

Digital Object Identifier

Continuous EEG Decoding of Pilots' Mental States using Multiple Feature Block-based Convolutional Neural Network

DAE-HYEOK LEE^{1,†}, JI-HOON JEONG^{1,†}, KIDUK KIM^{1,2,†}, BAEK-WOON YU¹, AND SEONG-WHAN LEE³, (FELLOW, IEEE)

¹Department of Brain and Cognitive Engineering, Korea University, Anam-dong, Seongbuk-ku, Seoul 02841, South Korea (e-mail: lee_dh@korea.ac.kr, jh_jeong@korea.ac.kr, dukekim@hanmail.net, bw_yu@korea.ac.kr)

²R&D Center, LG Display, Seoul, 07796, South Korea

³Department of Artificial Intelligence, Korea University, Anam-dong, Seongbuk-ku, Seoul 02841, South Korea

[†]These authors contributed equally to this work.

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

This work was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (No. 2017-0-00451, No. 2019-0-00079) and partly supported by Defense Acquisition Program Administration (DAPA) and Agency for Defense Development (ADD) of Korea (06-201-305-001, A Study on Human-Computer Interaction Technology for the Pilot Status Recognition).

ABSTRACT Non-invasive brain-computer interface (BCI) has been developed for recognizing and classifying human mental states with high performances. Specifically, classifying pilots' mental states accurately is a critical issue because their cognitive states, which are induced by mental fatigue, workload, and distraction, may be fundamental in catastrophic accidents. In this study, we present an electroencephalogram (EEG) classification of four mental states (fatigue, workload, distraction, and the normal state) from EEG signals in both offline and pseudo-online analyses. To the best of our knowledge, this study is the first attempt to classify pilots' mental states using only EEG signals during continuous decoding. We recorded EEG signals from seven pilots under various simulated flight conditions. We proposed a multiple feature block-based convolutional neural network (MFB-CNN) with temporal-spatio EEG filters to recognize the pilot's current mental states. We validated the proposed method for two analyses across all subjects. In the offline analysis, we confirmed the classification accuracy of 0.75 (± 0.04) and, in the pseudo-online analysis, we obtained the detection accuracy of 0.72 (± 0.20), 0.72 (± 0.27), and 0.61 (± 0.18) for fatigue, workload, and distraction, respectively. Hence, we demonstrate the feasibility of classifying various types of mental states for implementation in real-world environments.

INDEX TERMS Brain-computer interface (BCI), Electroencephalogram (EEG), Mental states, Deep convolutional neural network.

I. INTRODUCTION

BRAIN-COMPUTER interface (BCI) is one of the innovative technologies used for communication between humans and devices by recognizing users' status and intention. In particular, non-invasive BCI, one of practical BCI techniques, has attracted a lot of attention owing to its non-invasive nature and low cost [1]–[7]. Hence, non-invasive BCI systems to control a robotic arm [8]–[10], a wheelchair [11], and a BCI speller [5], [12], [13] have been developed. Recently, the non-invasive BCI has been investigated for recognizing human mental states, such as fatigue [14]–[18],

workload [14], [19]–[21], and distraction [22]–[24], with robust detection performances. Especially, classification of each state such as fatigue or not, workload or not, and distraction or not for autonomous driving systems (e.g., vehicles and aircraft) using physiological signals has been developed for a high level of artificial intelligence (AI) techniques [15], [25]. Hence, recent BCI advances have been focused on the classification of the users' mental state because it can directly reflect the actual condition of human mental states [26], [27].

Fatigue is commonly caused by a prolonged cognitive task, especially a repetitive or boring task [28], [29]. According

to the British airline pilots' association (BALPA), 56% of the approximately 500 commercial pilots responded that they usually fell asleep during a flight. In addition, mental workload is defined as the required mental cost to achieve a given task [14]. Task difficulty and mental workload of pilots could be influenced by various interruptions during flight journey. The division of national aeronautics and space administration (NASA) for human factors in aerospace published research indicating that as the pilot's workload increased, the ability to deal with flight-based information decreased [30]. Also, distraction occurs when pilots could not concentrate on the flight task. Civil aviation authority (CAA) of New Zealand reported that distraction during flight is the major reason for errors leading to accidents [31].

Recent BCI studies have investigated the classification of each state using various physiological signals, such as those captured by electrooculogram (EOG), electrocardiogram (ECG), functional magnetic resonance imaging [32], functional near-infrared spectroscopy (fNIRS), and electroencephalogram (EEG) signals. Hong *et al.* [33] detected drowsy state utilizing machine learning methods with various physiological signals. The highest performance was 0.98 kappa values using random forest method. Liu *et al.* [19] used EEG signals and fNIRS recorded during the n-back task to evaluate workload. The highest accuracy of 65.1% was observed when fusing all information. Horrey *et al.* [23] recognized distraction using near-infrared spectroscopy, heart monitoring, and eye-tracking systems. They obtained the highest accuracy of 93.3% when unheard material was rejected, and this method helped across 31 participants.

A few groups have conducted experiments using only EEG signals for classifying each mental state. Gao *et al.* [15] accurately detected fatigue utilizing a spatial-temporal convolutional neural network (CNN) model from EEG signals. They constructed a deep artificial neural network with 14 layers and used a relatively small number of EEG data while maintaining high performances. In addition, the discriminative point compared to other studies was that the number of parameters could be reduced by sharing the weight parameters among core blocks, and this was the critical factor for improving performances. Blanco *et al.* [20] evaluated workload using EEG signals. They compared binary classification using k-nearest neighbor (kNN), linear discriminant analysis (LDA), naïve Bayes, decision trees, and support vector machine (SVM), and the highest accuracy of 90.17% was observed using LDA. Sonnleitner *et al.* [22] recognized distraction from EEG signals using regularized LDA. They used two tasks for their experiments: a driving task and auditory secondary task. They achieved approximately 8% classification error across 20 subjects.

However, these studies have focused on binary classification for each mental state. Furthermore, most recent studies have not investigated continuous decoding for the direct application of accident prevention yet.

In this study, we designed experimental paradigms for inducing specific mental states under various simulated flight

conditions and also conducted both offline and pseudo-online analyses for classifying mental states. Additionally, we proposed a multiple feature block-based convolutional neural network (MFB-CNN) using temporal-spatio EEG filters to accurately classify fatigue, workload, distraction, and the normal state using only EEG signals. We evaluated the proposed method for the offline and pseudo-online analyses across all subjects. We could confirm the feasibility of classifying various mental states for implementation in real-world environments.

II. METHODS

A. SUBJECTS

Seven healthy subjects (S1-7, 6 males and 1 female, aged 27.8 (± 1.4)) with over 100 hr. of flight experience in Taean Flight Education Center, an affiliated education organization of Hanseo University, participated in our experiment. The experimental environment and protocols were reviewed and approved by the Institutional Review Board at Korea University [1040548-KU-IRB-18-92-A-2]. Before the experiment, all subjects were informed about the experimental protocols and consent according to the Declaration of Helsinki. They were asked to refrain from drinking alcohol and coffee and to sleep for 6~8 hr. the day before the experiment. At the end of the experiment, the subjects were instructed to complete the questionnaires regarding the recognition of their status and evaluation of the paradigms.

B. EXPERIMENTAL ENVIRONMENT

Fig. 1(a) shows the experimental environment for acquiring the pilots' EEG and EOG signals. The simulator system presented the flight environment using Cessna 172 aircraft. The environment included a cockpit, screen, wireless keypad, and signal amplifier. The cockpit consisted of a monitor display, flight yoke, and control panels. The screen offered a 210° view so that the pilots could see outside the aircraft. A wireless keypad was installed on the flight yoke for recording the pilots' response as the input. We used the signal amplifier (BrainAmp, Brain Products GmbH, Germany) to measure EEG and EOG signals [34]. We set up the sampling frequency to 1,000 Hz and a 60 Hz notch filter was used to remove DC noise.

30 EEG channels were placed on the subjects' scalp according to the international 10/20 system (Fp1, Fp2, F3, F4, Fz, FC1, FC2, FC5, FC6, T7, T8, C3, C4, Cz, CP1, CP2, CP5, CP6, TP9, TP10, P3, P4, P7, P8, Pz, PO9, PO10, O1, O2, and Oz) [29] as depicted in Fig. 1(b). In addition, 4 EOG channels (with electrodes vertically above (EOG1) and below (EOG2) the left eye and, horizontally at the outer canthus (left: EOG3, right: EOG4)) were also acquired (Fig. 1(b)). The reference electrode was placed at FCz channel, and the ground electrode was placed at AFz channel. Before the acquisition of EEG data, the skin impedance of EEG electrodes was adjusted below 10 k Ω by injecting conductive gel.

