deterioration. Additionally, in Fig. 2, we can observe how the range of frequency bands can be key to obtaining robust performance despite the signal variability. Here, the rearranged frequency indices were chosen according to the value of the mutual information, and most rearranged indices were located on narrow frequency bands. This finding appears to indicate that narrow frequency bands are regions that can extract discriminate features and thus provide solid classification performance and further could even reduce subject-wise signal variability.

Beyond the MI area, other EEG-based studies have also focused on these signal variability in individual subjects.

Reference [7] showed individual difference and subject variability of brain signal between the control and schizophrenia participants in the ERP task. Reference [65] showed clear individual difference of brain signal through visual and auditory ERP tasks. In summary, variability of brain signal in individual subjects is unavoidable and it is crucial to develop new algorithms to mitigate the performance deterioration and obtain robust performance.

G. Limitations and Future Study

This article has several limitations that call for future investigations. First, the proposed method is based on the CNN structure, and there are several new architectures that are worthwhile to explore [61], [62]. These new architectures might be useful to further boost the performance. Second, we could also use other types of inputs, such as a frequency-by-time matrix or channel-by-time matrix, to allow the deep neural nets to learn the most discriminant features for classification. Third, we need a more informative visualization tool to interpret the outcome of a DL model, such as in [60] and [66], for a better understanding of the neurophysiological patterns of the MI. Finally, a fast adaptation method based on a pretrained model can be developed for the long-term use of BCI [37], [39].

VII. CONCLUSION

In this article, we proposed a subject-independent CNN framework for an MI-based BCI system. Spectral–spatial input generation is used to represent the general brain signal patterns from a large-scale MI database. The experimental results indicate that the proposed method significantly outperformed previous subject-independent as well as conventional subject-dependent approaches. In conclusion, this article demonstrates the superior performance and promising potential of the proposed feature representation coupled with a deep neural network method. This article could pave the way for a practical implementation of a subject-independent BCI.

REFERENCES

- [1] G. Purtscheller and C. Neuper, "Motor imagery and direct brain-computer communication," *Proc. IEEE*, vol. 89, no. 7, pp. 1123–1134, Jul. 2001.
- [2] H. R. Leigh *et al.*, "Reach and grasp by people with tetraplegia using a neurally controlled robotic arm," *Nature*, vol. 485, no. 7398, pp. 372–375, May 2012.
- [3] S. Gaurav *et al.*, "Using an artificial neural bypass to restore cortical control of rhythmic movements in a human with quadriplegia," *Sci. Rep.*, vol. 6, Sep. 2016, Art. no. 33807.
- [4] K.-K. Ang and C. Guan, "Brain–computer interface for neurorehabilitation of upper limb after stroke," *Proc. IEEE*, vol. 103, no. 6, pp. 944–953, Jun. 2015.
- [5] A. Ramos-Murguialday *et al.*, "Brain-machine interface in chronic stroke rehabilitation: A controlled study," *Ann. Neurol.*, vol. 74, no. 1, pp. 100–108, Jul. 2013.
- [6] L. Li, G. Huang, Q. Lin, J. Liu, S. Zhang, and Z. Zhang, "Magnitude and temporal variability of inter-stimulus EEG modulate the linear relationship between laser-evoked potentials and fast-pain perception," *Frontiers Neurosci.*, vol. 12, p. 340, May 2018.
- [7] C. A. Mugruza-Vassallo and D. D. Potter, "Context dependence signature, stimulus properties and stimulus probability as predictors of ERP amplitude variability," *Frontiers Hum. Neurosci.*, vol. 13, p. 39, Feb. 2019.

