

Received July 7, 2020, accepted July 17, 2020, date of publication July 23, 2020, date of current version July 31, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3011140

SessionNet: Feature Similarity-Based Weighted Ensemble Learning for Motor Imagery Classification

BYEONG-HOO LEE¹, JI-HOON JEONG¹, AND SEONG-WHAN LEE², (Fellow, IEEE)

¹Department of Brain and Cognitive Engineering, Korea University, Seoul 02841, South Korea

²Department of Artificial Intelligence, Korea University, Seoul 02841, South Korea

Corresponding author: Seong-Whan Lee (sw.lee@korea.ac.kr)

This work was supported in part by the Institute for Information and Communications Technology Planning and Evaluation (IITP) funded by the Korea Government through the Development of Intelligent Pattern Recognition Softwares for Ambulatory Brain-Computer Interface under Grant 2015-0-00185, in part by the Development of Non-Invasive Integrated BCI SW Platform to Control Home Appliances and External Devices by User's Thought via AR/VR Interface under Grant 2017-0-00432, and in part by the Department of Artificial Intelligence, Korea University, under Grant 2019-0-00079.

ABSTRACT A brain-computer interface (BCI) provides a direct communication pathway between user and external devices. Motor imagery (MI) paradigm is widely used in non-invasive BCI to control external devices by decoding user intentions. The traditional MI-BCI problem is to obtain enough EEG data samples for adopting deep learning techniques, as electroencephalography (EEG) data have intricate and non-stationary properties that can cause a discrepancy between different sessions of data. Because of the discrepancy, the recorded EEG data with different sessions cannot be treated as the same. In this study, we recorded a large intuitive EEG dataset that contained nine types of movements of a single-arm across 12 subjects. We proposed a SessionNet that learns generality with EEG data recorded over multiple sessions using feature similarity to improve classification performance. Additionally, the SessionNet adopts the principle of a hierarchical convolutional neural network that shows robust classification performance regardless of the number of classes. The SessionNet outperforms conventional methods on 3-class, 5-class, and two types of 7-class and 9-class of a single-arm task. Hence, our approach could demonstrate the possibility of using feature similarity based on a novel ensemble learning method to train generality from multiple session data for better MI classification performance.

INDEX TERMS Brain-computer interface (BCI), electroencephalogram (EEG), motor imagery (MI), convolutional neural network (CNN), weighted ensemble learning.

I. INTRODUCTION

Brain-computer interface (BCI) has been widely studied to recover and replace the motor function of motor-disabled patients [1]–[4]. Recent advances in BCI have tried to provide healthy users with extended motor function capabilities to control external devices [5]–[7]. One type of brain signal is electroencephalography (EEG) that can be collected without any brain surgery. Therefore, EEG-based BCI has been investigated for using various paradigms such as event-related potential (ERP) [8], steady-state visual evoked potential (SSVEP) [9], movement-related cortical potential (MRCP) [10], and motor imagery (MI) [11]. According to

the BCI paradigms, ERP is the potential generated by direct brain responses such as P300 and N200 from a stimulus such as a target cognitive. Specifically, a P300-based speller system has been developed using the ERP paradigm [12], [13]. The SSVEP is a natural brain response corresponding to frequencies that are triggered by visual stimuli [14], [15]. The MRCP reflects the user's voluntary movement a few seconds before the movement, so it has been applied in exoskeleton walking experiments [16], [17]. The MI paradigm can induce an event-related desynchronization/synchronization pattern [18] over the supplementary motor area and premotor cortex when the user imagines him or herself performing movements. Because of its neurophysiological origin, MI is mainly used to control external devices [6], [19]. As an endogenous BCI, consistent

The associate editor coordinating the review of this manuscript and approving it for publication was Yizhang Jiang¹.

muscle imagination is required for high classification performance. However, it is difficult to obtain high-quality data in the MI paradigm because what the user imagines is not exactly known. Furthermore, repeated MI causes users to feel severely tired, making it difficult to collect sufficient data from a single subject [20], [21].

Decoding user intentions using EEG signals is one of the important issues because the non-stationary properties of EEG signals cause session and subject dependency [22], [23]. Therefore, many BCI researchers have investigated improving MI-BCI performance with novel machine learning algorithms and deep learning [24]. For example, a filter-bank common spatial pattern (FBCSP) algorithm has been adopted to extract spatial features measuring spectral power modulations from EEG data [25], [26]. For classifiers, they have used linear discriminant analysis (LDA) and support vector machine (SVM) as a classifier [27]. Recently, deep learning approach such as a convolutional neural networks (CNNs) [28]–[31] and long short-term memory (LSTM) [32]–[36], have been introduced to decode user intentions from EEG signals. Specifically, for the MI-based BCI, deep and shallow CNN architectures were designed to investigate the causal contributions of features from different frequency bands [37]. Some groups have attempted depthwise and separable convolutions, which summarizes individual features over time to consider more channel information [30]. However, these approaches have suffered from lack of data samples thereby giving several problems in training procedure. To overcome this constraint, recent studies have tried to use only a small amount of input data based on transfer learning [38]. In addition, a diverse depth of CNN can extract features and use them as an input data to exploit input information [39]. Through this approach, it has been demonstrated that BCI performance enhancement using deep learning is possible with a sufficient amount of feature input data.

In this paper, we propose a SessionNet based on weighted ensemble learning to exploit multi-session data for learning generality. The proposed SessionNet comprises several hierarchical CNNs to utilize multiple session data, where each CNN was assigned specific session data. SessionNet primarily trains using target session data. Furthermore, the proposed model learns generality from other session data (non-target session data) based on the structural similarity (SSIM) index [40]. For the experiment, we designed an intuitive MI paradigm for direct interaction between users and devices without artificial command matching for better performance [41] and collected various types of movements of a single-arm such as arm-related MI and hand-related MI.

Hence, our main contributions are three folds: *i*) We collected a large intuitive EEG data of single-arm MI; arm-reaching MI, hand-grasping MI, and wrist-twisting MI. *ii*) We proposed a novel ensemble learning method for learning generality from other session data. Additionally, we adopted the principle of a hierarchical CNN architecture for robust multi-class classification performance.

iii) The proposed SessionNet achieved remarkable improvement in MI classification performance, and we proved that learning generality using data from different sessions can improve multi-class data classification.

II. MATERIALS AND METHODS

A. DATA ACQUISITION

1) PARTICIPANTS

Twelve naïve subjects (nine males, all right-handed, age: 24 - 31 years) participated in the experiment. All subjects were healthy without any psychiatric or neurological disorders. Before the experiment, the subjects were asked to minimize their eye blanking and consistently imagine the muscle movements. The experimental protocols and environments were reviewed and approved by the Institutional Review Board at Korea University [1040548-KU-IRB-17-172-A-2].

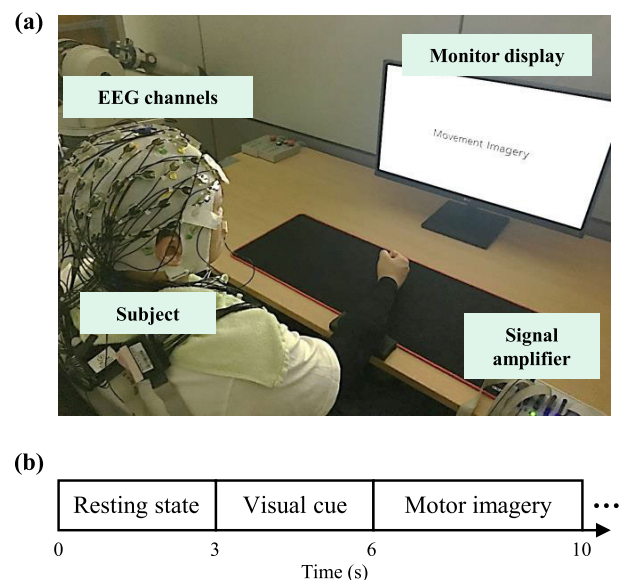


FIGURE 1. Experimental protocol for data acquisition.

(a) An experimental environment consisting of a signal amplifier, monitor display, and EEG channels. (b) Experimental paradigm for obtaining EEG signals with respect to upper-extremity MI.