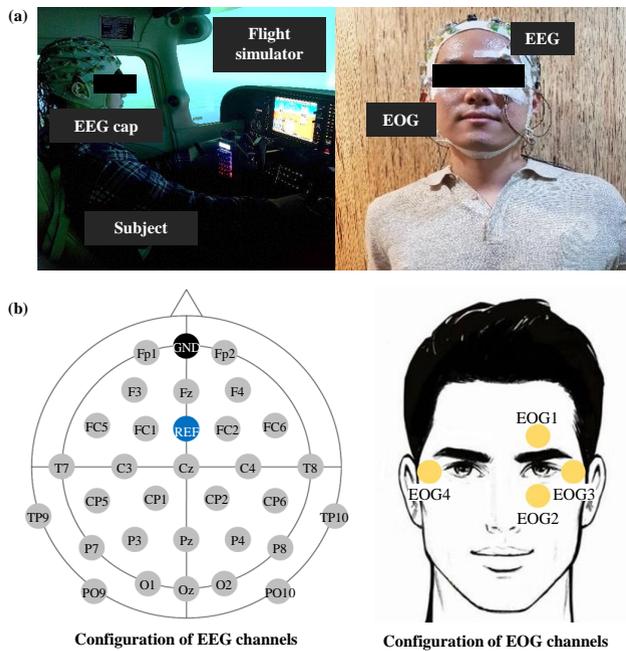


FIGURE 1: Experimental protocols for data acquisition using EEG and EOG signals. (a) Experimental environment in the flight simulator. (b) Data configuration of EEG and EOG channels.

C. EXPERIMENTAL PARADIGMS

We designed the experimental paradigms to effectively induce pilots' fatigue, workload, and distraction based on the conventional works [15], [22], [31], [35]. The pilot's mental state-related EEG signals were recorded during simulated flights based on our experimental paradigms.

1) Fatigue

The subjects performed a monotonous nighttime flight with the initial setting values of 3,000 feet (height), 0° (heading), and 100 knots (speed) for 1 hr. to induce drowsiness. The subjects entered the Karolinska sleepiness scale (KSS) value, a significant index of subjective drowsiness level [36], every minute using the keypad. KSS consists of nine levels (1: extremely alert, 2: very alert, 3: alert, 4: rather alert, 5: neither alert nor sleepy, 6: some signs of sleepiness, 7: sleepy, but no difficulty remaining alert, 8: sleepy, some effort to keep alert, 9: extremely sleepy). We considered level 1 to 6 as the normal state and level 8 to 9 as fatigue [37]. Level 7 was excluded to get a clearer distinction between the groups. See Fig. 2(a) for more details.

2) Workload

To induce a change of workload, flight instructions were given by an instructor who sat next to the pilots. The instructions consisted of three levels based on the complexity of the given task; 1: flight with the maintenance of given altitude and disregard for heading and velocity, 2: flight with the maintenance of continuously given altitude, heading, and velocity, 3: steep turn with given bank angle and the main-

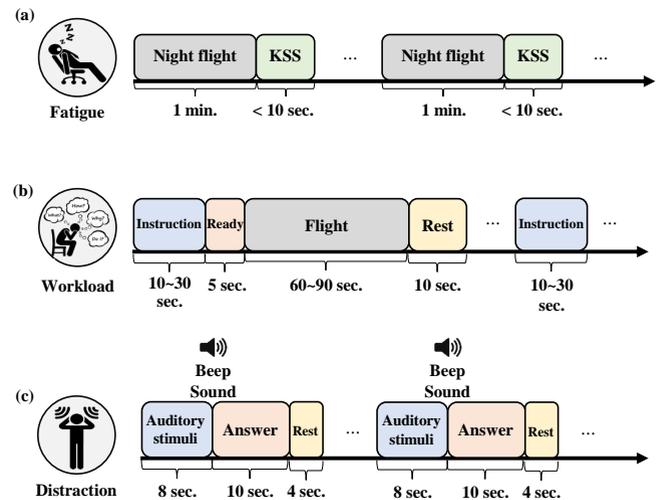


FIGURE 2: Experimental paradigms for inducing various pilot's mental states during simulated flight environment. (a) fatigue, (b) workload, and (c) distraction.

tenance of velocity toward given heading. The experiment consisted of 15 trials in level 1, 10 trials in level 2, and 10 trials in level 3. The label decision for 'Pass' and 'Fail' was marked by the instructor in the Rest interval (Fig. 2(b)).

3) Distraction

Fig. 2(c) illustrates the experimental paradigm for inducing distraction. While the subjects performed the simulated flight, a part of pre-recorded sentences (level 1: 4-9 words, level 2: 10-14 words, level 3: 15-22 words) of an air traffic control (ATC) message were presented at a regular time interval. Our designed experiment consisted of 40 trials in level 1, 40 trials in level 2, and 40 trials in level 3. The subjects were instructed to inwardly count the number of words in the ATC messages. After a beep sound, they entered the number of words using the keypad.

D. MFB-CNN

We proposed a CNN-based deep learning framework with temporal-spatio EEG filters to classify four mental states in the real-time environment as shown in Fig. 3. We adopted multiple feature blocks using various EEG features such as the spectral, temporal, and spatial information [38], [39]. For example, depending on the filter size, the CNN architecture could extract both temporal features and spatial information [39]. In addition, we designed a deep convolutional neural network owing to the characteristics of long EEG recording sessions compared to other BCI paradigm. Table 1 indicates the structure of the proposed MFB-CNN framework in detail.

Initially, EEG signals were down-sampled from 1,000 to 100 Hz and a bandpass filter was applied with 1-50 Hz using zero-phase, 2nd order, Butterworth filter. To obtain high-quality EEG data, we applied an independent component analysis (ICA) [40] to remove the contaminated components generated by unnecessary pilots' body movements, such as

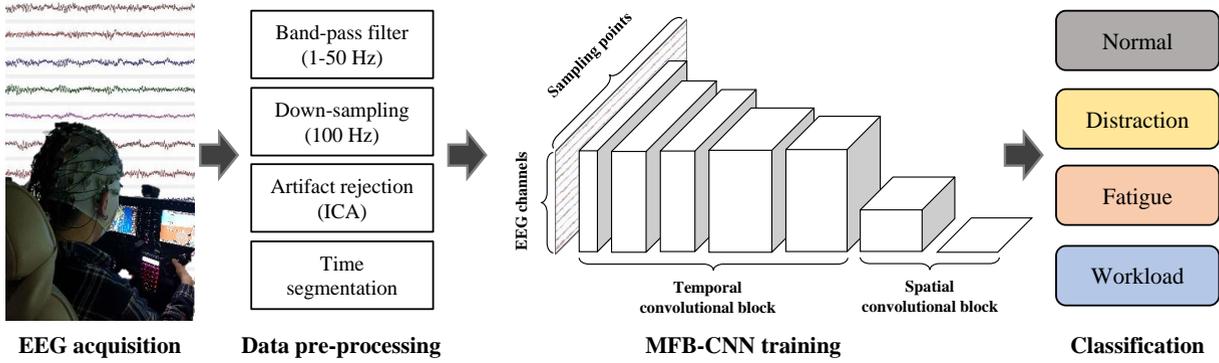


FIGURE 3: Overall flowchart of the proposed MFB-CNN for classifying various types of mental states.

TABLE 1: Structure of the proposed MFB-CNN framework

Block	Layer	Parameter	Output
-	Input	-	30×300
Convolutional block I	Convolution	Layer number: 3 Filter size: 1×5 Feature map: 16 Stride size: 1×1	$30 \times 16 \times 288$
	BatchNorm	-	-
Convolutional block II	Convolution	Layer number: 3 Filter size: 1×5 Feature map: 32 Stride size: 1×1	$30 \times 32 \times 276$
	BatchNorm	-	-
Convolutional block III	Convolution	Layer number: 3 Filter size: 1×5 Feature map: 64 Stride size: 1×1	$30 \times 64 \times 264$
	BatchNorm	-	-
Convolutional block IV	Convolution	Layer number: 4 Filter size: 1×5 Feature map: 64 Stride size: 1×1	$30 \times 64 \times 248$
	Max-pool	Filter size: 1×4 Stride size: 1×4	$30 \times 64 \times 62$
	BatchNorm	-	-
Convolutional block V	Convolution	Layer number: 3 Filter size: 1×5 Feature map: 64 Stride size: 1×1	$30 \times 64 \times 50$
	Avg-pool	Filter size: 1×5 Stride size: 1×5	$30 \times 64 \times 10$
	BatchNorm	-	-
	Dropout (0.5)	-	-
Convolutional block VI	Convolution	Layer number: 3 Filter size: 5×1 Feature map: 64 Stride size: 1×1	$18 \times 64 \times 10$
	Max-pool	Filter size: 2×1 Stride size: 2×1	$9 \times 64 \times 10$
	BatchNorm	-	-
Convolutional block VII	Convolution	Layer number: 3 Filter size: 3×1 Feature map: 64 Stride size: 1×1	$3 \times 64 \times 10$
	Avg-pool	Filter size: 3×1	-
	BatchNorm	-	-
	Activation (ELU)	-	$1 \times 64 \times 10$
	Dropout (0.5)	-	-
Classification	Fully connected	Hidden units: 128	128×1
	Fully connected	Hidden units: 64	64×1
	Fully connected	Hidden units: 4	4×1
	Softmax	-	-

eye blinks and movement of the head. ICA is simply one of the preprocessing steps to obtain clean EEG data. Since EEG signals are vulnerable to noise, most studies using deep learning or machine learning to detect user intentions or decode mental states have proceeded with feature extraction or classifier training after removing artifacts. Therefore, since our experimental environment was simulated flight where a lot of artifacts could occur; first, artifacts were removed from EEG signals, after which the training set and the test set was divided, finally the deep learning framework proceeded [41]–[45]. 4 EOG channels were used as the contaminated reference signals to reject the independent component (IC). All signal processing was conducted using a MATLAB 2019a software with a BCI toolbox [46].

