Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/bbe

Original Research Article

Classification of pilots' mental states using a multimodal deep learning network

Soo-Yeon Han [a,1], No-Sang Kwak [a,1], Taegeun Oh [b], Seong-Whan Lee [a,c,*]

[a] Department of Brain and Cognitive Engineering, Korea University, Seoul, Korea
[b] Agency for Defense Development, Daejeon, Korea
[c] Department of Artificial Intelligence, Korea University, Seoul, Korea

ARTICLE INFO

Article history:
Received 24 July 2019
Received in revised form
20 October 2019
Accepted 4 December 2019
Available online

Keywords:
MDL
EEG
ECG
Respiration
EDA
Pilot's mental states

ABSTRACT

An automation system for detecting the pilot's diversified mental states is an extremely important and essential technology, as it could prevent catastrophic accidents caused by the deteriorated cognitive state of pilots. Various types of biosignals have been employed to develop the system, since they accompany neurophysiological changes corresponding to the mental state transitions. In this study, we aimed to investigate the feasibility of a robust detection system of the pilot's mental states (i.e., distraction, workload, fatigue, and normal) based on multimodal biosignals (i.e., electroencephalogram, electrocardiogram, respiration, and electrodermal activity) and a multimodal deep learning (MDL) network. To do this, first, we constructed an experimental environment using a flight simulator in order to induce the different mental states and to collect the biosignals. Second, we designed the MDL architecture – which consists of a convolutional neural network and long short-term memory models – to efficiently combine the information of the different biosignals. Our experimental results successfully show that utilizing multimodal biosignals with the proposed MDL could significantly enhance the detection accuracy of the pilot's mental states.

1. Introduction

Many studies have demonstrated that pilot's cognitive capabilities could affect flight safety because aircraft control requires high cognitive skills [1–3]. Indeed, more than 70% of accidents in aviation are caused by human error, which is directly attributed to failures in cognitive performance [4]. Lowered cognitive capabilities may occur in various mental states, such as distraction, workload, and fatigue. Distraction is the process of diverting the attention, which blocks or diminishes the reception of desired information [5]. This state

* *Corresponding author at*: Department of Brain and Cognitive Engineering, Korea University, Seoul, Korea.
E-mail addresses: syhan09@korea.ac.kr (S.-Y. Han), nskwak@korea.ac.kr (N.-S. Kwak), tgoh@add.re.kr (T. Oh), sw.lee@korea.ac.kr (S.-W. Lee).

[1] These authors contributed equally to this work..

https://doi.org/10.1016/j.bbe.2019.12.002

may appear when pilots divert their attention away from the main flight task. Workload refers to the cognitive and psychological efforts required to complete given tasks [1,6]. This state is usually caused by excessive flight missions. Fatigue is defined as a sensation of boredom and unwillingness to carry on a task, and may lead to reduced work efficiency and increased possibility of accidents [2,3,6]. Such state could occur after a lengthy time of monotonous flight.

For preventing the accidents caused by deteriorated cognitive states, several different approaches based on biosignals have been used to explore three major aspects of the subjects' cognitive states, i.e., sensor modality, types and number of mental states, and machine learning method. Regarding the sensor modality, a single sensor [5,7–14] and multi-sensors [15–18] have been used for extracting information related to the subject's mental states. For example, electroencephalogram (EEG), functional near-infrared spectroscopy (fNIRS), electrocardiogram (ECG), electromyogram (EMG), electrodermal activity (EDA), and respiration have been widely used for detecting mental states because the dynamics of these signals are linked to the pilots' cognitive states.