2) EXPERIMENTAL ENVIRONMENT

We designed an experimental environment and protocol for recording EEG signals (Fig. 1). An EEG signal amplifier (BrainAmp, BrainProduct GmbH, Germany) and BrainVision software were used for recording the EEG signals from 64 Ag/AgCl electrodes according to the 10-20 international system. The FPz and FCz channels were selected as ground and reference electrodes. The sampling rate was set to 1,000 Hz and a 60 Hz notch filter was applied to remove the noise from the DC power supply. The channel impedance measured between the electrodes and the scalp was maintained at 15 k Ω . Visual instructions were given as cues presented on a monitor display. Subjects were asked to imagine the specific muscle movements following the

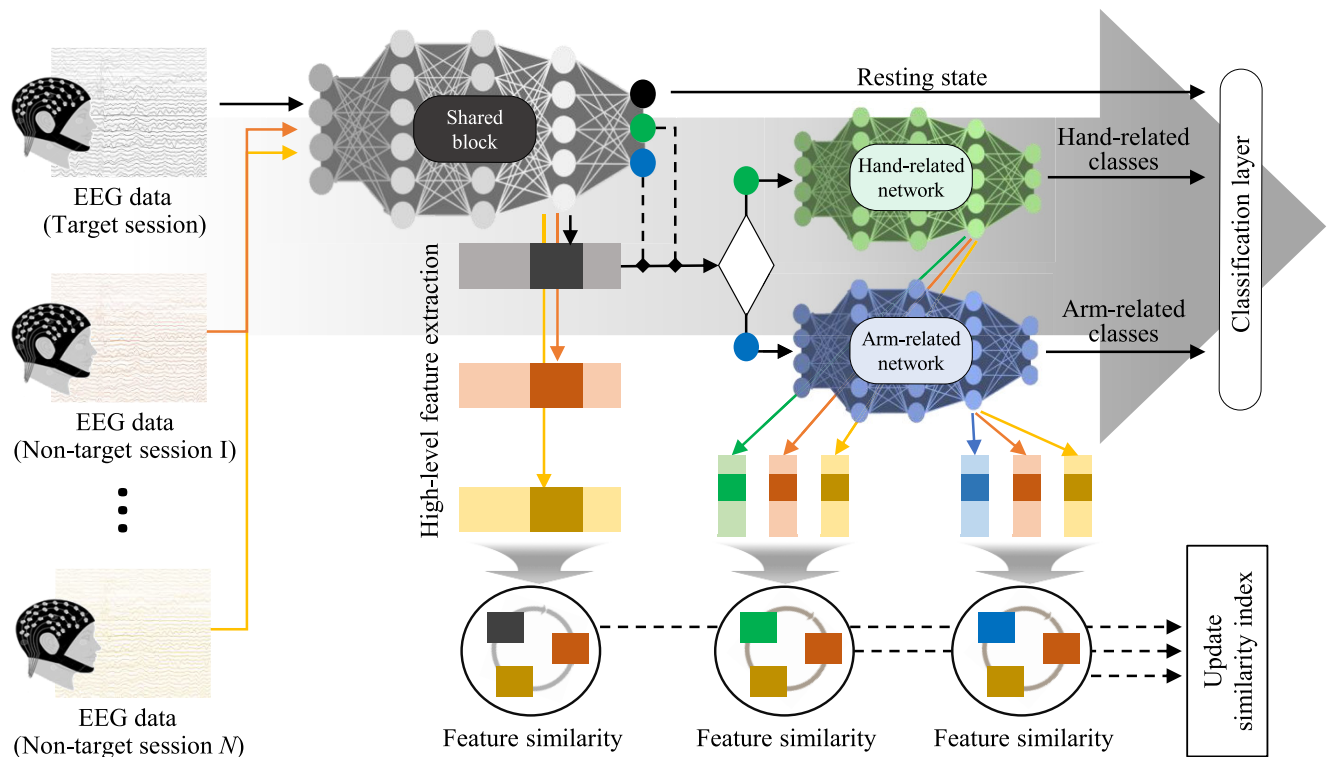


FIGURE 2. Overview of the proposed SessionNet for MI classification.

visual cues. They performed 50 trials according to each MI task. For multi-session recording, every subject participated in three recording sessions at one week intervals with the same experimental protocols.

3) DATASET DESCRIPTION

Six different types of datasets were organized as 3-class, 5-class, 7-class, and 9-class, where the 7-class and 9-class datasets consisted of two types each, based on vertical arm-reaching (upward and downward) and the horizontal arm-reaching (forward and backward). The 3-class dataset contained categorized arm-reaching MI, hand-related MI, and resting-state. The 5-class dataset consisted of left, right, supination, cylindrical grasp, and resting-state. The horizontal 7-class dataset included left, right, forward, backward, supination, cylindrical, and resting-state. The vertical 7-class dataset included left, right, upward, downward, supination, cylindrical grasp, and resting-state. Horizontal and vertical 9-class datasets added lumbrical grasp and pronation to the horizontal and vertical 7-class datasets, respectively.

B. PROPOSED METHOD

As a pre-processing procedure, we used the zero-phase band-pass filter in range of [4-40] Hz and selected 24 channels (F3, F1, Fz, F2, F4, FC3, FC1, FC2, FC4, C3, C1, Cz, C2, C4, CP3, CP1, CPz, CP2, CP4, P3, P1, Pz, P2, and P4) placed on the motor cortex [42]. SessionNet was inspired

by the ensemble learning, especially bagging technique. The fundamental CNN architecture of SessionNet consists of a shared block and two sub-networks. We designed identical CNNs to handle multiple session data. Each CNN was trained by different session data. Model training was organized to optimize each CNN on the assigned session data, and a CNN assigned target session is affected by other CNNs according to a designed loss function. SSIM is used for measuring the structural similarity between two images. EEG data are represented as vector like image, SSIM can measure the similarity between EEG data. Additionally, we assumed that non-target session data were structurally shifted compared to target session data. Therefore, the SSIM was selected as a similarity measurement because we assumed the shift in data as a structural distortion. Hence, CNN becomes committees for influencing training based on the SSIM index [43]. Additionally, SessionNet was designed as a hierarchical CNN architecture to decrease the workload of singular CNN (shared block and sub-networks) extracting frequency features. Hence, it performs step-wise classification splitting the classes as depicted in Fig. 2.

1) HIERARCHICAL CNN

We adopted the concept of a hierarchical CNN architecture, which has shown efficiency in multi-class classification [42]. The shared block is composed of two convolution-pooling blocks. The first block generates the receptive field

from [4-40] Hz and reduces the channel dimension to a single channel using the spatial filter. An additional pair of convolution and pooling layers perform convolution and pooling to extract features. The classification layer categorizes input data into either arm-reaching MI, hand-related MI, or the resting-state. The softmax function normalizes outputs into a probability distribution in the last layer. If SessionNet predicts that the input is a resting-state, it terminates the classification process. In this case, the prediction of the model will be classified as resting-state. Meanwhile, if the result is either arm-reaching or hand-related MI, sub-networks start exploiting the feature (i.e., shared feature) from the second convolution layer of shared block for further classification. Sub-networks are designated to classify either the hand-related MI or the arm-reaching MI. One sub-network was assigned hand-related MI classification, and it consists of three pairs of convolution and pooling layers. The other sub-network was assigned arm-reaching MI classification, and it consists of four convolution-pooling pairs. We conjectured that arm-reaching MI tasks are relatively more complicated than hand-related MI. Therefore, we added one more pair to obtain higher-level features. The parameters of the sub-networks were selected for the best accuracy. During the training, sub-networks classify the shared feature regardless of the shared block's predictions. Therefore, sub-networks can learn both correct and incorrect cases at the same time. Consequently, the two sub-networks become specialized in arm-reaching MI and hand-related MI.

Every convolution layer conducts an average pooling for smoothing and reducing the dimension of the features. In the experimental dataset, average pooling showed better performance than max pooling. An exponential linear unit (ELU) was used as an activation function [44]. Batch normalization was performed to avoid gradient vanishing and exploding. The detailed architecture design and filter sizes are given in Table 1.

TABLE 1. Architecture design of the SessionNet.

	Shared conv-pool block	Sub-network (Hand)	Sub-network (Arm)
Input	Raw EEG (1, 1, 24, 501)	Shared feature (1, 36, 1, 137)	Shared feature (1, 36, 1, 137)
Layers	Conv2D (1, 65, 36 filter)	Conv2D (1, 9, 72 filter)	Conv2D (1, 9, 72 filter)
	Conv2D (24, 1, 36 filter)	AvgPool (1, 3, stride: 1, 3)	AvgPool (1, 3, stride: 1, 3)
	AvgPool (1, 3, stride: 1, 3)	Conv2D (1, 9, 144 filter)	Conv2D (1, 9, 144 filter)
	Conv2D (1, 9, 36 filter)	AvgPool (1, 3, stride: 1, 3)	AvgPool (1, 3, stride: 1, 3)
	AvgPool (1, 10, stride: 1, 3)	Conv2D (1, 9, 288 filter)	Conv2D (1, 9, 288 filter)
	Fully-Connected Layer	AvgPool (1, 3, stride: 1, 3)	AvgPool (1, 3, stride: 1, 3)
	Softmax	Fully-Connected Layer	Conv2D (1, 9, 288 filter)
		Softmax	AvgPool (1, 3, stride: 1, 3)
			Fully-Connected Layer
			Softmax
Activation	ELU	ELU	ELU
Optimizer	Adam	Adam	Adam
Loss function	Cross entropy	Cross entropy	Cross entropy

2) WEIGHTED ENSEMBLE LEARNING

The proposed SessionNet adopted ensemble learning to improve MI classification performance exploiting multi-session data recorded by the same subject. Recently, several

studies have attempted to achieve better performance in MI classification using advanced deep learning. However, one of the challenging limitations of BCI is difficulty in recording sufficient data from a single subject. Additionally, due to the EEG recording method and characteristics of the endogenous paradigm, shifts in data occur. For these reasons, discrepancies in brain signals recorded in different sessions make it difficult to treat signals from different sessions and different subjects' brain signals as the same. As a result, the number of training data samples is limited. The lack of data samples leads to a complete but improper training causing serious problems, such as over- and under-fitting in BCI.