After the pre-processing, EEG data were composed of the total number of EEG channels and sampling points (30×300) as input. Seven convolutional blocks were incorporated into a hierarchical CNN for the deep application of high-level features. Each convolutional block had three or four convolution layers and a batch normalization layer with a size of 32. The filter size for extracting temporal features was 1×5 and the stride size was 1×1 . In blocks VI and VII, the filter sizes were 5×1 and 3×1 , respectively, to consider the spatial information. Convolutional blocks V and VII were set with a 0.5 dropout ratio to solve the overfitting problem. Max-pooling and avg-pooling layers were used to reduce the number of output features while minimizing the loss of information. We applied the exponential linear units (ELUs) as an activation function in convolutional block VII. The output function is defined as:

$$ELU(x) = \begin{cases} x & \text{if } x > 0 \\ a(e^x - 1) & \text{if } x < 0 \end{cases} \quad (1)$$

For model training, we performed 30 iterations and the classification accuracies were evaluated on the test data.

E. CLASSIFICATION PERFORMANCE EVALUATION

1) Offline analysis

For the offline analysis, we evaluated the classification accuracies based on the proposed MFB-CNN architecture for

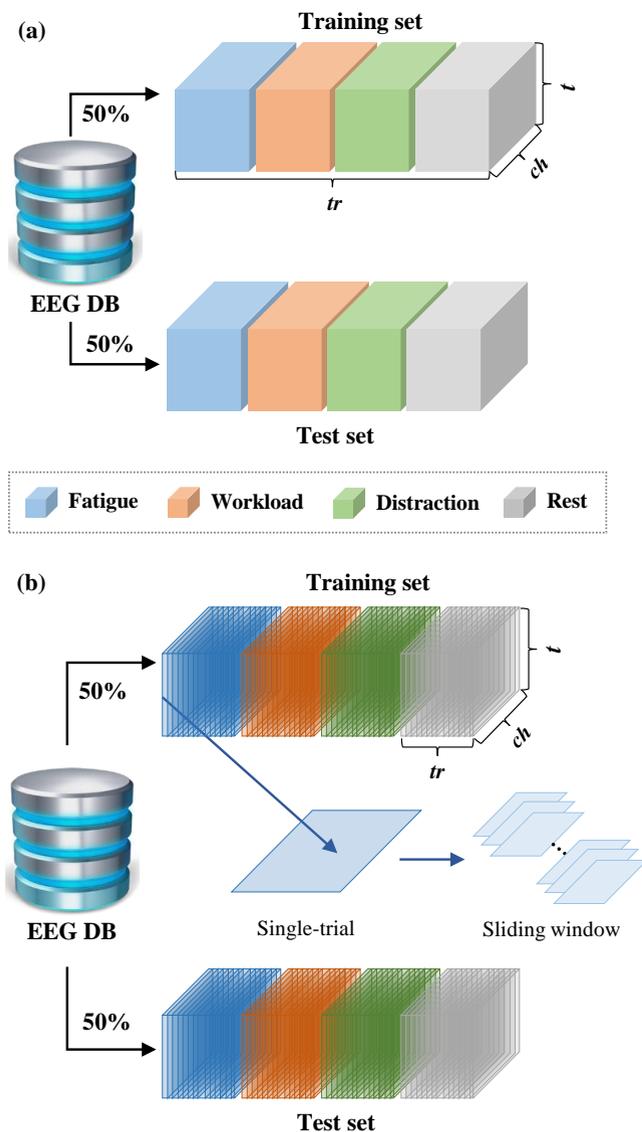


FIGURE 4: Schematic view of the training and test data for the (a) offline analysis and (b) pseudo-online analysis (tr : trial, ch : channel, t : time).

four different mental states across each subject. First, we randomly divided the entire dataset into a training set and a test set using 50:50 proportion. For a fair evaluation of classification accuracy, the cross-validation method was used. We evaluated the classification accuracies using the true label of samples and classification output each time.

In all paradigms, the data for each trial (1 min.) were segmented into 3 sec. data without overlap in the offline analysis as shown in Fig. 4(a). In fatigue experiment, subjects evaluated their condition themselves using KSS, subjectively. Since KSS could not be sometimes entered in a drowsy state occurred, each number of data samples have differed according to subjects. For example, S1 obtained 687 samples for fatigue and 427 samples for the normal state. In contrast, S2 recorded 482 samples for fatigue and 662 samples for the normal state. Each subject had 900 samples and 360 samples

TABLE 2: Number of samples for training and test sets in the offline analysis and pseudo-online analysis

Subject	Offline analysis		Pseudo-online analysis	
	Training	Test	Training	Test
S1	1,187	1,187	9,567	9,567
S2	1,202	1,202	9,676	9,676
S3	1,203	1,203	9,689	9,689
S4	1,212	1,212	9,599	9,599
S5	1,211	1,211	9,669	9,669
S6	1,210	1,210	9,732	9,732
S7	1,205	1,205	9,555	9,555

in workload and distraction experiments, respectively. Across all subjects, we could obtain 4,896 samples for fatigue, 6,300 samples for workload, 2,520 samples for distraction, and 3,144 samples for the normal state.

In the case of fatigue, we used 2,448 samples when conducting training and test, respectively. In terms of workload state, 3,150 samples were used when training and testing our proposed framework, respectively. In the case of distraction, 1,260 samples were used and, in the normal state, 1,572 samples were used during training and testing our model, respectively. We divided data samples evenly into 50:50 proportion for each mental state (Fig. 4(a)).

2) Pseudo-online analysis

We also evaluated the proposed MFB-CNN in the pseudo-online analysis to confirm the feasibility of continuous decoding for human mental states. In this study, due to the low commercialization of BCI technique and the environmental constraints of pilots' driving, the evaluation has been conducted through the pseudo-online analysis, which is similar to a real-time experiment. Generally, in this specific environment, most BCI investigators have adopted pseudo-online analysis instead of fully online tests for proving the possibility of real-time scenarios [47]–[49]. Therefore, we also confirmed the feasibility of continuous decoding using MFB-CNN through offline analysis first, then we re-trained MFB-CNN using all dataset assuming general calibration data for a pseudo-online analysis [49]. We conducted the pseudo-online analysis with a classifier that classified four mental states in the offline analysis.

A sliding window of 3 sec. duration with an overlap of 0.5 sec. was applied for continuous classification output as shown in Fig. 4(b). Each window was also processed with the same preprocessing step as shown in Fig. 3, and the parameters of the model for classification were used throughout the training phase. The classification results of the pseudo-online analysis were evaluated for each sub-trial between the true label of samples and predicted classification output.

We used 13,800 samples when conducting training and test, respectively, in the case of fatigue. Also, in terms of workload, 20,825 samples were used when training and testing our model, respectively. In the case of distraction, we

used 23,954 samples when conducting training and test and, in terms of the normal state, 8,908 samples were used when training and testing our model. We fairly selected the 8,908 samples from each mental state same as the offline analysis so that we used a total of 35,632 samples when conducting training and test.

III. RESULTS

A. CLASSIFICATION PERFORMANCES IN THE OFFLINE ANALYSIS

Table 3 shows the overall classification performances for each subject in the offline analysis. We applied a 2-fold cross-validation method to evaluate classification accuracy fairly; 50% of the samples were randomly selected for training and the remaining 50% were reserved for validation. Also, we repeated the two-fold cross-validation four times after adopting a different shuffle order in each time [50]. We obtained the grand-average accuracy of 0.75 (± 0.04) for decoding the four mental states. S6 showed the highest classification accuracy of 0.79, and S4 showed the lowest classification accuracy of 0.71.

Table 4 represents the comparison of classification accuracies between the conventional methods and the proposed network in the offline analysis. The conventional methods used for performance comparison were power spectral density-SVM (PSD-SVM) [51], PSD-kNN [20], channel-wise CNN (CCNN) [52], and deep long short-term memory (LSTM-D) [53]. PSD-SVM extracted the PSD of the theta, alpha, and beta bands while neglecting the spatial information of EEG signals as features, and SVM was used for the classification [51]. Similarly, PSD-kNN extracted the PSD of the theta, alpha, and beta bands. After extracting the PSD as a feature, it trained using the kNN with the number of neighbors set at 5 [20]. CCNN used a channel-wise filter to reflect the spatial information of EEG signals for classification [52]. LSTM-D method was comprised of a sequence-to-sequence and many-to-one LSTM layers [53].