A single sensor-based system could provide users with comfortable interfaces, while multi-sensors could enhance the system's performance by combining the various distinct modalities [19]. Note that, in comparison with other sensors, EEG appears to provide the most accurate information to classify mental states [17]. It is the most essential physiological signal to analyze mental states because it directly reflects cortical activities [20]. Moreover, compared with other brain monitoring methods, EEG is more versatile, easy to setup, comfortable, non-invasive, and safe [21–25]. Therefore, EEG has been investigated in various domains in addition to classification of mental states [26–29]. Despite its advantages, a significant limitation of EEG is its signal-to-noise ratio, as multiple kinds of noise may easily be induced by biological or environmental sources [30]; indeed, the main artifacts in EEG are caused by eye blinks, muscle contractions, and electronic devices [31]. Furthermore, some characteristics of EEG might vary from subject to subject [32]. To compensate for the shortages of EEG, various researchers have adopted a multimodality approach by combining EEG with non-brain measures (i.e., peripheral physiological measures (PPMs)) such as ECG, respiration, and EDA [1,16–18] for mental state assessment. For example, the combined features of EEG and ECG show a higher accuracy for workload classification than EEG alone [18]. Further, EEG, fNIRS, and physiological measures (ECG and respiration) provide better results than the EEG-based classification [17].

Various types of mental states such as distraction, workload, fatigue, drowsiness, and stress were selected to develop the system. However, most research aimed to distinguish each state from a normal state (i.e., two-class problem) [5,7,8,11,13,14,16], or to divide a single state into three or four levels in the single state [9,10,12,15,17,18]. To the best of our knowledge, there have been no attempts to recognize various mental states (i.e., multi-class classification) simultaneously using only biosignals. From the point of view of machine learning method, for the EEG research, various classification methods have been applied, including conventional machine learning techniques (e.g., linear discriminant analysis (LDA) [5,8,16,17], support vector machine (SVM) [18,33], etc.) and deep learning techniques (e.g., convolutional neural network (CNN) [13], Recurrent 3DCNN [11], autoencoder [12], etc.). However, previous multi-modality studies have used only conventional machine learning techniques (see details in Section 2).

Here, our study aimed to investigate the feasibility of detecting – with efficiency and robustness – various mental states of the pilot by means of multimodal biosignals (i.e., EEG, ECG, respiration, and EDA) and multimodal deep learning (MDL). MDL offers the advantage that a hierarchical representation can be automatically learned for each modality.

Hence, the main contribution of this work is threefold: first, we used a flight simulator to build an experimental environment that allows us to induce the four mental states in the pilot and to collect multimodal biosignals. Second, we acquired multimodal data from eight pilots who had flight experience. Finally, we designed the MDL architecture, which consists of CNN and LSTM models, in order to fuse the various neurophysiological signals efficiently. Our study successfully shows that the combined use of various biosignals and the MDL network could enhance the classification accuracy of the pilot's mental states.

2. Related works

As a single modality, EEG in particular, has been widely used in combination with various machine learning methods. For example, Lal et al. [34], Jap et al. [35], Kar et al. [36], and Trejo et al. [37] analyzed the statistical changes of EEG that occurred during a driver simulator task for fatigue detection. Sonnleitner et al. [5] used regularized LDA (rLDA) to investigate whether the EEG could predict the subject's distraction on a single trial analysis. Chaudhuri and Routray [7] focused on classifying the normal and fatigue states in a simulation environment using the source localization method and an SVM classifier. Dehais et al. [8] classified the mental workload and normal states using the frequency features based on shrinkage LDA (sLDA). However, these studies used handcrafted EEG features to build classifiers. Also, ECG has been widely used in recognizing mental states. Rogado et al. [38] and Jung et al. [39] designed an embedded ECG sensor on the steering wheel for analyzing driver fatigue and drowsiness. These studies did not classify between normal state and fatigue state, but performed a statistical analysis of each state.