In this study, similarity between extracted features could be higher than between raw EEG data depending on the method. For similarity measurement, we used feature maps as images for measuring SSIM. Thus, if the target session (that we wish to classify) is set, then other sessions can be exploited as extra training data. SessionNet extracts the same size of feature maps from target session data and other session data for measuring similarity. As shown in Fig. 2, SessionNet extracts three feature maps from a session dataset. One is extracted by a shared block and the others are from two sub-networks. Before backpropagation, it calculates similarities between extracted feature maps from the same networks based on the SSIM. These indices are used in loss function as described in the loss function equation.

During training, the proposed model proceeds to classify target session data. A CNN assigned target session data trains only as much as the SSIM index using target session data and non-target session data whereas, each hierarchical CNN assigned non-target session data trains using only the assigned session data. As the training progresses, each CNN becomes biased to the assigned session data, causing a lower similarity index. Through several experiments, we empirically found that feature maps with too low value of an SSIM index adversely affect training. Therefore, we set a threshold SSIM index to avoid performance degradation. For our experimental dataset, SessionNet shows the best performance with a threshold SSIM index of 0.6. As shown in Fig. 3, the similarity indices between extracted feature maps are measured higher than similarity indices between raw EEG data. When the pattern representation of feature maps becomes apparent as the training progresses, the SSIM index is measured higher. Feature maps with SSIM index higher than the threshold are selected for additional training, and feature maps observed lower SSIM index than the threshold are excluded from the training process. As a result, SessionNet can learn target session data, and further training can be done using the non-target session datasets for extended generality.

3) LOSS FUNCTION

We used a weighted loss function based on the cross-entropy loss function [45]. SessionNet generates three separate terms of loss values from the shared block and sub-networks. At first, the shared block classifies input data to select sub-networks based on the probability of each class.

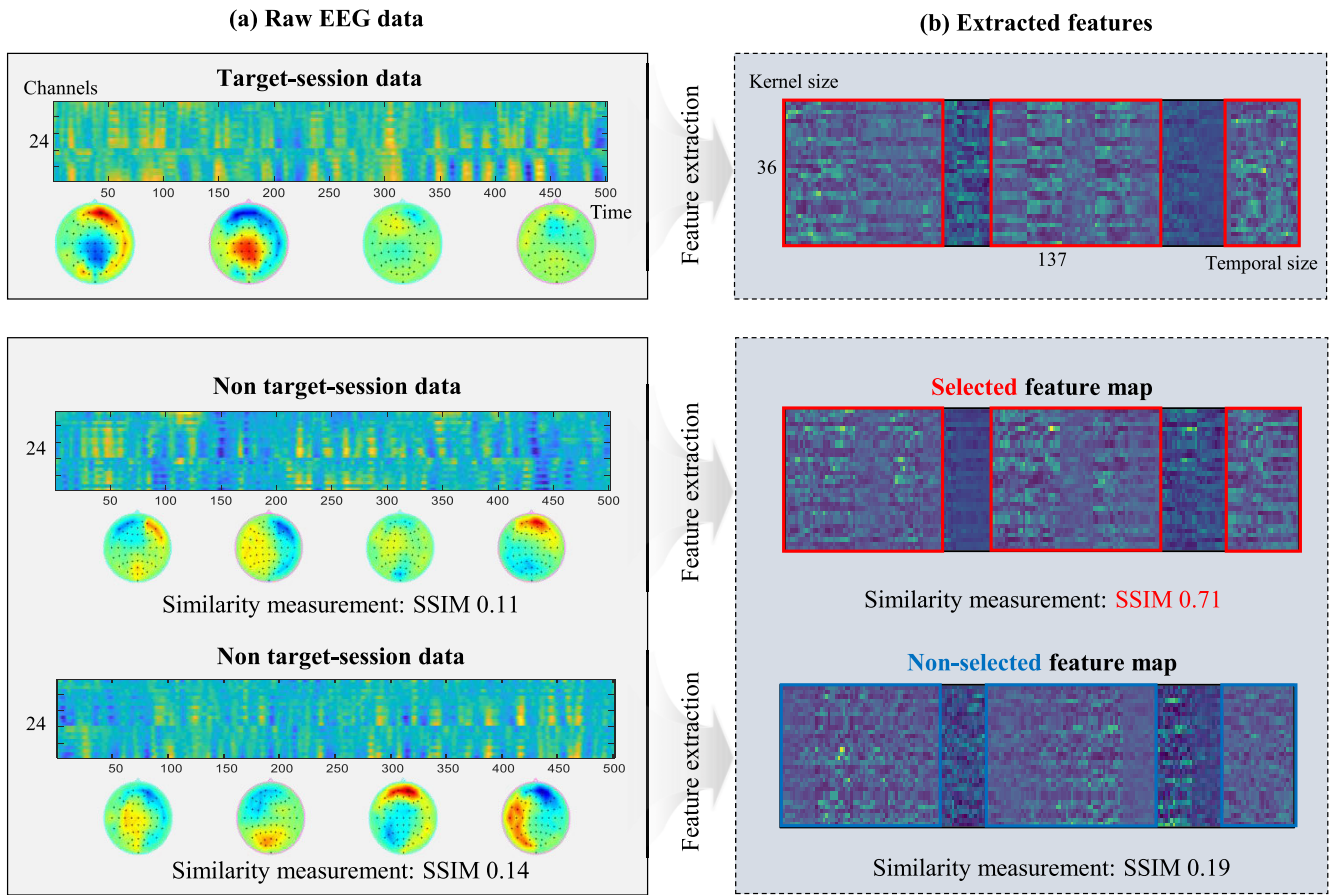


FIGURE 3. Visualized raw EEG data (a) and the shared feature maps (b). Through visualization, it has been confirmed that features can be extracted for high similarity unlike raw EEG data. The reference feature map used to measure similarity is a feature map extracted from the target session data (top). Feature maps with similarity index higher than a threshold value are included in the training (middle), and feature maps with an index lower than the threshold value are excluded from training (bottom).

Therefore, the probabilities are multiplied to each sub-network loss function term to consider the uncertainty of a prediction. All loss terms are summed to the loss function of the target session. Therefore, the loss function is defined as:

$$loss = \hat{L}_s + p_a \hat{L}_a + p_h \hat{L}_h \quad (1)$$

where p_a and p_h are probabilities of arm-reaching MI and hand-related MI (shared block output). L_s is a loss value generated by the shared block. L_a and L_h are the loss of the sub-networks assigned to arm-reaching MI and hand-related MI respectively. The modified loss function was designed based on the cross-entropy loss function. Details of \hat{L}_s , \hat{L}_a , and \hat{L}_h are as follows:

$$\hat{L}_s = L_{s.target} + \lambda_1 L_{s.non-target_1} + \lambda_2 L_{s.non-target_2} \quad (2)$$

$$\hat{L}_a = L_{a.target} + \lambda_1 L_{a.non-target_1} + \lambda_2 L_{a.non-target_2} \quad (3)$$

$$\hat{L}_h = L_{h.target} + \lambda_1 L_{h.non-target_1} + \lambda_2 L_{h.non-target_2} \quad (4)$$

where λ is the SSIM index between the feature maps of target session data and non-target session data. $L_{s.target}$ is the loss of a shared block assigned target session, and $L_{s.non-target}$ is the loss of the shared block assigned non-target sessions.

Likewise, $L_{a.target}$, $L_{h.target}$, $L_{h.non-target}$, and $L_{h.non-target}$ are loss values of target session and non-target sessions data generated by each sub-network. Details are as follows:

$$L_{s.target} = - \sum_{1}^3 y_s \log \bar{y}_{s.tar} \quad (5)$$

$$L_{s.non-target} = - \sum_{1}^3 y_s \log \bar{y}_{s.non} \quad (6)$$

$$L_{a.target} = - \sum_{1}^M y_a \log \bar{y}_{a.tar} \quad (7)$$

$$L_{a.non-target} = - \sum_{1}^M y_a \log \bar{y}_{a.non} \quad (8)$$

$$L_{h.target} = - \sum_{1}^N y_h \log \bar{y}_{h.tar} \quad (9)$$

$$L_{h.non-target} = - \sum_{1}^N y_h \log \bar{y}_{h.non} \quad (10)$$

where y_s is a label of a shared block. y_a and y_h are labels of the arm-reaching MI and hand-related MI. \bar{y}_s , \bar{y}_a and \bar{y}_h are outputs of the shared block, arm-related sub-network and hand-related sub-network, respectively. The number of arm-reaching MI and hand-related MI classes determines the parameters M and N .

III. RESULTS

Table 2 indicates the classification accuracies and standard deviations of each dataset. To avoid under and overfitting, 5-fold cross-validation was used as a statistical technique of evaluation. The performances of each target classification session were obtained using the target session data and the remaining two non-target session data as depicted in Fig. 2.