The comparison of classification accuracies showed that our proposed MFB-CNN with temporal-spatio EEG filters had the highest accuracy for classifying various mental states. To verify the classification performance difference between the conventional methods and the proposed model, we applied the analysis of variance (ANOVA) with the Bonferroni correction for multiple comparisons. Initially, we validated the normality and homoscedasticity owing to a small number of samples. The normality for each conventional method applying the Shapiro–Wilk test was satisfied with a null hypothesis (H_0), and the assumption of homoscedasticity based on Levene's test was also met for each group. Hence, we conducted a statistical analysis between the conventional methods which satisfied these conditions and MFB-CNN using multiple comparisons with a Bonferroni correction. MFB-CNN had the most statistically significant difference in performance among the machine learning method such as PSD-SVM and PSD-kNN with p -value below 0.001.

TABLE3: Classification accuracies of four mental states according to each subject

Subject	Offline analysis				Average
	1 st acc.	2 nd acc.	3 rd acc.	4 th acc.	
S1	0.75	0.78	0.79	0.79	0.78
S2	0.76	0.78	0.75	0.75	0.76
S3	0.69	0.71	0.74	0.74	0.72
S4	0.71	0.70	0.72	0.72	0.71
S5	0.73	0.74	0.73	0.73	0.73
S6	0.76	0.82	0.80	0.80	0.79
S7	0.76	0.78	0.77	0.77	0.77
Grand-average	0.74	0.76	0.76	0.76	0.75
Std.	0.03	0.04	0.03	0.03	0.04

*acc.: Accuracy

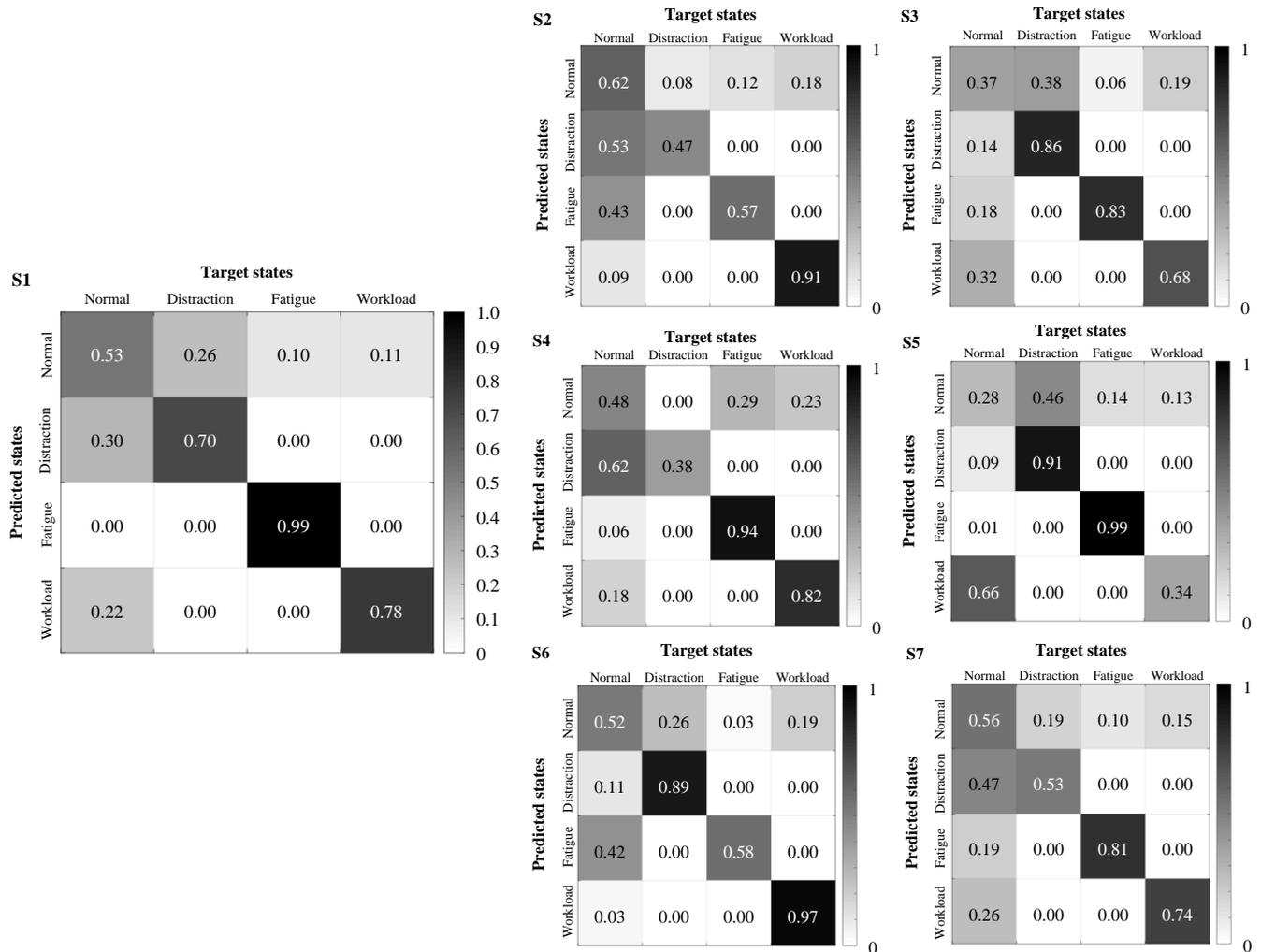
TABLE4: Comparison of classification accuracies between the conventional methods and the proposed MFB-CNN

Subject	Offline analysis				Average
	1 st acc.	2 nd acc.	3 rd acc.	4 th acc.	
PSD-SVM [51]	0.41	0.42	0.40	0.41	0.41
PSD-kNN [20]	0.48	0.43	0.46	0.45	0.46
CCNN [52]	0.59	0.60	0.64	0.61	0.61
LSTM-D [53]	0.61	0.60	0.65	0.62	0.62
MFB-CNN	0.74	0.76	0.76	0.76	0.75

*acc.: Accuracy

In addition, we also evaluated the degree of confusion across each subject. Fig. 5 shows the confusion matrices when classifying the four mental states using MFB-CNN for all subjects in the offline analysis. Each column of the matrix contains the target states and each row represents the predicted states. The representative subject (S1) had the highest true positive rate (TPR) for detecting fatigue, and the value was 0.99. The subjects S2 and S6 had the highest TPR for detecting workload, and the values were 0.91 and 0.97, respectively. Subject S3 had the highest TPR for decoding distraction of 0.86. S4, S5, and S7 had the highest TPR for recognizing fatigue, and the values were 0.94, 0.99, and 0.81, respectively.

Through the experimental results, we confirmed that the results for all subjects had some common characteristics. When the target state was the distraction, the rates of misclassification of the predicted state as fatigue or workload were 0. Similarly, when the target label was fatigue, the rates of misclassification of the predicted label as distraction or workload were 0. Additionally, when the target state was workload, the rates of misclassification of the predicted state as fatigue or distraction were 0. Finally, most of the subjects could not detect the normal state correctly. The TPR for detecting the normal state was the lowest accuracy, in other words, most subjects confused their state as the normal state.



FIGURES: Confusion matrices of classification accuracy for each class in the offline analysis.

B. CLASSIFICATION PERFORMANCES IN THE PSEUDO-ONLINE ANALYSIS

We obtained the evaluation results of decoding the mental states applying the One-Versus-Rest (OVR) strategy. We conducted an evaluation between one of the mental states and rest states. Unlike the typical BCI research, in the research of mental state detection, the recording session of each mental state is different because it takes a long recording time to obtain data related to the state change. If we artificially concatenate data of recording sessions for pseudo-online analysis, continuous decoding of mental states could not be conducted at the concatenated data points owing to not formed continuous EEG signals. Therefore, we applied the OVR strategy as state detection and used the trained classifier capable of four state classification instead. The grand-average classification accuracy was 0.72 (± 0.20) for ‘fatigue’ vs. ‘rest states’, 0.72 (± 0.27) for ‘workload’ vs. ‘rest states’, and 0.61 (± 0.18) for ‘distraction’ vs. ‘rest states’ as depicted in Table 5. In addition, S1 showed significantly high accuracy in terms of detecting workload and distraction. Similarly, S2

TABLE5: Detection accuracies following to each mental state for all subjects in the pseudo-online analysis

Subject	Pseudo-online analysis		
	Fatigue	Workload	Distraction
S1	0.80	0.94	0.93
S2	0.94	0.68	0.75
S3	0.45	0.82	0.51
S4	0.92	0.55	0.46
S5	0.47	0.39	0.41
S6	0.65	0.91	0.73
S7	0.78	0.62	0.49
Grand-average	0.72	0.72	0.61
Std.	0.20	0.27	0.18

and S4 represented high performance for detecting fatigue. Also, S6 showed high accuracy for detecting workload.