Recently, deep learning has been utilized, as it can learn hierarchical features without any outside assistance. For example, Patel et al. [14] applied a neural network using a set of ECG data for detecting early onset of fatigue in drivers. Bashivan et al. [9] proposed a deep recurrent convolutional neural network (RCNN) to detect four workload states from multi-channel EEG signals. Hajinoroozi et al. [40] proposed a channel-wise convolutional neural network (CCNN) and a CCNN variation that uses a restricted Boltzmann machine (CCNN-R) to detect poor performance in the driver. Jiao et al. [10] proposed a deep CNN method for detecting four mental workload levels from the EEG data. Furthermore, they introduced a pointwise gated Boltzmann machine (PGBM)-based fusion strategy of two CNN models for different EEG

inputs. Zhang et al. [11] utilized recurrent 3D CNN (R3DCNN) to learn spatial-spectral-temporal EEG features for cross-task mental workload assessment. Wu et al. [12] proposed a deep stacked contractive autoencoder network (DCAEN) to learn the fatigue-related features from raw EEG data in order to recognize the pilot's fatigue status. Gao et al. [13] developed an EEG-based spatial-temporal convolutional neural network (ESTCNN) to detect the subject's state of fatigue with high accuracy. However, these studies used the single modality alone to detect the cognitive states.

Using multimodal sensors is an effective way to improve the detection performance compared with a single sensor-based recognition. Also, various combinations of biosignal modalities (e.g., EEG, ECG, photoplethysmogram, EOG, EDA, respiration, and EMG) were used for analyzing the fatigue states [41–46]. Hogervorst et al. [15] tested the combined information of EEG, skin conductance, respiration, ECG, pupil size, and eye blinks for mental workload estimation. Ahn et al. [16] collected EEG, fNIRs, and ECG data simultaneously to develop algorithms that allow researchers to explore the neurophysiological correlates of subjects' state of fatigue. The combination of LDA methods yielded substantial improvements in the ability to discriminate between well-rested (i.e., normal) state and sleep deprived (i.e., fatigue) state. Liu et al. [17] integrated EEG, fNIRS, and physiological measures for the classification of three workload levels in an *n*-back working memory task. They showed that the fusion of these modalities could improve the classification performance. Zhang et al. [18] used EEG and ECG signals to validate the effectiveness of the interactive mutual information modeling (IMIM), which is a feature-weight-driven signal-fusion method based on mutual information. However, to the extent of our knowledge, multimodal biosignals have not been combined with MDL methods (See the summary of related works in Table 1).

From these multi-modality studies, hand-crafted features for each biosignal were extracted, and were then used to classify the mental states. For example, in EEG signals, power spectral density (PSD) features in specific frequency bands including delta (1-4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (30–40 Hz) were normally used [6,12]. In ECG signals, time-domain features (e.g., mean heart rate (MHR) and standard deviation of normal-to-normal intervals (SDNN)) and frequency-domain features (e.g., PSD from 0.04 to 0.15 Hz frequency range for low frequency band (LF) and 0.15 to 0.40 Hz range for high frequency band (HF), and that of the ratio (LFHF)) were employed [16,47–49]. In respiration signals, the standard deviations of the amplitude from the abdomen and thorax channels (i.e., SDAbd and SDThor, respectively) were used as time-domain features [17]. Further, the PSD of the dominant respiration frequency band from the abdomen and thorax channels (i.e., DRFAbd and DRFThor, respectively) were extracted as frequency-domain respiration features. For the EDA signals, the mean amplitude of EDA (MEDA) and the standard deviation of the amplitude (SDEDA) were used as time-domain features. Also, the PSD extracted from the EDA index of the frequency bands of the sympathetic nervous system (EDASymp) was investigated for frequency-domain features [15,50].

3. Methods

3.1. Experiment

3.1.1. Participants

Eight healthy subjects (6 males and 2 females, age: 25.7 ± 2.6) underwent flight experience for over 100 hours in the Taean Flight Education Center. This study was reviewed and approved by the Institutional Review Board at Korea University [1040548-KU-IRB-18-92-A-2], and written informed consent was obtained from all participants before the experiments. All subjects had a normal or corrected-to-normal vision, normal hearing, and no history of psychiatric or neurological diseases. They were asked to refrain from alcohol and coffee and to sleep (6–8 h) before the experiment. They were instructed to fill out the questionnaires for recording the subjects' status and for evaluating our experimental paradigm.