For 3-class classification, only shared convolution-pooling blocks performed classification because the dataset consists of three categorized classes. The highest accuracies were observed as 0.97, 0.95, and 0.93, respectively. The average accuracy of target session III was at least 0.07 points lower than the other target sessions. Target session I shows the highest averaged classification accuracy as 0.77 in the 5-class classification, whereas the highest averaged accuracy was recorded in target session III for 7-class horizontal classification. In the other datasets, the differences between average accuracies shrink to 0.01 points. Furthermore, in all target sessions, horizontal classifications show lower accuracies than vertical classifications. More specifically, 7-class vertical classification shows better results than a horizontal classification by up to 0.09 points in target session I. Interestingly, the performances of 9-class classifications were slightly better than 7-class classifications. We investigated statistical abnormality to figure out using confusion matrices.

Fig. 4 shows representative confusion matrices of SessionNet in the target session I task. All true-positive values recorded higher than true-negatives. In all classifications, SessionNet tended to confuse left and right arm-reaching MI with cylindrical and supination MI, meanwhile, the other classes show consistent accuracy. Specifically, in 3-class classification, SessionNet misclassified the resting-state as an arm-reaching MI task by 0.21 points. In 5-class classification, SessionNet misclassified an average 0.20 points of the left and right arm-reaching classes as cylindrical grasping and supination. Cylindrical grasping caused wrong predictions prominently in some arm-reaching MIs, such as left, right, and forward arm-reaching in 7-class datasets. As the number of classes increased, there was a tendency to confuse left and right arm-reaching MI with the other MI classes.

In addition, decoding resting-state is fundamental to achieving asynchronous BCI and furthering its practical uses. SessionNet properly classifies the resting-state, although it shows somewhat low accuracy in 7-class (HOR). Similar to left and right arm-reaching MI, the resting-state tends to be confused with cylindrical and supination MI.

Fig. 5 shows classification performance comparisons by using the different amount of non-target session dataset for training. As the amount of non-target session data increases,

TABLE 2. Classification performances using the SessionNet according to the classes types for all subjects.

(a) Target classification session: Session I						
Subjects	3-class	5-class	7-class (HOR)	7-class (VER)	9-class (HOR)	9-class (VER)
S1	0.97 ± 0.01	0.79 ± 0.02	0.62 ± 0.03	0.83 ± 0.02	0.83 ± 0.03	0.83 ± 0.02
S2	0.86 ± 0.01	0.82 ± 0.01	0.68 ± 0.02	0.81 ± 0.03	0.81 ± 0.01	0.83 ± 0.01
S3	0.87 ± 0.02	0.60 ± 0.01	0.66 ± 0.01	0.64 ± 0.06	0.69 ± 0.04	0.69 ± 0.05
S4	0.86 ± 0.03	0.77 ± 0.01	0.72 ± 0.02	0.79 ± 0.02	0.80 ± 0.01	0.79 ± 0.01
S5	0.90 ± 0.02	0.87 ± 0.01	0.73 ± 0.02	0.89 ± 0.01	0.87 ± 0.01	0.90 ± 0.01
S6	0.95 ± 0.03	0.88 ± 0.02	0.70 ± 0.02	0.83 ± 0.05	0.85 ± 0.03	0.87 ± 0.02
S7	0.95 ± 0.02	0.77 ± 0.02	0.68 ± 0.06	0.80 ± 0.03	0.74 ± 0.03	0.80 ± 0.02
S8	0.85 ± 0.02	0.78 ± 0.03	0.67 ± 0.01	0.68 ± 0.01	0.70 ± 0.01	0.73 ± 0.03
S9	0.83 ± 0.01	0.77 ± 0.03	0.67 ± 0.01	0.77 ± 0.01	0.72 ± 0.01	0.79 ± 0.01
S10	0.87 ± 0.01	0.75 ± 0.02	0.68 ± 0.02	0.75 ± 0.02	0.73 ± 0.04	0.78 ± 0.03
S11	0.83 ± 0.01	0.61 ± 0.04	0.64 ± 0.03	0.71 ± 0.02	0.71 ± 0.01	0.72 ± 0.03
S12	0.88 ± 0.02	0.79 ± 0.02	0.74 ± 0.01	0.77 ± 0.01	0.78 ± 0.03	0.82 ± 0.02
Avg.	0.89 ± 0.02	0.77 ± 0.02	0.68 ± 0.02	0.77 ± 0.02	0.77 ± 0.02	0.80 ± 0.02
(b) Target classification session: Session II						
Subjects	3-class	5-class	7-class (HOR)	7-class (VER)	9-class (HOR)	9-class (VER)
S1	0.95 ± 0.01	0.77 ± 0.01	0.78 ± 0.02	0.83 ± 0.01	0.80 ± 0.02	0.82 ± 0.01
S2	0.94 ± 0.01	0.78 ± 0.02	0.79 ± 0.08	0.79 ± 0.03	0.81 ± 0.01	0.81 ± 0.03
S3	0.86 ± 0.02	0.72 ± 0.01	0.71 ± 0.04	0.71 ± 0.02	0.74 ± 0.02	0.78 ± 0.01
S4	0.82 ± 0.03	0.70 ± 0.02	0.83 ± 0.04	0.74 ± 0.01	0.76 ± 0.02	0.81 ± 0.01
S5	0.89 ± 0.01	0.79 ± 0.03	0.75 ± 0.05	0.86 ± 0.01	0.85 ± 0.02	0.87 ± 0.03
S6	0.95 ± 0.02	0.86 ± 0.02	0.68 ± 0.09	0.84 ± 0.04	0.86 ± 0.01	0.88 ± 0.01
S7	0.93 ± 0.01	0.76 ± 0.03	0.60 ± 0.04	0.75 ± 0.02	0.80 ± 0.04	0.82 ± 0.01
S8	0.93 ± 0.04	0.59 ± 0.06	0.64 ± 0.06	0.77 ± 0.02	0.69 ± 0.01	0.74 ± 0.02
S9	0.83 ± 0.01	0.77 ± 0.03	0.67 ± 0.01	0.77 ± 0.01	0.72 ± 0.01	0.79 ± 0.01
S10	0.89 ± 0.01	0.73 ± 0.03	0.73 ± 0.01	0.75 ± 0.01	0.76 ± 0.02	0.78 ± 0.03
S11	0.81 ± 0.02	0.62 ± 0.02	0.67 ± 0.01	0.70 ± 0.02	0.71 ± 0.01	0.72 ± 0.02
S12	0.91 ± 0.02	0.77 ± 0.02	0.72 ± 0.01	0.75 ± 0.01	0.74 ± 0.01	0.75 ± 0.02
Avg.	0.90 ± 0.02	0.73 ± 0.02	0.71 ± 0.04	0.77 ± 0.02	0.77 ± 0.02	0.79 ± 0.02
(c) Target classification session: Session III						
Subjects	3-class	5-class	7-class (HOR)	7-class (VER)	9-class (HOR)	9-class (VER)
S1	0.88 ± 0.04	0.74 ± 0.01	0.73 ± 0.04	0.71 ± 0.02	0.83 ± 0.02	0.89 ± 0.02
S2	0.80 ± 0.03	0.79 ± 0.02	0.73 ± 0.05	0.79 ± 0.02	0.82 ± 0.02	0.83 ± 0.02
S3	0.82 ± 0.03	0.73 ± 0.03	0.68 ± 0.02	0.72 ± 0.01	0.73 ± 0.03	0.78 ± 0.04
S4	0.78 ± 0.06	0.71 ± 0.04	0.70 ± 0.02	0.74 ± 0.01	0.74 ± 0.02	0.73 ± 0.01
S5	0.93 ± 0.03	0.87 ± 0.01	0.86 ± 0.02	0.82 ± 0.03	0.84 ± 0.02	0.86 ± 0.01
S6	0.89 ± 0.02	0.83 ± 0.01	0.84 ± 0.03	0.81 ± 0.05	0.86 ± 0.03	0.88 ± 0.03
S7	0.89 ± 0.02	0.71 ± 0.04	0.68 ± 0.04	0.73 ± 0.03	0.75 ± 0.02	0.75 ± 0.01
S8	0.70 ± 0.03	0.54 ± 0.05	0.72 ± 0.01	0.80 ± 0.03	0.74 ± 0.01	0.73 ± 0.02
S9	0.64 ± 0.02	0.69 ± 0.01	0.66 ± 0.01	0.72 ± 0.02	0.75 ± 0.01	0.74 ± 0.02
S10	0.83 ± 0.01	0.75 ± 0.02	0.69 ± 0.03	0.71 ± 0.02	0.81 ± 0.03	0.84 ± 0.02
S11	0.78 ± 0.03	0.73 ± 0.01	0.67 ± 0.05	0.70 ± 0.04	0.74 ± 0.02	0.75 ± 0.01
S12	0.91 ± 0.01	0.84 ± 0.01	0.78 ± 0.02	0.81 ± 0.03	0.80 ± 0.01	0.84 ± 0.01
Avg.	0.82 ± 0.03	0.74 ± 0.02	0.73 ± 0.03	0.76 ± 0.03	0.78 ± 0.02	0.80 ± 0.02

consistent performance improvements were observed in both target session I and session II classifications. Compared to the experiment using only the target session dataset, the