Table 6 shows the comparison of classification accuracies between the conventional methods and MFB-CNN. We used the same conventional methods used in the offline analy-

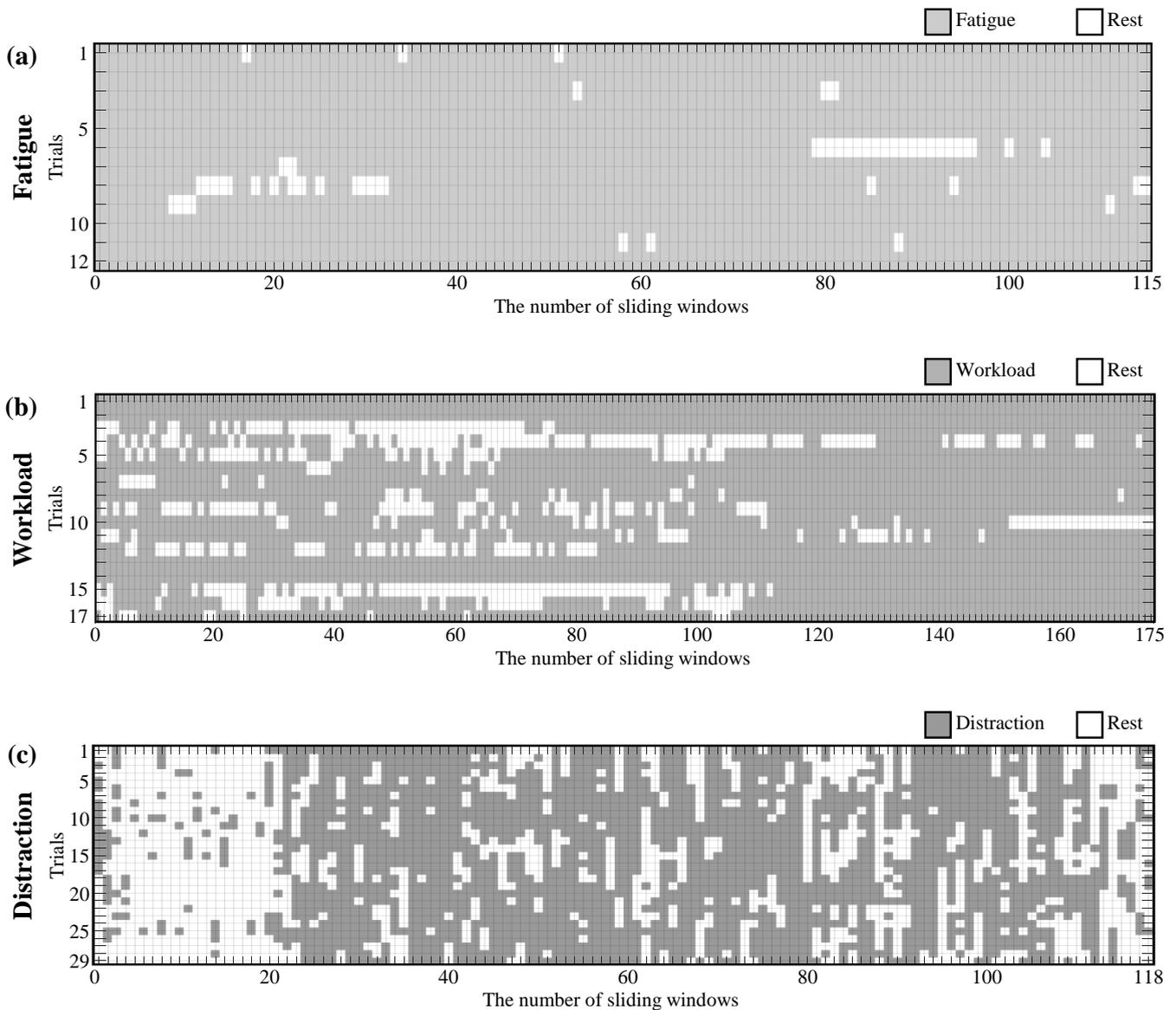


FIGURE 6: Representation of the detection outputs for (a) fatigue, (b) workload, and (c) distraction in the pseudo-online analysis.

TABLE 6: Comparison of detection accuracies between the conventional methods and MFB-CNN in the pseudo-online analysis

Subject	Pseudo-online analysis		
	Fatigue	Workload	Distraction
PSD-SVM [51]	0.48	0.45	0.39
PSD-kNN [20]	0.51	0.49	0.42
CCNN [52]	0.61	0.60	0.52
LSTM-D [53]	0.60	0.62	0.52
MFB-CNN	0.72	0.72	0.61

sis. The comparison of classification accuracies showed that MFB-CNN had the highest accuracy for detecting each mental state. As in the offline analysis, we conducted a statistical analysis employing ANOVA with the Bonferroni correction. ‘MFB-CNN’ vs. ‘PSD-kNN’ had the most statistically sig-

nificant difference in performance with p -value below 0.002.

CCNN is a specialized method that considers spatial information. The accuracy of our proposed model is higher than that of CCNN because MFB-CNN considers not only spatial information but also temporal information. LSTM-D is a specialized method for considering sequential information. However, MFB-CNN has higher accuracy than LSTM-D because it considers both spatial and spectral information as well as sequential information.

Fig. 6 represents the detection output for each state during the pseudo-online experiment. We used a 3 sec. sliding window with an overlap of 0.05 sec.. Each gray square indicates the detection results for the class of the corresponding mental state. The white squares represent the detection output for the rest states. S2, S7, and S6 showed relatively clear results for fatigue, workload, and distraction, respectively, compared

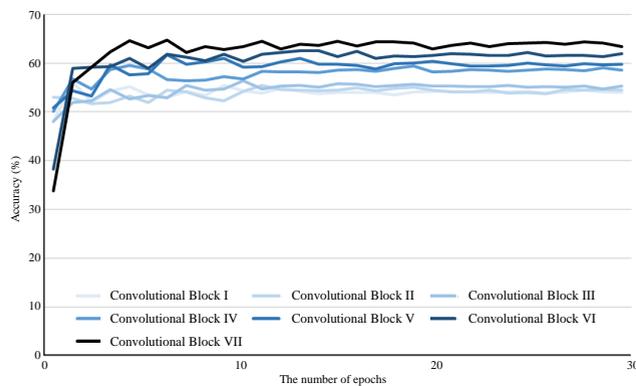


FIGURE 7: Convergence curve of the training accuracy on the representative subject.

with the other subjects.

In case of workload, we found out that the majority of sliding windows were classified as rest states in a few number of trials. This is considered to be a sufficient result when the task is entered as a failure since the instructor sitting next to the subject determines whether the result of each task is 'Pass' or 'Fail'.

In terms of distraction, we confirmed that the early parts of many trials were classified as rest states. Before we started the distraction experiment, we fully explained to subjects how to experiment and transmitted the precautions (minimize the movements of body and eye, inwardly count the number of words in the ATC messages without using fingers, and etc.). However, in the early part of the experiment, subjects had difficulty performing the experiment and counted the number of words in the ATC messages not inwardly.

C. TRAINING CONVERGENCE CURVE FOR MFB-CNN

To observe the convergence process of MFB-CNN, the accuracy change curve of the representative subject (S2), which measures the percentage of the correctly classified samples in the training set along with epochs is shown in Fig. 7. For all subjects, the training process converges within about 10 epochs. Our proposed model showed that it could train using a small number of epochs. Additionally, we obtained shorter computational times for model training (40.76 sec.) and testing (0.0029 sec.), respectively, compared with other conventional methods.

D. NEUROPHYSIOLOGICAL ANALYSIS

Moreover, we have tried to analyze the neurophysiological phenomenon using temporal, spectral, and spatial information of EEG features according to each mental state (fatigue, workload, and distraction). The proposed MFB-CNN was designed to consider the neural information using the temporal-spatio filters.

Fig. 8 shows the grand-average spectrograms of mental states for each channel according to the spectral bands (the delta, theta, alpha, and beta bands) of EEG signals. The spectrograms represent the grand-average power to visualize

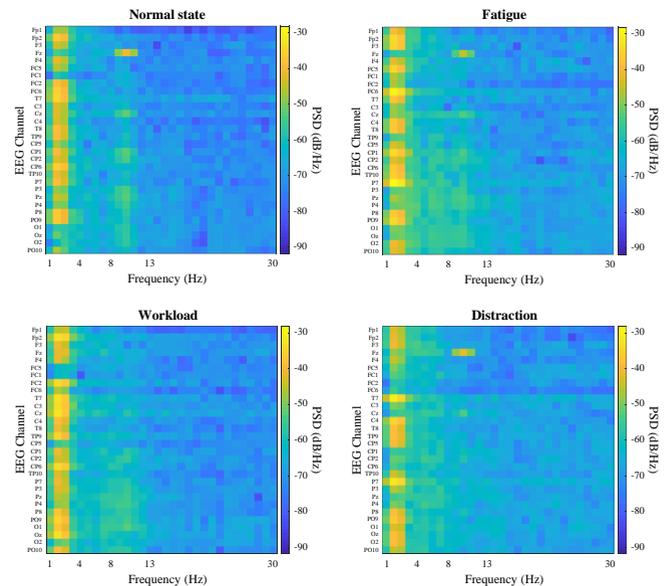


FIGURE 8: Grand-average spectrograms of each mental state across all subjects.

brain activation. Each column of the matrix reflects the frequency and each row represents the channels. The PSD was computed for all EEG channels and each frequency band across all subjects. The spectrograms show that the PSD of each spectral band and each channel were hardly different among mental states.