3.1.2. Experimental setup

We designed an experimental environment using a flight simulator system (Cessna 172, FRASCA International, Inc.) (see Fig. 1). The cockpit consisted of the wide visual display

Table 1 – Summary of related works: N, D, W, and F indicate normal, distraction, workload, and fatigue, respectively. Moreover, W_i and F_i indicate *i* levels of the corresponding mental states. Note that performance values are the representative results of the referred studies. Performance could be different depending on the experimental conditions.

References	Modalities	# of class (types)	# of subjects	Methods	Performance
Sonnleitner et al. [5]	EEG	2 (D/N)	20	rLDA	92%
Chaudhuri et al. [7]	EEG	2 (F/N)	12	Source feature, SVM	86.8%
Dehais et al.[8]	EEG	2 (W/N)	18	sLDA	70.8%
Bashivan et al. [9]	EEG	4 (W_1/W_2/W_3/W_4)	13	RCNN	91.1%
Jiao et al. [10]	EEG	4 (W_1/W_2/W_3/W_4)	13	CNN, PGBM	92.4%
Zhang et al. [11]	EEG	2 (W/N)	20	R3DCNN	88.9%
Wu et al. [12]	EEG	3 (F_1/F_2/N)	40	DCAEN	91.7%
Gao et al. [13]	EEG	2 (F/N)	8	ESTCNN	97.4%
Patel et al. [14]	ECG	2 (F/N)	12	Neural network	90%
Hogervorst et al. [15]	EEG, eye-tracking measures	2 (W_1/W_2)	14	Logistic regression	91%
Ahn et al. [16]	EEG, fNIRs, ECG	2 (F/N)	11	LDA	75.9%
Liu et al. [17]	EEG, fNIRS, ECG, respiration	3 (W_1/W_2/W_3)	21	LDA, Naive-Bayes	65.1%
Zhang et al. [18]	EEG, ECG	3 (W_1/W_2/W_3)	10	IMIM, SVM	89.9%