- [8] 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.
- [9] 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.
- [10] D.-O. Won, H.-J. Hwang, S. Dähne, K.-R. Müller, and S.-W. Lee, "Effect of higher frequency on the classification of steady-state visual evoked potentials," *J. Neural Eng.*, vol. 13, no. 1, Feb. 2016, Art. no. 016014.
- [11] M.-H. Lee, J. Williamson, D.-O. Won, and S.-W. Lee, "A high performance spelling system based on EEG-EOG signals with visual feedback," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 7, pp. 1443–1459, Jul. 2018.
- [12] K. LaFleur, K. Cassady, A. Doud, K. Shades, E. Rogin, and B. He, "Quadcopter control in three-dimensional space using a noninvasive motor imagery-based brain-computer interface," *J. Neural Eng.*, vol. 10, no. 4, Aug. 2013, Art. no. 046003.
- [13] R. Caraballona, F. Grossi, A. Tessadri, P. Castiglioni, A. Caracciolo, and I. de Munari, "Light on! Real world evaluation of a P300-based brain-computer interface (BCI) for environment control in a smart home," *Ergonomics*, vol. 55, no. 5, pp. 552–563, Mar. 2012.
- [14] J. Meng, S. Zhang, A. Bekyo, J. Olsoe, B. Baxter, and B. He, "Non-invasive electroencephalogram based control of a robotic arm for reach and grasp tasks," *Sci. Rep.*, vol. 6, Dec. 2016, Art. no. 38565.
- [15] N.-S. Kwak and S.-W. Lee, "Error correction regression framework for enhancing the decoding accuracies of ear-EEG brain-computer interfaces," *IEEE Trans. Cybern.*, to be published.
- [16] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review," *Sensors*, vol. 12, no. 2, pp. 1211–1279, Jan. 2012.
- [17] M.-H. Lee, S. Fazli, J. Mehnert, and S.-W. Lee, "Subject-dependent classification for robust idle state detection using multi-modal neuroimaging and data-fusion techniques in BCI," *Pattern Recognit.*, vol. 48, pp. 2725–2737, Aug. 2015.
- [18] S.-K. Yeom, S. Fazli, K.-R. Müller, and S.-W. Lee, "An efficient ERP-based brain-computer interface using random set presentation and face familiarity," *PLoS ONE*, vol. 9, no. 11, Nov. 2014, Art. no. e111157.
- [19] 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, Art. no. e0172578, Feb. 2017.
- [20] C. Wu, K. Lin, W. Wu, and X. Gao, "A novel algorithm for learning sparse spatio-spectral patterns for event-related potentials," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 4, pp. 862–872, Apr. 2017.
- [21] H. Cecotti, M. P. Eckstein, and B. Giesbrecht, "Single-trial classification of event-related potentials in rapid serial visual presentation tasks using supervised spatial filtering," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 11, pp. 2032–2042, Nov. 2014.
- [22] F. Lotte, "Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain-computer interfaces," *Proc. IEEE*, vol. 103, no. 6, pp. 871–890, Jun. 2015.
- [23] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 8, no. 4, pp. 441–446, Dec. 2000.
- [24] F. Lotte, M. Congedo, A. Lecuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain-computer interfaces," *J. Neural Eng.*, vol. 4, no. 2, pp. R1–R13, Jun. 2007.
- [25] B. Blankertz, G. Dornhege, M. Krauledat, K.-R. Müller, and G. Curio, "The non-invasive Berlin brain-computer interface: Fast acquisition of effective performance in untrained subjects," *NeuroImage*, vol. 37, no. 2, pp. 539–550, Aug. 2007.
- [26] B. Blankertz, M. Kawanabe, R. Tomioka, F. U. Hohlefeld, V. Nikulin, and K.-R. Müller, "Invariant common spatial patterns: Alleviating nonstationarities in brain-computer interfacing," in *Proc. Neural Inf. Process. Syst. (NIPS)*, 2008, pp. 113–120.
- [27] H. Wang and W. Zheng, "Local temporal common spatial patterns for robust single-trial EEG classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 16, no. 2, pp. 131–139, Apr. 2008.
- [28] K.-K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jun. 2008, pp. 2391–2398.
- [29] S. Lemm, B. Blankertz, G. Curio, and K.-R. Müller, "Spatio-spectral filters for improving the classification of single trial EEG," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 9, pp. 1541–1548, Sep. 2005.
- [30] 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.
- [31] A. M. Ray *et al.*, "A subject-independent pattern-based brain-computer interface," *Frontiers Behav. Neurosci.*, vol. 9, p. 269, Oct. 2015.
- [32] F. Lotte, C. Guan, and K.-K. Ang, "Comparison of designs towards a subject-independent brain-computer interface based on motor imagery," in *Proc. 31st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Sep. 2009, pp. 4543–4546.
- [33] S. Fazli, F. Popescu, M. Danóczy, B. Blankertz, K.-R. Müller, and C. Grozea, "Subject-independent mental state classification in single trials," *Neural Netw.*, vol. 22, no. 9, pp. 1305–1312, Nov. 2009.
- [34] M. Alamgir, M. Grosse-Wentrup, and Y. Altun, "Multitask learning for brain-computer interfaces," in *Proc. 13th Int. Conf. Artif. Intell. Statist. (AISTATS)*, 2010, pp. 17–24.
- [35] V. Jayaram, M. Alamgir, Y. Altun, B. Scholkopf, and M. Grosse-Wentrup, "Transfer learning in brain-computer interfaces," *IEEE Comput. Intell. Mag.*, vol. 11, no. 1, pp. 20–31, Feb. 2016.
- [36] W. Samek, F. C. Meinecke, and K.-R. Müller, "Transferring subspaces between subjects in brain-computer interfacing," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 8, pp. 2289–2298, Aug. 2013.
- [37] C. Vidaurre, M. Kawanabe, P. von Büna, B. Blankertz, and K.-R. Müller, "Toward unsupervised adaptation of LDA for brain-computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 3, pp. 587–597, Mar. 2011.
- [38] B. Reuderink, J. Farquhar, M. Poel, and A. Nijholt, "A subject-independent brain-computer interface based on smoothed, second-order baselining," in *Proc. 33rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug./Sep. 2011, pp. 4600–4604.
- [39] A. Llera, V. Gómez, and H. J. Kappen, "Adaptive multiclass classification for brain computer interfaces," *Neural Comput.*, vol. 26, no. 6, pp. 1108–1127, Jun. 2014.
- [40] L. Arnold, S. Rebecchi, S. Chevallier, and H. Paugam-Moisy, "An introduction to deep learning," in *Proc. Eur. Symp. Artif. Neural Netw. Comput. Intell. Mach. Learn. (ESANN)*, 2011, pp. 477–488.
- [41] Y. Sun, Y. Chen, X. Wang, and X. Tang, "Deep learning face representation by joint identification-verification," in *Proc. Neural Inf. Process. Syst. (NIPS)*, 2014, pp. 1988–1996.
- [42] L. Deng and D. Yu, "Deep learning: Methods and applications," in *Foundations and Trends in Signal Processing*. Hanover, MA, USA: Now, 2014, pp. 197–387.
- [43] Y. LeCun, K. Kavukcuoglu, and C. Farabet, "Convolutional networks and applications in vision," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May/Jun. 2010, pp. 253–256.
- [44] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 1097–1105.
- [45] N. Liu, J. Han, T. Liu, and X. Li, "Learning to predict eye fixations via multiresolution convolutional neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 2, pp. 392–404, Feb. 2018.
- [46] J. Liu, M. Gong, K. Qin, and P. Zhang, "A deep convolutional coupling network for change detection based on heterogeneous optical and radar images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 3, pp. 545–559, Mar. 2018.
- [47] W. Shi, Y. Gong, X. Tao, J. Wang, and N. Zheng, "Improving CNN performance accuracies with min-max objective," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 7, pp. 2872–2885, Jul. 2018.
- [48] Y. R. Tabar and U. Halici, "A novel deep learning approach for classification of EEG motor imagery signals," *J. Neural Eng.*, vol. 14, no. 1, p. 016003, Feb. 2017.
- [49] N. Lu, T. Li, X. Ren, and H. Miao, "A deep learning scheme for motor imagery classification based on restricted Boltzmann machines," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 6, pp. 566–576, Jun. 2017.
- [50] R. T. Schirrmester *et al.*, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Hum. Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.
- [51] S. Sakhavi, C. Guan, and S. Yan, "Learning temporal information for brain-computer interface using convolutional neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5619–5629, Nov. 2018.
- [52] H.-J. Hwang, K. Kwon, and C.-H. Im, "Neurofeedback-based motor imagery training for brain-computer interface (BCI)," *J. Neurosci. Methods*, vol. 179, no. 1, pp. 150–156, Apr. 2009.
- [53] M.-H. Lee, S. Fazli, K.-T. Kim, and S.-W. Lee, "Development of an open source platform for brain-machine interface: OpenBMI," in *Proc. 4th Int. Winter Conf. Brain-Comput. Interface (BCI)*, Feb. 2016, pp. 1–2.