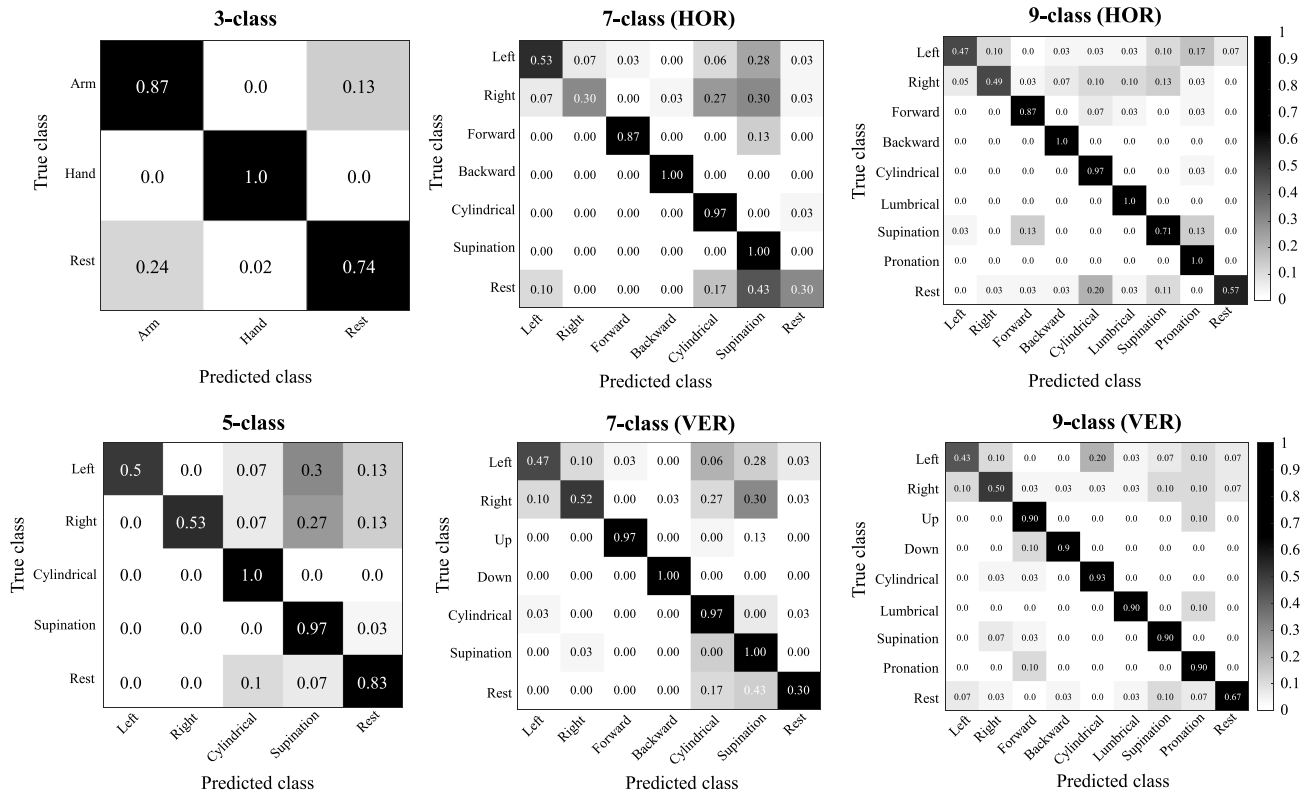


FIGURE 4. Representative confusion matrices using all subjects in session I dataset. All confusion matrices from each session data showed similar tendencies across all subjects.

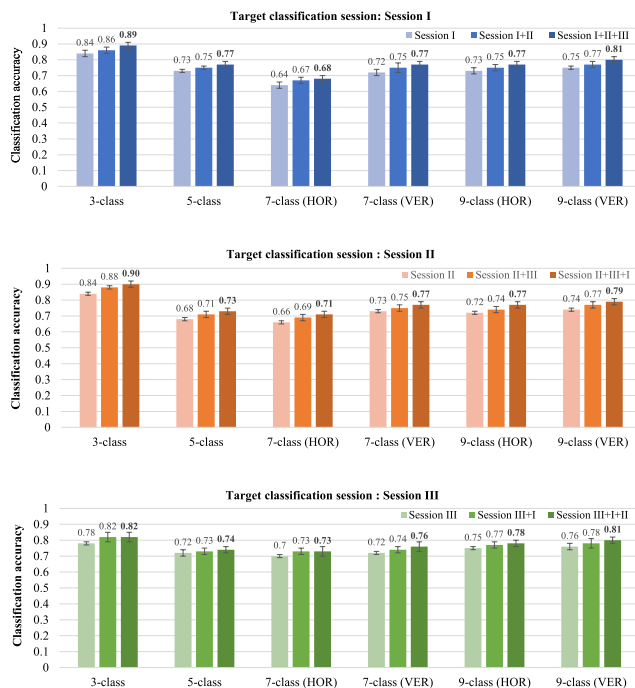


FIGURE 5. Comparison of classification accuracies according to the training session data.

performance increased up to 0.06 points. For target session III classification, non-target session data did not derive considerable performance improvement in all experiments. Though it

derived better performances in some experiments, the classification accuracies could not increase even where it exploited additional non-target session datasets in specific classifications (e.g., 3-class and 7-class classifications).

Table 3 indicates performance comparisons with existing methods. All experiments were conducted in the same computational environment. The FBCSP with the RLDA model is a commonly used machine learning method across all paradigms. The FBCSP+RLDA had the lowest accuracy of all experiments and under the condition of a large number of classification classes, the classification performance rapidly decreased close to the chance rate. The other CNN-based models show similar performances in 3-class and 5-class classification. In target session II and III classifications, DeepConvNet and ShallowConvNet recorded the same accuracies as SessionNet in 3-class classification (i.e., 0.91 and 0.81, respectively). As the number of classes increases, singular architecture models could not maintain competitive performance. Among these comparison models, ShallowConvNet and MCNN show slightly higher consistency than the others. However, ERA-CNN, designed as a hierarchical architecture, shows relatively robust performance in both 7-class and 9-class classifications. We designed SessionNet as a hierarchical CNN similar to the ERA-CNN. Additionally, ensemble learning was used to train more generality from other session data to handle unseen data. According to Table 3, SessionNet achieved a performance improvement up to 0.05 points compared to ERA-CNN. The differences in overall classification

TABLE 3. Classification performance comparison with existing methods and the SessionNet according to the target classification session.

Target classification session: Session I						
Methods	3-class	5-class	7-class (HOR)	7-class (VER)	9-class (HOR)	9-class (VER)
FBCSP+RLDA [25]	0.58 ± 0.04	0.42 ± 0.07	0.27 ± 0.05	0.32 ± 0.04	0.19 ± 0.02	0.24 ± 0.04
DeepConvNet [37]	0.85 ± 0.01	0.64 ± 0.05	0.57 ± 0.06	0.62 ± 0.07	0.63 ± 0.02	0.62 ± 0.01
ShallowConvNet [37]	0.87 ± 0.01	0.74 ± 0.05	0.61 ± 0.03	0.63 ± 0.07	0.66 ± 0.03	0.66 ± 0.03
EEGNet [30]	0.79 ± 0.04	0.70 ± 0.03	0.56 ± 0.07	0.55 ± 0.03	0.57 ± 0.06	0.56 ± 0.03
MCNN [39]	0.86 ± 0.05	0.76 ± 0.01	0.62 ± 0.03	0.65 ± 0.01	0.65 ± 0.03	0.66 ± 0.06
ERA-CNN [42]	0.85 ± 0.02	0.75 ± 0.03	0.65 ± 0.10	0.74 ± 0.03	0.76 ± 0.05	0.75 ± 0.01
SessionNet	0.89 ± 0.02	0.77 ± 0.02	0.68 ± 0.02	0.77 ± 0.02	0.77 ± 0.02	0.80 ± 0.02

Target classification session: Session II						
Methods	3-class	5-class	7-class (HOR)	7-class (VER)	9-class (HOR)	9-class (VER)
FBCSP+RLDA [25]	0.56 ± 0.02	0.46 ± 0.04	0.27 ± 0.02	0.31 ± 0.07	0.23 ± 0.05	0.29 ± 0.02
DeepConvNet [37]	0.91 ± 0.01	0.65 ± 0.04	0.63 ± 0.10	0.62 ± 0.03	0.64 ± 0.03	0.65 ± 0.08
ShallowConvNet [37]	0.86 ± 0.05	0.69 ± 0.03	0.63 ± 0.04	0.62 ± 0.04	0.66 ± 0.02	0.63 ± 0.01
EEGNet [30]	0.72 ± 0.03	0.64 ± 0.04	0.58 ± 0.04	0.57 ± 0.05	0.52 ± 0.05	0.52 ± 0.02
MCNN [39]	0.85 ± 0.01	0.71 ± 0.03	0.65 ± 0.01	0.64 ± 0.03	0.66 ± 0.01	0.67 ± 0.02
ERA-CNN [42]	0.87 ± 0.05	0.72 ± 0.03	0.70 ± 0.04	0.72 ± 0.03	0.73 ± 0.02	0.74 ± 0.01
SessionNet	0.91 ± 0.04	0.73 ± 0.02	0.71 ± 0.04	0.77 ± 0.02	0.77 ± 0.02	0.79 ± 0.02