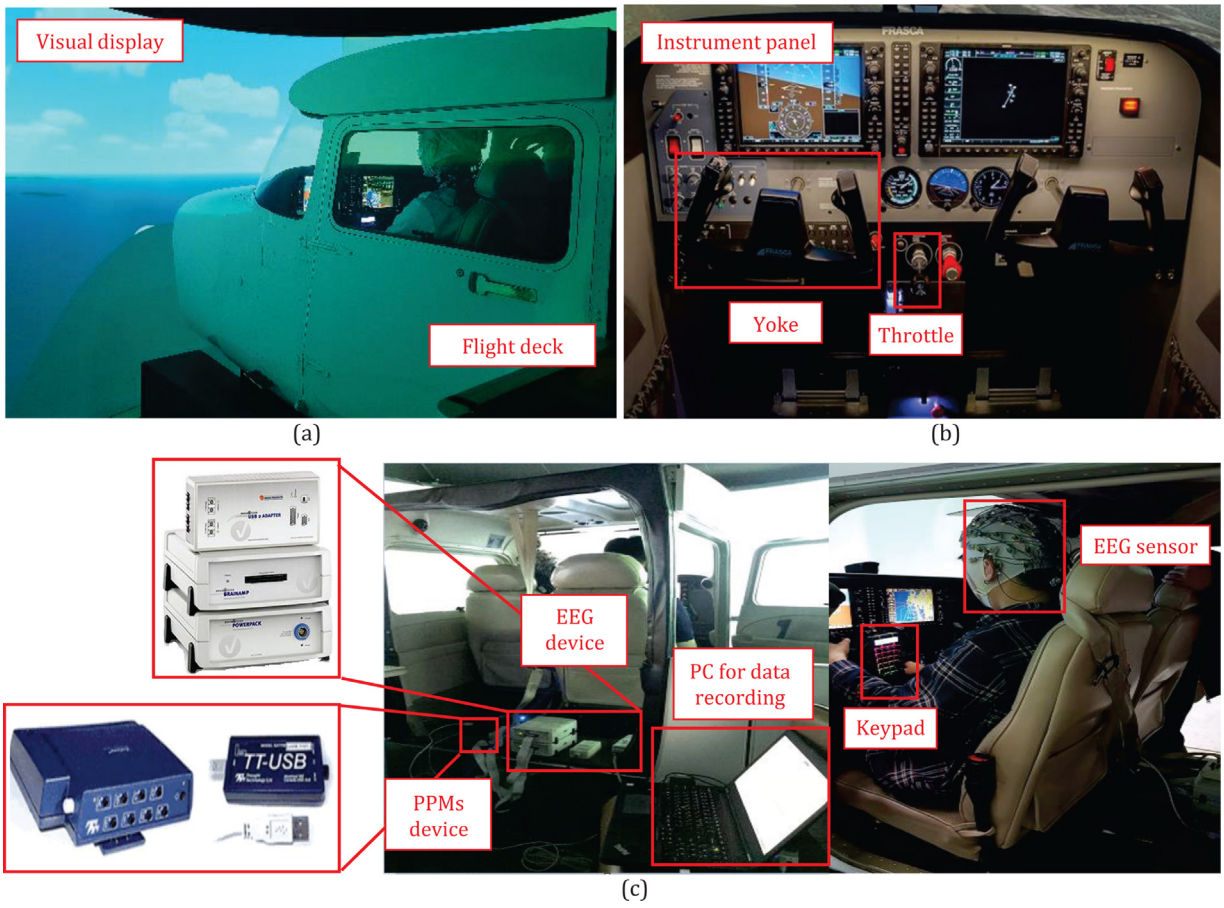

Fig. 1 – Experimental environment. (a) External environment of the flight simulator. (b) Internal environment of the flight deck. (c) Experimental setup in the flight simulator.

providing an angle view of 210°, a flight yoke, an electric control loading system, other control panels, etc. A wireless keypad was attached on the yoke to record the pilot's responses. We measured EEG signals using the EEG system (BrainAmp, Brain Products GmbH). A total of 60 electrodes were located on the scalp, according to the international 10–20 system, as shown in Fig. 2(a). Additionally, we acquired the EOG signals using four electrodes simultaneously; two electrodes were placed at the outer canthi to measure the horizontal component and two electrodes, one above and the other below the left eye, were used for the vertical component [51]. The reference electrode was placed at the FCz, and the ground was placed at the AFz. The sampling frequency of EEG and EOG was 250 Hz. The peripheral physiological measures (PPMs) (i.e., ECG, respiration, and EDA) were carried out using the Flexcomp system with biograph Infiniti software (Thought Technology Ltd.). Three ECG electrodes were attached to the chest wall equidistant from the heart related to the specific limb, such as left arm (LA), right arm (RA), and left leg (LL) according to the standard 3-lead configuration. A respiration sensor was fastened around the thorax and abdomen. EDA was measured at the index and ring fingers of the left hand. The sampling frequency of PPMs was 256 Hz. We developed the proposed method using Python 3.7, based on the Pytorch library.

3.1.3. *Experimental paradigm*

We designed experimental paradigms to induce different cognitive states in the pilots, including distraction, workload, and fatigue (see Fig. 3) following a similar procedure as in previous studies [5,13,52,53]. The data were collected during a period of at least 3.5 h. After finishing each paradigm, subjects were asked how difficult it was to complete the task. Subjects were allowed to rest if they wished to do so. Detailed descriptions of each paradigm are provided in the following paragraphs.

Distraction paradigm: Distraction was induced by an auditory cue that signaled the subject when it was time to perform a secondary task provided in an air traffic control (ATC) message while performing a primary task, which consisted of maintaining the aircraft under predefined conditions. During the simulated flight, subjects were asked to maintain an altitude of 3000 feet and a heading of 0°, while prerecorded sentences of the ATC message were presented at a regular time interval. The pilot was instructed to mentally count the number of words of the ATC message. ATC messages were divided into three levels based on the number of words (Level 1: 4–9 words, Level 2: 10–14 words, Level 3: 15–22 words). A beep sound signaled the pilot when it was time to type the number using the keypad (see Fig. 4).