Fig. 9 represents the grand-average PSD of each spectral band of EEG signals corresponding to four mental states. Each mental state has the highest PSD of a particular band in a certain brain region. The highest PSD of a particular band means the highest activation degree for a specific mental state in that band. In the temporal region, the activation degree of fatigue was higher in the delta, theta, and alpha bands compared with that of other mental states. In the central region, the notable highest activation degree was observed for workload, represented in the alpha and beta bands. In the parietal region, the activation degree of distraction was the highest in the theta band compared with other mental states. In addition, we could confirm the statistical significance through the ANOVA with the Bonferroni correction. In specific, at the temporal region, 'fatigue' vs. 'the other mental states (workload, distraction, and the normal state)' has a statistically significant difference in the delta band ($p < 0.001$). Hence, the delta band at the temporal region has meaningful spatial and spectral information when detecting 'fatigue' among various mental states. Also, at the temporal region, 'fatigue' vs. 'normal' has the statistical significance in the theta band ($p < 0.001$). Additionally, at the central region, 'the normal state' vs. 'the abnormal states (fatigue, workload, and distraction)' has a statistically significant difference in the theta and alpha bands ($p < 0.001$).

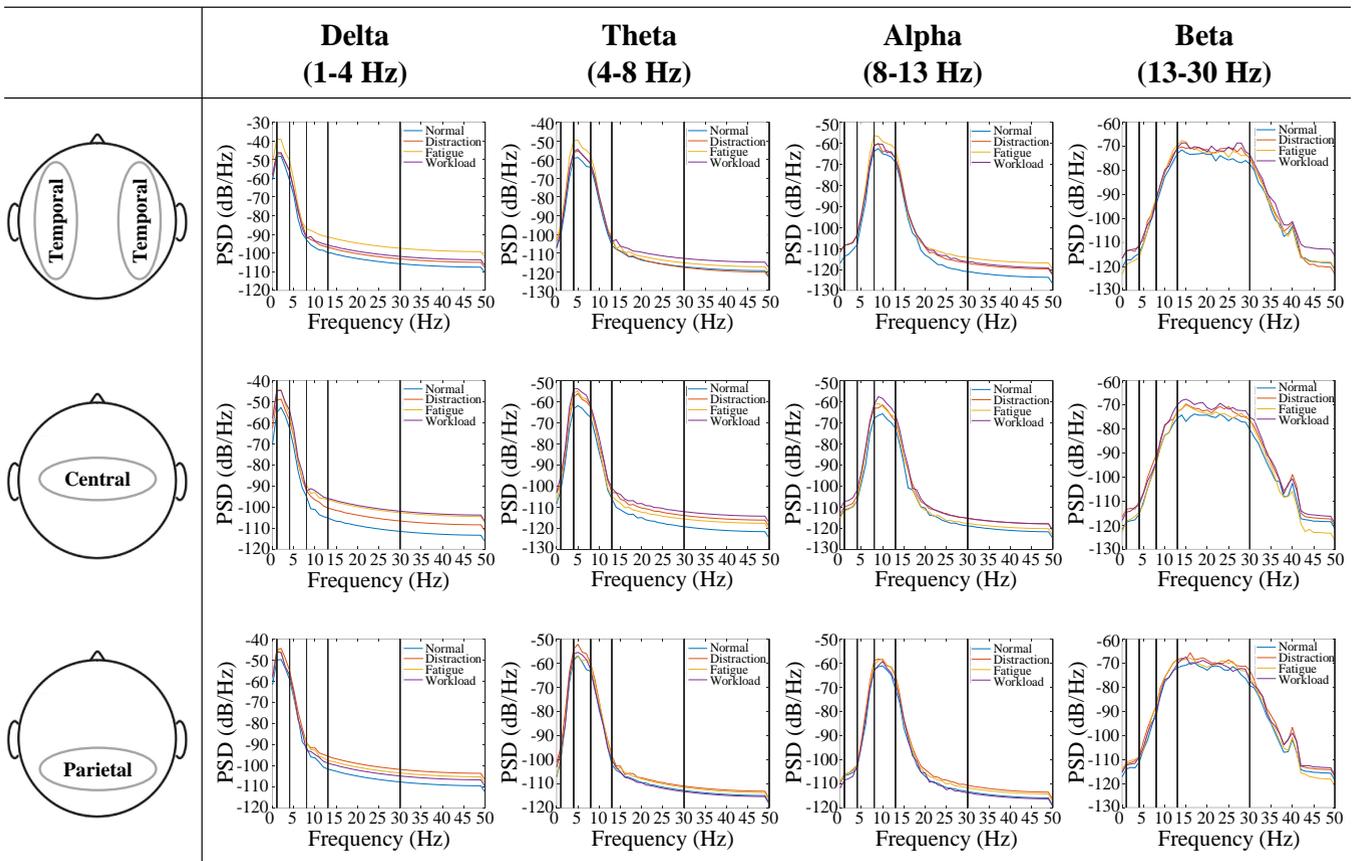


FIGURE 9: Grand-average spectral peak of each power band using the PSD for mental states across all subjects.

IV. DISCUSSION

In this paper, we demonstrated the feasibility of continuous decoding of various mental states (fatigue, workload, distraction, and the normal state) based on a deep learning method in a pilot environment. The proposed MFB-CNN was based on deep CNN architecture using temporal-spatio EEG filters related to the pilots' mental states. To the best of our knowledge, this study is one of the novel attempts to conduct continuous decoding of various mental states with robust classification performance.

Recent conventional works have focused on the classification of user mental states with high performances using advanced machine learning algorithms and deep learning methods such as [54]–[60]. Zhang *et al.* [54] introduced two deep learning-based frameworks with novel spatio-temporal preserving representations of raw EEG streams to identify human intentions with high performances. The two frameworks consisted of both convolutional and recurrent neural networks effectively exploring the preserved spatial and temporal information in either a cascade or a parallel manner. Also, they reflected the matrix transformation in 2-D data based on the channel location of the brain regions. Bashivan *et al.* [55] proposed a novel approach for learning such representations from multi-channel EEG time-series, and demonstrated its advantages in the context of mental

workload classification task. Their proposed approach was designed to preserve the spatial, spectral, and temporal structure of EEG. In addition, they used topographical maps based on the channel location of the brain regions. In this work, we decoded the pilots' mental states through the method of applying the temporal filter and spatial filter in the preprocessing step for deep learning training.

Accurate detection for mental states of drivers or pilots has been attracted as a critical issue in AI fields owing to the technical advances of autonomous vehicles and the autopilot system. Therefore, some researchers have focused on detecting mental states [15], [17], [22], [25], [61], while a few others have performed the analysis of physiological signals for detecting mental states of users in the aircraft environment [14], [18].

In specific, while aviation accidents do not happen frequently as vehicle accidents, when they occur, they cause far worse casualties and larger explosions. These aviation accidents not only produce immediate effects (e.g., casualties and aircraft explosion) but could induce secondary impacts, such as harming the surrounding environment [62]. Statistically, more than 70% of aviation accidents can be attributed to human factors, such as pilot fatigue; drowsiness may also be an important contributor to a large number of aviation accidents [63]. Therefore, in this study, we focused on analyzing EEG

data regarding fatigue, workload, and distraction in simulated flight experience with human pilots. In the field of detecting or classifying human mental states, the pilot environment is a very rare environment. Only volunteers with over 100 hr. of flight experience were allowed to participate in this study. We designed experimental paradigms to induce various mental states. In order to obtain high-quality EEG signals for various mental states, we conducted experiments very strictly, and collecting experimental data was quite challenging owing to the difficulty in constructing the experimental environment. Our experimental data were able to indicate the pilots' mental condition, and we showed that MFB-CNN could contribute to the prevention of large aviation and vehicle accidents in the real world, as it can robustly detect users' mental conditions.