Target classification session: Session III						
Methods	3-class	5-class	7-class (HOR)	7-class (VER)	9-class (HOR)	9-class (VER)
FBCSP + RLDA [25]	0.44 ± 0.05	0.40 ± 0.02	0.22 ± 0.02	0.25 ± 0.03	0.18 ± 0.01	0.20 ± 0.01
DeepConvNet [37]	0.79 ± 0.04	0.64 ± 0.07	0.60 ± 0.06	0.61 ± 0.07	0.64 ± 0.07	0.65 ± 0.03
ShallowConvNet [37]	0.81 ± 0.06	0.70 ± 0.07	0.64 ± 0.04	0.63 ± 0.06	0.63 ± 0.07	0.67 ± 0.03
EEGNet [30]	0.71 ± 0.01	0.65 ± 0.02	0.58 ± 0.05	0.59 ± 0.10	0.57 ± 0.03	0.57 ± 0.02
MCNN [39]	0.74 ± 0.02	0.66 ± 0.03	0.66 ± 0.05	0.60 ± 0.02	0.65 ± 0.04	0.69 ± 0.02
ERA-CNN [42]	0.77 ± 0.01	0.68 ± 0.03	0.68 ± 0.02	0.72 ± 0.03	0.73 ± 0.05	0.76 ± 0.03
SessionNet	0.82 ± 0.03	0.74 ± 0.02	0.73 ± 0.03	0.76 ± 0.03	0.78 ± 0.02	0.80 ± 0.02

accuracy of the proposed SessionNet over the existing methods were found to be significant using a paired t -test (p -value < 0.01) [29].

Many studies have been done using only single session data to avoid the EEG non-stationary problem. Calculating the epoch until the model reaches convergence is one of the important indicators whether the training is well completed or not. Therefore, if a simple or a sophisticated model can converge the training within a few iterations, it shows that the model learns more efficiently exploiting the given data. Fig. 6 indicates the epoch that obtained the best accuracy within a subject in target session II, for 9-class classification. SessionNet derived the highest accuracy before 100 epochs iterations in five subjects. The averaged epoch for SessionNet was 115.25 and the minimum value was 76 epochs.

The second-ranked model is ERA-CNN and its mean epoch was 152.58 epochs. In terms of accuracy, ERA-CNN showed slightly lower accuracies than SessionNet. However, the averaged epoch of SessionNet was 37.33 points lower than ERA-CNN. Accordingly, we confirmed that SessionNet converged the training substantially faster than ERA-CNN. ShallowConvNet shows almost 10 points lower mean epoch than the other CNN-based models (154.25). Even though it consisted of only three layers, it converges the training quickly and shows competitive performance. DeepConvNet, EEGNet, and MCNN take a similar number of iterations to converge the training, 163.08, 161.67, and 165, respectively.

Fig. 7 shows the representative training and test losses along with the iterations (epochs) using a 5-class classification dataset. Loss values were calculated to obtain the

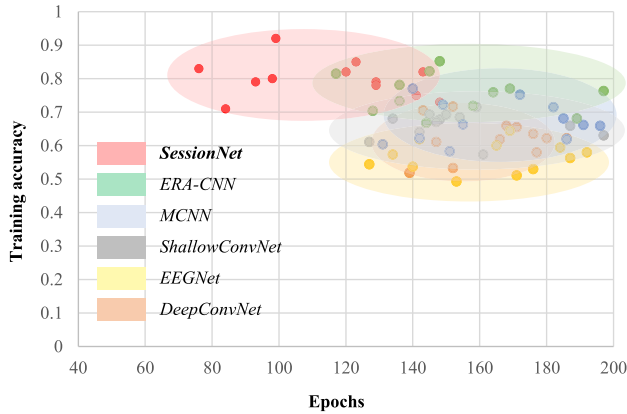


FIGURE 6. Training accuracy among the conventional methods according to the number of epochs in 9-class classification.

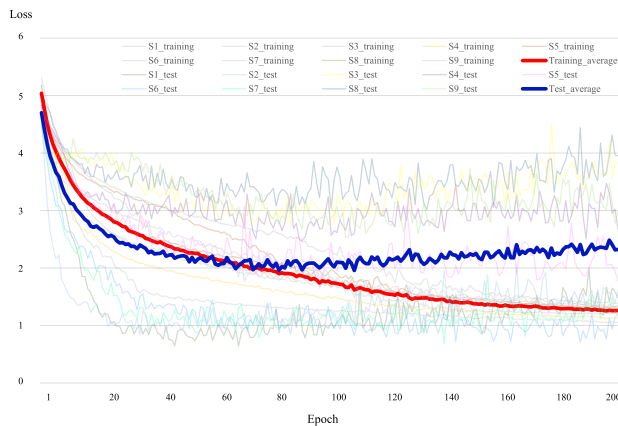


FIGURE 7. Convergence curves of training and test losses across all subjects. The loss values were calculated using a 5-class classification dataset where the target session II.

optimized weights of networks during the training. The optimizer updates the model parameters to minimize the losses. Training loss values converged as the training goes on. On the other hand, the test loss values had decreased until around 100 epochs, but the loss values had since increased due to overfitting.

IV. DISCUSSION

In this study, we investigated the possibility of robust multi-class classification performance related to intuitive MI using a single upper-extremity. We performed MI classification using a deep learning strategy across healthy subjects. We also presented a novel approach for robust MI classification. We proposed the SessionNet architecture, based on feature similarity-based weighted ensemble learning as depicted in Fig. 2. To evaluate the proposed method, we collected a diverse intuitive EEG data samples from a single upper-extremity, such as arm-reaching MI (6-class), hand-grasping MI (2-class), and wrist-twisting MI (2-class). The experimental results showed that SessionNet could achieve high classification accuracies in multi-class classifications. This result

demonstrates that learning generality from other session data was effective in intuitive MI classifications. Hence, deep learning is an important method in the development of BCI.

Recent BCIs have adopted various deep learning techniques such as ERP detection [46], [47], mental state detection [48]–[50], and MI classification [31], [37], [51], [52]. Especially, for MI decoding, some groups have tried to adopt deep learning techniques with brain signals. Schirrneister *et al.* [37] designed the shallow and deep depth of CNN architecture (i.e., ShallowConvNet and DeepConvNet). The proposed ConvNets are well suited to most EEG paradigms and have proven to have high classification performance on raw EEG data. Xu *et al.* [31] also proposed a transfer learning framework for EEG classification. They showed that their framework could improve accuracy and efficiency compared to the traditional methods. Tayeb *et al.* [52] validated MI classification using a variety of deep neural networks and conventional machine learning techniques through online experiments. They demonstrated that deep learning techniques can have robust decoding performance in real-time BCIs. Lei *et al.* [51] proposed a multi-view multi-level deep polynomial network (MMDPN) for decoding walking imagery. They compared the decoding performance using a virtual environment paradigm and traditional text-based paradigm. MMDPN outperformed other deep learning methods in terms of classification accuracy in both paradigms. Jiao *et al.* [53] proposed a novel sparse group representation model (SGRM) that extracts features using CSP. They explicitly exploited within-group sparse and group-wise sparse constraints. Through this method, SGRM proved that it can reduce the required training samples from the target subject effectively. Zhang *et al.* [54] also proposed a temporally constrained sparse group spatial pattern (TSGSP) using CSP-based spatial filtering to classify MI data. For better performance, TSGSP optimizes filter bands and time windows simultaneously within the CSP. Recently, intuitive MI decoding has been investigated for natural interactions between humans and devices. Schwarz *et al.* [55] decoded hand movements such as palmar grasping, lateral grasping, and supination using MRCP. They evaluated MRCP decoding in a simulated environment and demonstrated the possibility of robotic arm control using EEG signals. Jeong *et al.* [41] also proved the feasibility of a brain-controlled robotic arm system based on intuitive MI using a multi-directional CNN-Bidirectional long-short term memory (BiLSTM) network. Furthermore, they have tried to decode complex kinematics information such as forearm rotation from EEG signals. They proposed a hierarchical flow CNN and evaluated the decoding performance using not only their experimental dataset but also a public dataset (BNCI Horizon 2020) [29]. Xu *et al.* [56] have experienced MI tasks, such as hand grip, forearm extension/flexion, reach-and-grasp. They confirmed that phase synchronization information is effective in discriminating between different MI tasks with the same limbs.