Workload paradigm: The pilots were instructed to perform a monotonous flight maintaining an altitude of 3000 feet and a

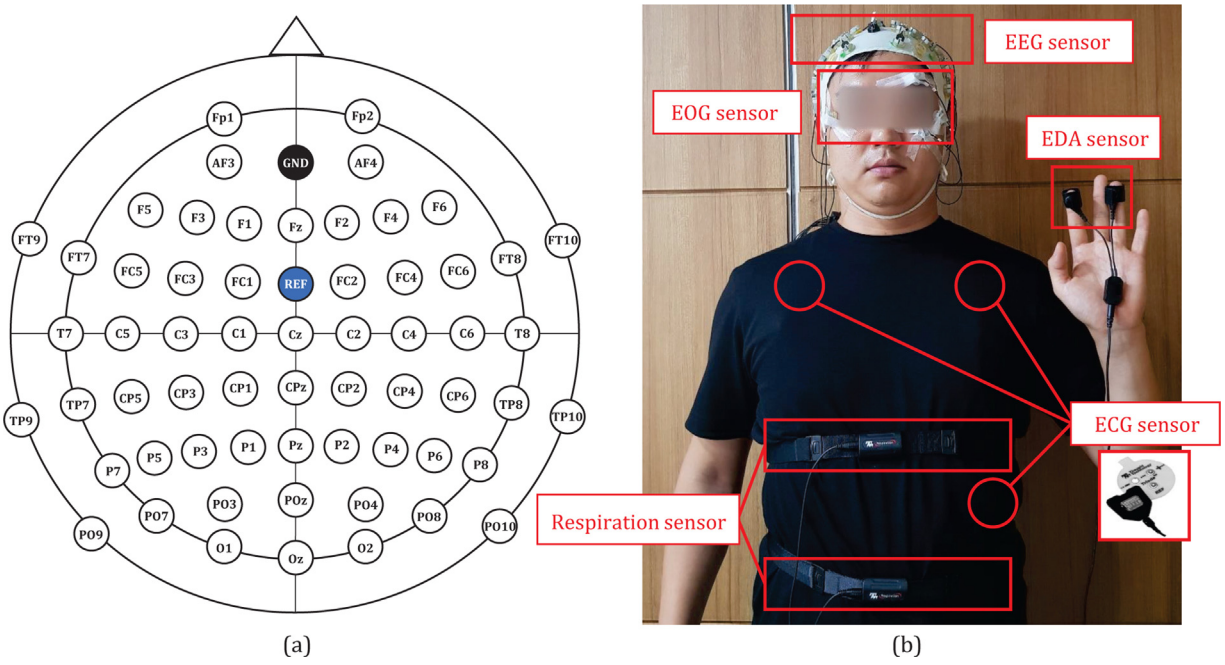

Fig. 2 – Experimental setup. (a) Setup of EEG electrodes. (b) Recording setup of physiological sensors.

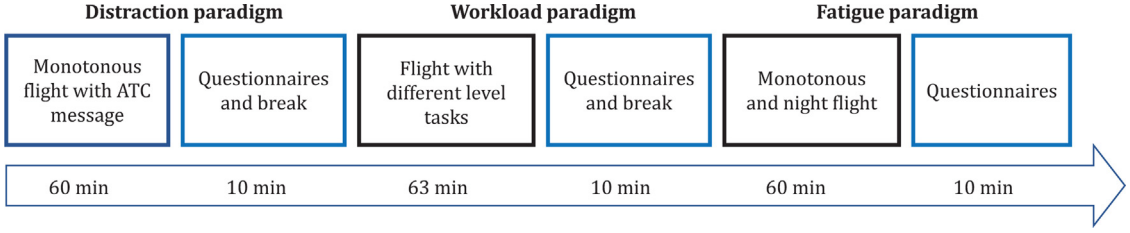

Fig. 3 – Timeline of the experiment.