Although direct performance comparison with related studies was impossible due to different experimental protocols and paradigms, we conducted the comparison of classification performances between the conventional methods and MFB-CNN using our experimental data. MFB-CNN showed the highest classification accuracies for both offline and pseudo-online analyses. In the offline analysis, we showed the grand-average classification accuracy of $0.75 (\pm 0.04)$. In the pseudo-online analysis, we obtained the detection accuracy of $0.72 (\pm 0.20)$, $0.72 (\pm 0.27)$, and $0.61 (\pm 0.18)$ for each fatigue, workload, and distraction, respectively. The reason that the detection of distraction is relatively low is that the subjects tended to count the number of words in the ATC not inwardly although we provided the instruction to inwardly count the number of words in the ATC messages. In aspects of classification accuracy, we could confirm that the proposed MFB-CNN showed a much higher performance of approximately 0.13 than the conventional methods in the offline analysis. Also, in the pseudo-online analysis, we could confirm that MFB-CNN presented a higher detection accuracy of 0.11, 0.10, and 0.09 for each mental state, respectively. Also, it reached high training accuracy using only a small number of epochs as depicted in Fig. 7.

In addition, the standard deviation of the subject's performance in the offline analysis presented the lowest value when using MFB-CNN compared with other existing methods. We could confirm that the standard deviation of the grand-average performance using machine learning methods such as PSD-SVM and PSD-kNN tends to have a high variation of performance among the subjects (0.09). On the other hand, the classification performance using deep learning methods such as CCNN and LSTM-D presented low performance variation tendencies across all subjects (0.06). Among deep learning methods, the standard deviation of MFB-CNN was the lowest, and the value was 0.04. However, the assumption of homoscedasticity based on Levene's test could not be met since the differences between the standard deviation of machine learning and deep learning methods showed high values. Hence, we could not adopt statistical multiple comparisons between machine learning and deep learning methods. Also, we represented the grand-average spectral peak of each band power using the PSD for mental states

as shown in Fig. 9. At the temporal region, the delta band can be used to classify the 'fatigue' vs. 'the other mental states'. Moreover, at the temporal region, the theta band can be used to classify the 'fatigue' vs. 'the normal state'. Finally, at the central region, the theta and alpha bands can be used to classify the 'the normal state' vs. 'the abnormal states'.

MFB-CNN architecture is composed of several convolution blocks compared to the existing deep learning architectures in the BCI field; therefore, in the model training, the architecture has high computation cost. Owing to the limitation, the current MFB-CNN architecture could be difficult to applying real-world BCI scenarios. Since the long calibration time makes the subjects could induce fatigue states and spread low attention state. Therefore, to classify mental states more accurately and rapidly in real-world BCI scenarios, we plan to modify MFB-CNN architecture with a shallower design but maintaining classification performances [15], [39], [64]; and adapt a weight-parameter sharing method [65] and data augmentation method [66] to the model. As depicted in Fig. 7, we confirmed the possibility of model swallowing since a stable performance was shown despite the small number of convolutional blocks. Additionally, each state activates different brain regions. At present, our proposed model used EEG data by extracting features from the whole brain area (Fig. 9). In order to improve the robustness of its performance, we plan to adopt the principle of a hierarchical deep learning model that can fully utilize the spatial characteristics of each state [65].

V. CONCLUSIONS AND FUTURE WORKS

In this study, we investigated the feasibility of classifying the pilots' various mental states (fatigue, workload, distraction, and the normal state) using only EEG signals. Also, we proposed MFB-CNN for classifying mental states with high accuracy. We obtained EEG data corresponding to the four mental states from seven subjects using our designed paradigms in the simulated flight environment. Our proposed model successfully enhanced classification accuracy of 0.75 for the offline analysis. Also, MFB-CNN improved the detection accuracy of 0.72, 0.72, and 0.61 for each fatigue, workload, and distraction state, respectively, during continuous decoding.

In future works, acquiring randomized sequential modes of each mental state should be considered. Additionally, an inter-session variation of physiological signals should be explored in order to develop a robust real-time BCI system. Hence, we will implement the proposed model in real-world environments.

ACKNOWLEDGMENT

We would like to thank S.-Y. Han for designing experimental paradigms and H.-J. Kim for useful discussion.

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, 2003.
- [2] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clin. Neurophysiol.*, vol. 113, no. 6, pp. 767–791, 2002.
- [3] 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, p. e0172578, 2017.
- [4] K. K. Ang and C. Guan, "EEG-based strategies to detect motor imagery for control and rehabilitation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 4, pp. 392–401, 2017.
- [5] 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, p. 016014, 2015.
- [6] M. Alhussein, G. Muhammad, and M. S. Hossain, "EEG pathology detection based on deep learning," *IEEE Access*, vol. 7, pp. 27781–27788, 2019.
- [7] 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, 2013.
- [8] 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, p. 38565, 2016.
- [9] 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.*, 2020.
- [10] B. Edelman, J. Meng, D. Suma, C. Zurn, E. Nagarajan, B. Baxter, C. Cline, and B. He, "Noninvasive neuroimaging enhances continuous neural tracking for robotic device control," *Sci. Robot.*, vol. 4, no. 31, p. eaaw6844, 2019.
- [11] 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, 2018.
- [12] P. Stawicki, F. Gembler, A. Rezeika, and I. Volosyak, "A novel hybrid mental spelling application based on eye tracking and SSVEP-based BCI," *Brain Sci.*, vol. 7, no. 4, p. 35, 2017.
- [13] 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, p. e111157, 2014.
- [14] G. Borghini, L. Astolfi, G. Vecchiato, D. Mattia, and F. Babiloni, "Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness," *Neurosci. Biobehav. Rev.*, vol. 44, pp. 58–75, 2014.
- [15] Z. Gao, X. Wang, Y. Yang, C. Mu, Q. Cai, W. Dang, and S. Zuo, "EEG-based spatio-temporal convolutional neural network for driver fatigue evaluation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 9, pp. 2755–2763, 2019.
- [16] S. K. Lal, A. Craig, P. Boord, L. Kirkup, and H. Nguyen, "Development of an algorithm for an EEG-based driver fatigue countermeasure," *J. Safety Res.*, vol. 34, no. 3, pp. 321–328, 2003.
- [17] S. Ahn, T. Nguyen, H. Jang, J. G. Kim, and S. C. Jun, "Exploring neuro-physiological correlates of drivers' mental fatigue caused by sleep deprivation using simultaneous EEG, ECG, and fNIRS data," *Front. Hum. Neurosci.*, vol. 10, p. 219, 2016.
- [18] E. Q. Wu, X. Y. Peng, C. Z. Zhang, J. X. Lin, and R. S. F. Sheng, "Pilot's fatigue status recognition using deep contractive autoencoder network," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 10, pp. 3907–3919, 2019.
- [19] Y. Liu, H. Ayaz, and P. A. Shewokis, "Multisubject "learning" for mental workload classification using concurrent EEG, fNIRS, and physiological measures," *Front. Hum. Neurosci.*, vol. 11, p. 389, 2017.
- [20] J. A. Blanco, M. K. Johnson, K. J. Jaquess, H. Oh, L.-C. Lo, R. J. Gentili, and B. D. Hatfield, "Quantifying cognitive workload in simulated flight using passive, dry EEG measurements," *IEEE Trans. Cogn. Develop. Syst.*, vol. 10, no. 2, pp. 373–383, 2018.
- [21] R. N. Roy, S. Charbonnier, A. Campagne, and S. Bonnet, "Efficient mental workload estimation using task-independent EEG features," *J. Neural Eng.*, vol. 13, no. 2, p. 026019, 2016.
- [22] A. Sonleitner, M. S. Sonleitner, M. Simon, S. Willmann, A. Ewald, A. Buchner, and M. Schrauf, "EEG alpha spindles and prolonged brake reaction times during auditory distraction in an on-road driving study," *Accid. Anal. Prev.*, vol. 62, pp. 110–118, 2014.
- [23] W. J. Horrey, M. F. Lesch, A. Garabet, L. Simmons, and R. Maikala, "Distraction and task engagement: how interesting and boring information impact driving performance and subjective and physiological responses," *Appl. Ergon.*, vol. 58, pp. 342–348, 2017.
- [24] Y.-K. Wang, T.-P. Jung, and C.-T. Lin, "Theta and alpha oscillations in attentional interaction during distracted driving," *Front. Behav. Neurosci.*, vol. 12, p. 3, 2018.
- [25] C. J. de Naurois, C. Bourdin, A. Stratulat, E. Diaz, and J.-L. Vercher, "Detection and prediction of driver drowsiness using artificial neural network models," *Accid. Anal. Prev.*, vol. 126, pp. 95–104, 2019.
- [26] I.-H. Kim, J.-W. Kim, S. Haufe, and S.-W. Lee, "Detection of braking intention in diverse situations during simulated driving based on EEG feature combination," *J. Neural. Eng.*, vol. 12, no. 1, p. 016001, 2014.
- [27] M. Lee, R. D. Sanders, S.-K. Yeom, D.-O. Won, K.-S. Seo, H.-J. Kim, G. Tononi, and S.-W. Lee, "Network properties in transitions of consciousness during propofol-induced sedation," *Sci. Rep.*, vol. 7, no. 1, p. 16791, 2017.
- [28] M. A. Boksem and M. Tops, "Mental fatigue: costs and benefits," *Brain Res. Rev.*, vol. 59, no. 1, pp. 125–139, 2008.
- [29] J.-H. Jeong, B.-W. Yu, D.-H. Lee, and S.-W. Lee, "Classification of drowsiness levels based on a deep spatio-temporal convolutional bidirectional LSTM network using electroencephalography signals," *Brain Sci.*, vol. 9, no. 12, p. 348, 2019.
- [30] B. H. Kantowitz and P. A. Casper, "Human workload in aviation," in *Human Error in Aviation*. Routledge, 2017, pp. 123–153.
- [31] M. P. Çakır, M. Vural, S. Ö. Koç, and A. Toktaş, "Real-time monitoring of cognitive workload of airline pilots in a flight simulator with fNIR optical brain imaging technology," in *Int. Conf. Aug. Cognit.*, 2016, pp. 147–158.
- [32] X. Chen, H. Zhang, L. Zhang, C. Shen, and S.-W. Lee, "Extraction of dynamic functional connectivity from brain grey matter and white matter for MCI classification," *Hum. Brain Mapp.*, vol. 38, no. 10, pp. 5019–5034, 2017.
- [33] S. Hong, H. Kwon, S. H. Choi, and K. S. Park, "Intelligent system for drowsiness recognition based on ear canal electroencephalography with photoplethysmography and electrocardiography," *Inf. Sci.*, vol. 453, pp. 302–322, 2018.
- [34] 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, 2018.
- [35] M. Choi, G. Koo, M. Seo, and S. W. Kim, "Wearable device-based system to monitor a driver's stress, fatigue, and drowsiness," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 3, pp. 634–645, 2018.
- [36] I. Daza, L. Bergasa, S. Bronte, J. Yebes, J. Almazán, and R. Arroyo, "Fusion of optimized indicators from advanced driver assistance systems (ADAS) for driver drowsiness detection," *Sensors*, vol. 14, no. 1, pp. 1106–1131, 2014.
- [37] H. Mårtensson, O. Keelan, and C. Ahlström, "Driver sleepiness classification based on physiological data and driving performance from real road driving," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 2, pp. 421–430, 2018.
- [38] 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, 2018.
- [39] 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, p. 056013, 2018.
- [40] T.-P. Jung, C. Humphries, T.-W. Lee, S. Makeig, M. J. McKeown, V. Iragui, and T. J. Sejnowski, "Extended ICA removes artifacts from electroencephalographic recordings," in *AProc. Adv. Neural Inf. Process. Syst. (NIPS)*, 1998, pp. 894–900.
- [41] N. Jiang, L. Gizzi, N. Mrachacz-Kersting, K. Dremstrup, and D. Farina, "A brain-computer interface for single-trial detection of gait initiation from movement related cortical potentials," *Clin. Neurophysiol.*, vol. 126, no. 1, pp. 154–159, 2015.
- [42] C. Y. Sai, N. Mokhtar, H. Arof, P. Cumming, and M. Iwahashi, "Automated classification and removal of EEG artifacts with SVM and wavelet-ICA," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 3, pp. 664–670, 2017.
- [43] E. Dong, K. Zhou, J. Tong, and S. Du, "A novel hybrid kernel function relevance vector machine for multi-task motor imagery EEG classification," *Biomed. Signal Process. Control*, vol. 60, p. 101991, 2020.
- [44] L.-W. Ko, O. Komarov, and S.-C. Lin, "Enhancing the hybrid BCI performance with the common frequency pattern in dual-channel EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 7, pp. 1360–1369, 2019.