Our proposed approach can produce well-trained networks because it can extract high-level features from different sessions for the same tasks and for the same subjects, and it could be applied as a data augmentation strategy. Most groups studying conventional data augmentation methods in BCIs have focused on data augmentation within a trial [28], [57]. As a result, the conventional methods confirmed the performance enhancement, but the data length for time-series analysis tends to reflect user intention insufficiently because of the short epochs of EEG data (e.g., 2-second [30], [37]). Furthermore, singular models classify entire classes at once, in contrast the proposed SessionNet first categorizes input data into arm-reaching MI, hand-related MI, and the resting-state. Hence, it performs a series of a small number of classifications (e.g., 3-class to 4-class).

Additionally, SessionNet learns more generality from other session data based on feature similarity for improving classification performance. SessionNet can use more data for training and this training strategy can explain the increase in performance. Therefore, we focused on enhancing the classification performance of the target session using multiple session data learning for more generality. To avoid the non-stationarity EEG problem, we strictly designed the experimental protocols and obtained the EEG data in each session under the same experimental conditions. Our experimental results confirmed that the proposed SessionNet using feature similarity was more effective for multi-class MI classification and consisted of complex tasks than conventional approaches. Additionally, we conducted a statistical analysis of classification performance between the conventional methods and the proposed method. We confirmed the statistically significant difference between the methods ($p < 0.01$). Hence, we proved the superiority and novelty of the proposed network.

However, the SessionNet tends to confuse the resting-state with hand-related MI (e.g., cylindrical grasping and supination). When the subjects generate brain signals, it is necessary to imagine the specific muscle movements. That is, we conjecture that subjects might find it difficult to imagine small muscle movements, such as in cylindrical grasping so that the discriminative brain areas are not activated and it appears similar to resting-state. To overcome this, the proposed SessionNet will be modified to consider the relationship between learned features and brain functional activities during MI. It is one of the advanced strategies to design a brain-inspired network. The deep learning in BCI studies has been designed to learn features from brain activities. Therefore, various kinds of filters as temporal, spatial, and spectral aspects have been developed for including information on brain activity. Furthermore, we will demonstrate the relationship between the learned features and the brain functional network by adding an interpretation function on the network.

In addition, we compared the computational time for classification, which is critical in real-time BCI scenarios. We measured the total elapsed time for model training,

validation, and testing [17]. On average, the elapsed time of SessionNet was approximately 28 s, based on 200 epochs during all procedures with a high-performance computer (i.e., Intel i7 CPU, a 64-GB RAM, 1-TB SSD, and a TITAN XP GPU). Achieving competitive classification performance is a critical issue at the same time, a short computational time is also essential for real-time BCI scenarios. Especially, the MI-based BCI paradigm has endogenous characteristics that generate the EEG patterns by their respective subjects. Therefore, compared with other BCI paradigms such as exogenous potentials (e.g., SSVEP and ERP), it is difficult to obtain exact EEG patterns owing to various external and internal factors like subjects' mental conditions and attention level. Long calibration time would induce fatigue and distraction to the subject and it might decrease the BCI performance in real-time scenarios [58]. In this point of view, the proposed SessionNet was designed to utilize the EEG signals from the different recording sessions. We expect that calibration time can be reduced within a day using the SessionNet.

V. CONCLUSION AND FUTURE WORKS

In this paper, we proposed SessionNet that trains more generality using data recorded multiple sessions based on ensemble learning and a feature similarity. Additionally, SessionNet adopted the principle of a hierarchical architecture that can extract features from different body regions such as the arm and hand of a single-arm. In applying a deep learning technique, one of the challenging tasks is recording sufficient EEG data from a single subject. Furthermore, MI classification performance depends on the quality of data. With low-quality data, there is a limit to improving performance by changing only filter sizes or designing more layers. We demonstrated that SessionNet achieved meaningful improvement in MI classification performance by training more generality from multiple sessions. This enhancement showed the possibility of using data from different sessions with a proper similarity index for improving classification performance.

Hence, in future work, we will develop SessionNet to control the external devices intuitively, such as robotic arms, with a high degree of freedom. Therefore, we will demonstrate that the proposed SessionNet could ultimately help improve the autonomy of people with movement disabilities and support the daily life of healthy people. Furthermore, at present, the SessionNet could be considered only MI domain but we will plan to expand to the domain-independent network applicable to different types of brain signals.

ACKNOWLEDGMENT

The authors thank Prof. C. Guan and Prof. D.-J. Kim for their useful discussion of the data analysis and Mr. J.-H. Cho, Mr. B.-H. Kwon, and Ms. D.-Y. Lee for the help with EEG database construction. (*Byeong-Hoo Lee and Ji-Hoon Jeong contributed equally to this work.*)

REFERENCES

- [1] T. M. Vaughan, W. J. Heetderks, L. J. Trejo, W. Z. Rymer, M. Weinrich, M. M. Moore, A. Kübler, B. H. Dobkin, N. Birbaumer, E. Donchin, E. W. Wolpaw, and J. R. Wolpaw, "Brain-computer interface technology: A review of the second international meeting," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 2, pp. 94–109, Jun. 2003.
- [2] J. Meng, S. Zhang, A. Bekyo, J. Olsoe, B. Baxter, and B. He, "Noninvasive electroencephalogram based control of a robotic arm for reach and grasp tasks," *Sci. Rep.*, vol. 6, no. 1, p. 38565, Dec. 2016.
- [3] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "BCI2000: A general-purpose brain-computer interface (BCI) system," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1034–1043, Jun. 2004.
- [4] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review," *Sensors*, vol. 12, no. 2, pp. 1211–1279, 2012.
- [5] C. I. Penalzoza and S. Nishio, "BMI control of a third arm for multitasking," *Sci. Robot.*, vol. 3, no. 20, Jul. 2018, Art. no. eaat1228.
- [6] 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.
- [7] Y. Chen, A. D. Atnafu, I. Schlattner, W. T. Weldtsadik, M.-C. Roh, H. J. Kim, S.-W. Lee, B. Blankertz, and S. Fazli, "A high-security EEG-based login system with RSVP stimuli and dry electrodes," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 12, pp. 2635–2647, Dec. 2016.
- [8] 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.
- [9] Z. Lin, C. Zhang, W. Wu, and X. Gao, "Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 12, pp. 2610–2614, Dec. 2006.
- [10] I. K. Niazi, N. Jiang, O. Tiberghien, J. F. Nielsen, K. Dremstrup, and D. Farina, "Detection of movement intention from single-trial movement-related cortical potentials," *J. Neural Eng.*, vol. 8, no. 6, Oct. 2011, Art. no. 066009.
- [11] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication," *Proc. IEEE*, vol. 89, no. 7, pp. 1123–1134, Jul. 2001.
- [12] D. J. Krusienski, E. W. Sellers, F. Cabestaing, S. Bayouduh, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "A comparison of classification techniques for the P300 speller," *J. Neural Eng.*, vol. 3, no. 4, p. 299, 2006.
- [13] M.-H. Lee, J. Williamson, D.-O. Won, S. Fazli, and S.-W. Lee, "A high performance spelling system based on EEG-EOG signals with visual feedback," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 7, pp. 1443–1459, Jul. 2018.
- [14] 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.
- [15] N.-S. Kwak, K.-R. Müller, and S.-W. Lee, "A lower limb exoskeleton control system based on steady state visual evoked potentials," *J. Neural Eng.*, vol. 12, no. 5, Oct. 2015, Art. no. 056009.
- [16] F. Karimi, J. Kofman, N. Mrachacz-Kersting, D. Farina, and N. Jiang, "Detection of movement related cortical potentials from EEG using constrained ICA for brain-computer interface applications," *Frontiers Neurosci.*, vol. 11, p. 356, Jun. 2017.
- [17] J.-H. Jeong, N.-S. Kwak, C. Guan, and S.-W. Lee, "Decoding movement-related cortical potentials based on subject-dependent and section-wise spectral filtering," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 3, pp. 687–698, Mar. 2020.
- [18] C. Neuper, M. Wörtz, and G. Pfurtscheller, "ERD/ERS patterns reflecting sensorimotor activation and deactivation," *Progr. Brain Res.*, vol. 159, pp. 211–222, Jan. 2006.
- [19] B. J. Edelman, J. Meng, D. Suma, C. Zurn, E. Nagarajan, B. S. Baxter, C. C. Cline, and B. He, "Noninvasive neuroimaging enhances continuous neural tracking for robotic device control," *Sci. Robot.*, vol. 4, no. 31, Jun. 2019, Art. no. eaaw6844.
- [20] M. Alhussein, G. Muhammad, and M. S. Hossain, "EEG pathology detection based on deep learning," *IEEE Access*, vol. 7, pp. 27781–27788, 2019.
- [21] H.-I. Suk and S.-W. Lee, "Subject and class specific frequency bands selection for multiclass motor imagery classification," *Int. J. Imag. Syst. Technol.*, vol. 21, no. 2, pp. 123–130, Jun. 2011.
- [22] S. R. Liyanage, C. Guan, H. Zhang, K. K. Ang, J. Xu, and T. H. Lee, "Dynamically weighted ensemble classification for non-stationary EEG processing," *J. Neural Eng.*, vol. 10, no. 3, Jun. 2013, Art. no. 036007.
- [23] 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, no. 8, pp. 2725–2737, Aug. 2015.
- [24] X. Zhang, L. Yao, X. Wang, J. Monaghan, D. Mcalpine, and Y. Zhang, "A survey on deep learning based brain computer interface: Recent advances and new frontiers," 2019, *arXiv:1905.04149*. [Online]. Available: <http://arxiv.org/abs/1905.04149>
- [25] K. Keng Ang, Z. Yang Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Jun. 2008, pp. 2390–2397.
- [26] T.-E. Kam, H.-I. Suk, and S.-W. Lee, "Non-homogeneous spatial filter optimization for ElectroEncephaloGram (EEG)-based motor imagery classification," *Neurocomputing*, vol. 108, pp. 58–68, May 2013.
- [27] A. Subasi and M. Ismail Gursoy, "EEG signal classification using PCA, ICA, LDA and support vector machines," *Expert Syst. Appl.*, vol. 37, no. 12, pp. 8659–8666, Dec. 2010.
- [28] G. Dai, J. Zhou, J. Huang, and N. Wang, "HS-CNN: A CNN with hybrid convolution scale for EEG motor imagery classification," *J. Neural Eng.*, vol. 17, no. 1, Jan. 2020, Art. no. 016025.
- [29] J.-H. Jeong, B.-H. Lee, D.-H. Lee, Y.-D. Yun, and S.-W. Lee, "EEG classification of forearm movement imagery using a hierarchical flow convolutional neural network," *IEEE Access*, vol. 8, pp. 66941–66950, 2020.
- [30] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, Oct. 2018, Art. no. 056013.
- [31] G. Xu, X. Shen, S. Chen, Y. Zong, C. Zhang, H. Yue, M. Liu, F. Chen, and W. Che, "A deep transfer convolutional neural network framework for EEG signal classification," *IEEE Access*, vol. 7, pp. 112767–112776, 2019.
- [32] P. Wang, A. Jiang, X. Liu, J. Shang, and L. Zhang, "LSTM-based EEG classification in motor imagery tasks," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 11, pp. 2086–2095, Nov. 2018.
- [33] S. Alhagry, A. A. Fahmy, and R. A. El-Khoribi, "Emotion recognition based on EEG using LSTM recurrent neural network," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 10, pp. 355–358, 2017.
- [34] D. Freer and G.-Z. Yang, "Data augmentation for self-paced motor imagery classification with C-LSTM," *J. Neural Eng.*, vol. 17, no. 1, Jan. 2020, Art. no. 016041.
- [35] 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.
- [36] R. H. Eleessawy, S. Eldawlatly, and H. M. Abbas, "A long short-term memory autoencoder approach for EEG motor imagery classification," in *Proc. Int. Conf. Comput., Autom. Knowl. Manage. (ICCAKM)*, Jan. 2020, pp. 79–84.
- [37] R. T. Schirrmester, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggenberger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Human Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, Nov. 2017.
- [38] A. M. Azab, L. Mihaylova, K. Keng Ang, and M. Arvaneh, "Weighted transfer learning for improving motor imagery-based brain-computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 7, pp. 1352–1359, Jul. 2019.
- [39] S. U. Amin, M. Alsulaiman, G. Muhammad, M. A. Mekhtiche, and M. S. Hossain, "Deep learning for EEG motor imagery classification based on multi-layer CNNs feature fusion," *Future Gener. Comput. Syst.*, vol. 101, pp. 542–554, Dec. 2019.
- [40] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [41] J.-H. Jeong, K.-H. Shim, D.-J. Kim, and S.-W. Lee, "Brain-controlled robotic arm system based on multi-directional CNN-BiLSTM network using EEG signals," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 5, pp. 1226–1238, May 2020.
- [42] B.-H. Lee, J.-H. Jeong, K.-H. Shim, and S.-W. Lee, "Classification of high-dimensional motor imagery tasks based on an end-to-end role assigned convolutional neural network," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 1359–1363.
- [43] G. I. Webb and Z. Zheng, "Multistrategy ensemble learning: Reducing error by combining ensemble learning techniques," *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 8, pp. 980–991, Aug. 2004.