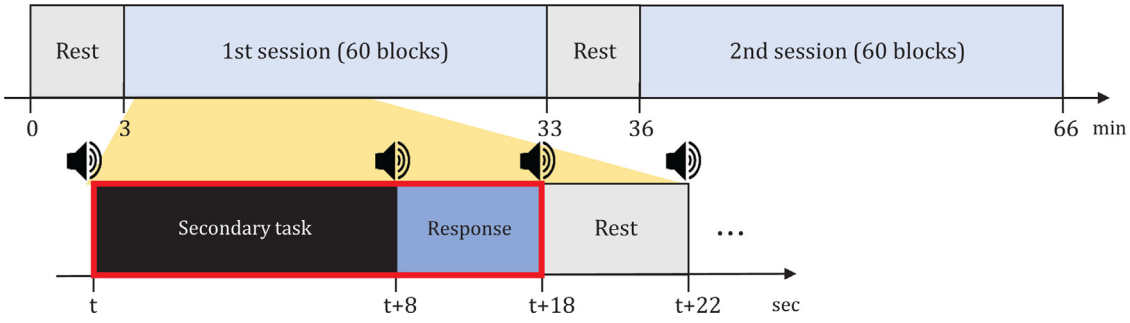

Fig. 4 – Distraction paradigm: data for the distraction class were extracted, as indicated by a red box.

heading of 0°. According to the complexity of the given task, the instructions were divided into three levels (Level 1: changing the altitude at a given speed, Level 2: changing the altitude, heading, and speed, Level 3: steep turn at a given bank angle, and roll-out with a given heading). After the execution of the given tasks, the subject was asked to mark the endpoint using the keypad. The success or failure of the tasks was reported by the instructor. The time to complete each task could vary among levels (see details in Fig. 5).

Fatigue paradigm: Subjects were instructed to take a monotonous night flight, and to maintain an altitude of 3000 feet and a heading of 0°. The beep sound was presented every 1 minute, and the subjects had to report a subjective sleepiness score using the keypad. Score values were based on the Karolinska sleepiness scale (KSS). Note that the KSS has a 9-point scale (e.g., 1 = extremely alert, 5 = neither alert nor sleepy, and 9 = extremely sleepy, fighting sleep). Fig. 6 shows the fatigue experimental paradigm.

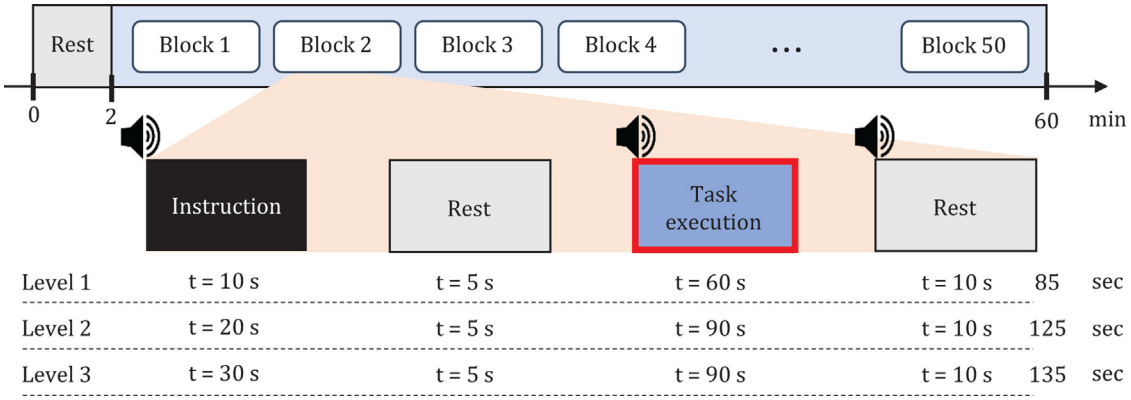

Fig. 5 – Workload paradigm: data for the workload class were extracted, as indicated by a red box.