- [54] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.
- [55] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
- [56] D. Rathee, H. Cecotti, and G. Prasad, "Single-trial effective brain connectivity patterns enhance discriminability of mental imagery tasks," *J. Neural Eng.*, vol. 14, no. 5, Oct. 2017, Art. no. 056005.
- [57] I. Daly, S. J. Nasuto, and K. Warwick, "Brain computer interface control via functional connectivity dynamics," *Pattern Recognit.*, vol. 45, no. 6, pp. 2123–2136, Jun. 2012.
- [58] K. Casimo, K. E. Weaver, J. Wander, and J. G. Ojemann, "BCI use and its relation to adaptation in cortical networks," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 10, pp. 1697–1704, Oct. 2017.
- [59] A. Kabbara, M. Khalil, W. El-Falou, H. Eid, and M. Hassan, "Functional brain connectivity as a new feature for P300 speller," *PLoS ONE*, vol. 11, no. 1, Jan. 2016, Art. no. e0146282.
- [60] J. Yosinski, J. Clune, A. Nguyen, T. Fuchs, and H. Lipson, "Understanding neural networks through deep visualization," in *Proc. 32nd Int. Conf. Mach. Learn. (ICML)*, 2015, pp. 1–12.
- [61] C. Feichtenhofer, A. Pinz, and A. Zisserman, "Convolutional two-stream network fusion for video action recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1933–1941.
- [62] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.
- [63] B. Blankertz *et al.*, "Neurophysiological predictor of SMR-based BCI performance," *NeuroImage*, vol. 51, no. 4, pp. 1303–1309, Jul. 2010.
- [64] M.-H. Lee *et al.*, "EEG dataset and OpenBMI toolbox for three BCI paradigms: An investigation into BCI illiteracy," *GigaScience*, vol. 8, no. 5, pp. 1–16, May 2019.
- [65] C. Mugruza-Vassallo, "Different regressors for linear modelling of ElectroEncephaloGraphic recordings in visual and auditory tasks," in *Proc. 13th Int. Conf. Wearable Implant. Body Sensor Netw. (BSN)*, Jun. 2016, pp. 260–265.
- [66] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2129–2921.