- [45] 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, 2018.
- [46] B. Blankertz, M. Tangermann, C. Vidaurre, S. Fazli, C. Sannelli, S. Haufe, C. Maeder, L. E. Ramsey, I. Sturm, G. Curio, and K.-R. Müller, "The Berlin brain-computer interface: non-medical uses of BCI technology," *Front. Neurosci.*, vol. 4, p. 198, Dec. 2010.
- [47] 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, 2020.
- [48] B. Blankertz, G. Curio, and K.-R. Müller, "Classifying single trial EEG: Towards brain computer interfacing," in *Adv. Neural Inf. Process. Syst.*, 2002, pp. 157–164.
- [49] 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, 2015.
- [50] D. Berrar, "Cross-validation," *Encyclopedia of Bioinformatics and Computational Biology*, pp. 542–545, 2019.
- [51] X. Zhang, J. Li, Y. Liu, Z. Zhang, Z. Wang, D. Luo, X. Zhou, M. Zhu, W. Salman, G. Hu, and C. Wang, "Design of a fatigue detection system for high-speed trains based on driver vigilance using a wireless wearable EEG," *Sensors*, vol. 17, no. 3, p. 486, 2017.
- [52] M. Hajinoroozi, Z. Mao, T.-P. Jung, C.-T. Lin, and Y. Huang, "EEG-based prediction of driver's cognitive performance by deep convolutional neural network," *Signal Process. Image Commun.*, vol. 47, pp. 549–555, 2016.
- [53] R. G. Hefron, B. J. Borghetti, J. C. Christensen, and C. M. S. Kabbani, "Deep long short-term memory structures model temporal dependencies improving cognitive workload estimation," *Pattern Recognit. Lett.*, vol. 94, pp. 96–104, 2017.
- [54] D. Zhang, L. Yao, K. Chen, S. Wang, X. Chang, and Y. Liu, "Making sense of spatio-temporal preserving representations for EEG-based human intention recognition," *IEEE Trans. Cybern.*, vol. 50, no. 7, pp. 3033–3044, 2019.
- [55] P. Bashivan, I. Rish, M. Yeasin, and N. Codella, "Learning representations from EEG with deep recurrent-convolutional neural networks," arXiv preprint arXiv:1511.06448, 2015.
- [56] O. N. Rahma and A. Rahmatillah, "Drowsiness analysis using common spatial pattern and extreme learning machine based on electroencephalogram signal," *J. Med. Signals Sens.*, vol. 9, no. 2, p. 130, 2019.
- [57] 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. 112 767–112 776, 2019.
- [58] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert, "Deep learning-based electroencephalography analysis: a systematic review," *J. Neural Eng.*, vol. 16, no. 5, p. 051001, 2019.
- [59] H. Zeng, C. Yang, G. Dai, F. Qin, J. Zhang, and W. Kong, "EEG classification of driver mental states by deep learning," *Cogn. Neurodyn.*, vol. 12, no. 6, pp. 597–606, 2018.
- [60] Y. Ma, B. Chen, R. Li, C. Wang, J. Wang, Q. She, Z. Luo, and Y. Zhang, "Driving fatigue detection from EEG using a modified PCANet method," *Comput. Intell. Neurosci.*, vol. 2019, 2019.
- [61] J. Chen, H. Wang, and C. Hua, "Assessment of driver drowsiness using electroencephalogram signals based on multiple functional brain networks," *Int. J. Psychophysiol.*, vol. 133, pp. 120–130, 2018.
- [62] J.-R. Yen, C.-C. Hsu, H. Yang, and H. Ho, "An investigation of fatigue issues on different flight operations," *J. Transp. Manag.*, vol. 15, no. 5, pp. 236–240, 2009.
- [63] D. A. Wiegmann and S. A. Shappell, *A human error approach to aviation accident analysis*. Aldershot, UK: Ashgate, 2003.
- [64] R. T. Schirrmester, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Hum. Brain Mapp.*, vol. 38, no. 11, pp. 5391–5420, 2017.
- [65] 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*, 2020.
- [66] 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. 15 945–15 954, 2019.



DAE-HYEOK LEE received his B.S. degree in bioengineering from UNIST, South Korea, in 2018. He is a master student in the Department of Brain and Cognitive Engineering at Korea University, South Korea. His research interests include signal processing, deep learning, and brain-computer interface.



JI-HOON JEONG received his B.S. degree in computer information science from Korea University, South Korea, in 2015. He is an Ph.D. student in the Department of Brain and Cognitive Engineering at Korea University, South Korea. His research interests include machine learning, deep learning, and brain-machine interfaces.



KIDUK KIM received his B.S. and M.S. degrees in electronics from Kyungpook National University, South Korea, in 1994 and 1996, respectively. He is currently pursuing the Ph.D. degree in brain and cognitive engineering at Korea University, South Korea. From 1994 to 2001, he worked for Daewoo Electronics Co., Ltd., where he was involved in the improvement of image and audio quality in consumer electronics. He has worked at LG Display Co., Ltd., since 2002, where he is involved in R&D for a variety of human–display interfaces. His main research interests include flat-panel displays as well as new human–display interface technologies.



BAEK-WOON YU received his B.S. degree in robotics (information control) from Kwangwoon University, South Korea, in 2015. He received his M.S. degree in brain and cognitive engineering at Korea University, South Korea. His research interests include brain-computer interface, physiological signal processing, and deep learning.



SEONG-WHAN LEE (S'84–M'89–SM'96–F'10) Seong-Whan Lee 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. Dr. Lee is a fellow of the International Association of Pattern Recognition and the Korea Academy of Science and Technology.

...