- [44] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (ELUs)," 2015, *arXiv:1511.07289*. [Online]. Available: <http://arxiv.org/abs/1511.07289>
- [45] Z. Zhang and M. Sabuncu, "Generalized cross entropy loss for training deep neural networks with noisy labels," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2018, pp. 8778–8788.
- [46] J. Li, Z. L. Yu, Z. Gu, W. Wu, Y. Li, and L. Jin, "A hybrid network for ERP detection and analysis based on restricted Boltzmann machine," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 3, pp. 563–572, Mar. 2018.
- [47] A. Dittaphon, N. Banluesombatkul, S. Kettrat, E. Chuangsuwanich, and T. Wilaiprasitporn, "Universal joint feature extraction for P300 EEG classification using multi-task autoencoder," *IEEE Access*, vol. 7, pp. 68415–68428, 2019.
- [48] J.-H. Jeong, B.-W. Yu, D.-H. Lee, and S.-W. Lee, "Classification of drowsiness levels based on a deep spatio-temporal convolutional bidirectional LSTM network using electroencephalography signals," *Brain Sci.*, vol. 9, no. 12, p. 348, Nov. 2019.
- [49] P. Zhang, X. Wang, W. Zhang, and J. Chen, "Learning spatial–spectral–temporal EEG features with recurrent 3D convolutional neural networks for cross-task mental workload assessment," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 1, pp. 31–42, Jan. 2019.
- [50] Z. Jiao, X. Gao, Y. Wang, J. Li, and H. Xu, "Deep convolutional neural networks for mental load classification based on EEG data," *Pattern Recognit.*, vol. 76, pp. 582–595, Apr. 2018.
- [51] B. Lei, X. Liu, S. Liang, W. Hang, Q. Wang, K.-S. Choi, and J. Qin, "Walking imagery evaluation in brain computer interfaces via a multi-view multi-level deep polynomial network," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 3, pp. 497–506, Mar. 2019.
- [52] Z. Tayeb, J. Fedjaev, N. Ghaboosi, C. Richter, L. Everding, X. Qu, Y. Wu, G. Cheng, and J. Conradt, "Validating deep neural networks for online decoding of motor imagery movements from EEG signals," *Sensors*, vol. 19, no. 1, p. 210, Jan. 2019.
- [53] Y. Jiao, Y. Zhang, X. Chen, E. Yin, J. Jin, X. Wang, and A. Cichocki, "Sparse group representation model for motor imagery EEG classification," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 2, pp. 631–641, Mar. 2019.
- [54] Y. Zhang, C. S. Nam, G. Zhou, J. Jin, X. Wang, and A. Cichocki, "Temporally constrained sparse group spatial patterns for motor imagery BCI," *IEEE Trans. Cybern.*, vol. 49, no. 9, pp. 3322–3332, Sep. 2019.
- [55] A. Schwarz, M. K. Höller, J. Pereira, P. Ofner, and G. R. Müller-Putz, "Decoding hand movements from human EEG to control a robotic arm in a simulation environment," *J. Neural Eng.*, vol. 17, no. 3, May 2020, Art. no. 036010.
- [56] B. Xu, Z. Wei, A. Song, C. Wu, D. Zhang, W. Li, G. Xu, H. Li, and H. Zeng, "Phase synchronization information for classifying motor imagery EEG from the same limb," *IEEE Access*, vol. 7, pp. 153842–153852, 2019.
- [57] Z. Zhang, F. Duan, J. Solé-Casals, J. Dinarès-Ferran, A. Cichocki, Z. Yang, and Z. Sun, "A novel deep learning approach with data augmentation to classify motor imagery signals," *IEEE Access*, vol. 7, pp. 15945–15954, 2019.
- [58] A. Singh, S. Lal, and H. Guesgen, "Reduce calibration time in motor imagery using spatially regularized symmetric positive-definite matrices based classification," *Sensors*, vol. 19, no. 2, p. 379, Jan. 2019.



BYEONG-HOO LEE received the B.S. degree in electronic engineering from Hanyang University, South Korea, in 2019. He is currently pursuing the master's degree with the Department of Brain and Cognitive Engineering, Korea University, South Korea. His research interests include deep learning, brain-computer interfaces, and signal processing.



JI-HOON JEONG received the B.S. degree in computer information science from Korea University, South Korea, in 2015, where he is currently pursuing the Ph.D. degree with the Department of Brain and Cognitive Engineering. His research interests include machine learning, deep learning, and brain-machine interfaces.



SEONG-WHAN LEE (Fellow, IEEE) received the B.S. degree in computer science and statistics from Seoul National University, South Korea, in 1984, and the M.S. and Ph.D. degrees in computer science from the Korea Advanced Institute of Science and Technology, 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. He is a Fellow of the International Association of Pattern Recognition and the Korea Academy of Science and Technology.

• • •