O-Yeon Kwon received the B.S. degree in biomedical engineering from Yonsei University, Seoul, South Korea, in 2015, and the M.S. degree in brain and cognitive engineering from Korea University, Seoul, in 2018.

He is currently working in VUNO, South Korea. His current research interests include machine learning, brain–computer interfaces, and deep learning.



Min-Ho Lee received the B.S. degree in computer science from Seokyeong University, Seoul, South Korea, in 2012, and the integrated master's and Ph.D. degree in brain and cognitive engineering from Korea University, Seoul, in 2019.

He is currently an Assistant Professor with the Department of Computer Science, Nazarbayev University, Astana, Kazakhstan. His current research interests include machine learning, brain–computer interfaces, and neurofeedback systems.



Cuntai Guan (S'91–M'92–SM'03–F'18) received the Ph.D. degree from Southeast University, Nanjing, China, in 1993.

He is currently a Professor with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. His current research interests include brain–computer interfaces, neural engineering, machine learning, data analytics, and neurotechnologies.

Dr. Guan was a recipient of the Annual BCI Research Award, the IES Prestigious Engineering Achievement Award, the Achiever of the Year (Research) Award, the Finalist of President Technology Award, and the National Infocomm Award. He is an Associate Editor of the *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING*, *Frontiers in Neuroscience*, and *Brain Computer Interfaces* and a Guest Editor of the *IEEE Computational Intelligence Magazine*.



Seong-Whan Lee (S'84–M'89–SM'96–F'10) received the B.S. degree in computer science and statistics from Seoul National University, Seoul, South Korea, in 1984, and the M.S. and Ph.D. degrees in computer science from the Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 1986 and 1989, respectively.

He is currently the Head of the Department of Artificial Intelligence, Korea University, Seoul.

His current research interests include artificial intelligence, pattern recognition, and brain engineering.

Dr. Lee is a fellow of the IAPR and the Korea Academy of Science and Technology.