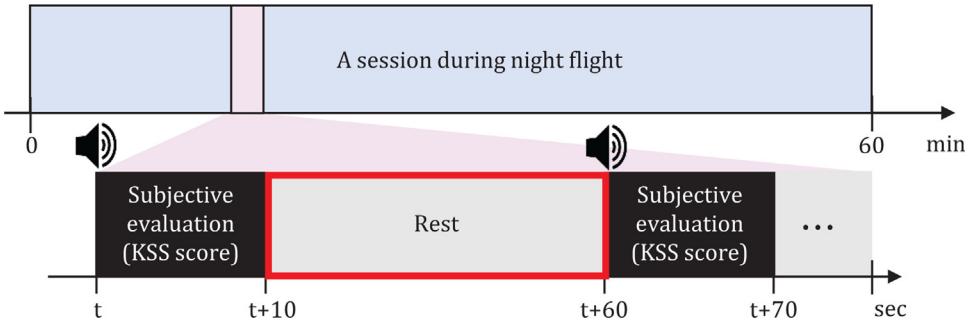

Fig. 6 – Fatigue paradigm: data for the fatigue class were extracted, as indicated by a red box.

3.2. *Data acquisition and preprocessing*

3.2.1. *Data preprocessing*

EEG signals were band-pass filtered at 0.1–50 Hz to remove the direct current voltage and high-frequency artifacts. We also employed an independent component analysis (ICA) algorithm to remove eye-related artifacts of vertical and horizontal components from four EOG channels [16]. ICA is widely used to calibrate the noise sources for artifact removal [54]. For the signals corresponding to the PPMs, we used a low-pass filter at 50 Hz, and the detrending algorithm to eliminate the linear trend, which is the systematic increase or decrease in the PPM data [17]. After that, the collected data from each experimental paradigm were segmented into trials of 5 s with an overlap of 4 s, and we evaluated the performance based on the threefold cross validation (see details in Section 3.4). Note that we totally separated training data and test data preserving sequential information, which means there are no any overlapping data between training and test data. In this study, we obtained the distraction data from every secondary task to the corresponding response in the distraction paradigm (red box in Fig. 4). We analyzed the workload data from the every task execution of each task in the workload paradigm (red box in Fig. 5). We obtained the fatigue data during the rest period of the fatigue paradigm (red box in Fig. 6); here we selected the data when the score of the subjective evaluation was greater than or equal to 6 points. The normal state is obtained during the rest periods in the distraction and workload paradigms. In the fatigue paradigm, we selected the normal state data when the KSS score was less than 5 points. Finally, we obtained a total of approximately 6000 samples per subject. Note that a sample indicates 5 s length of data from four modalities (i.e., EEG, ECG, respiration and EDA) corresponding to four mental states. Hence, 6000 samples per subject means that we acquired 1500 data samples from specific mental state from single subject.

For the EEG input data of the proposed method, we transformed the preprocessed EEG data into three topographic maps. First, we calculated the power spectrum density in five frequency bands (i.e., delta, theta, alpha, beta, and gamma bands) based on fast Fourier transform (FFT) [55,56]. Then, we selected three discriminant frequency bands, with high mutual information between class label and the corresponding frequency band power from the training data [18]. As follows [9], we applied a polar projection to project electrode locations on 2D surface. Also, we applied a Clough–Tocher scheme for interpolating the scattered power measurements over the scalp and for estimating the values in-between the electrodes over a 32×32. The three maps were provided as an input to the CNN model in the proposed method. The preprocessed PPMs were provided to the LSTM models (see details in Section 3.3). Thus, the dimension of each input of the EEG was $3@32 \times 32$ and that from each PPM was 1×1280 (256 sampling rate $\times$ 5 s).

3.3. *The proposed network*

Different types of sensors have inherent representations and statistical properties, which complicates the detection of com-

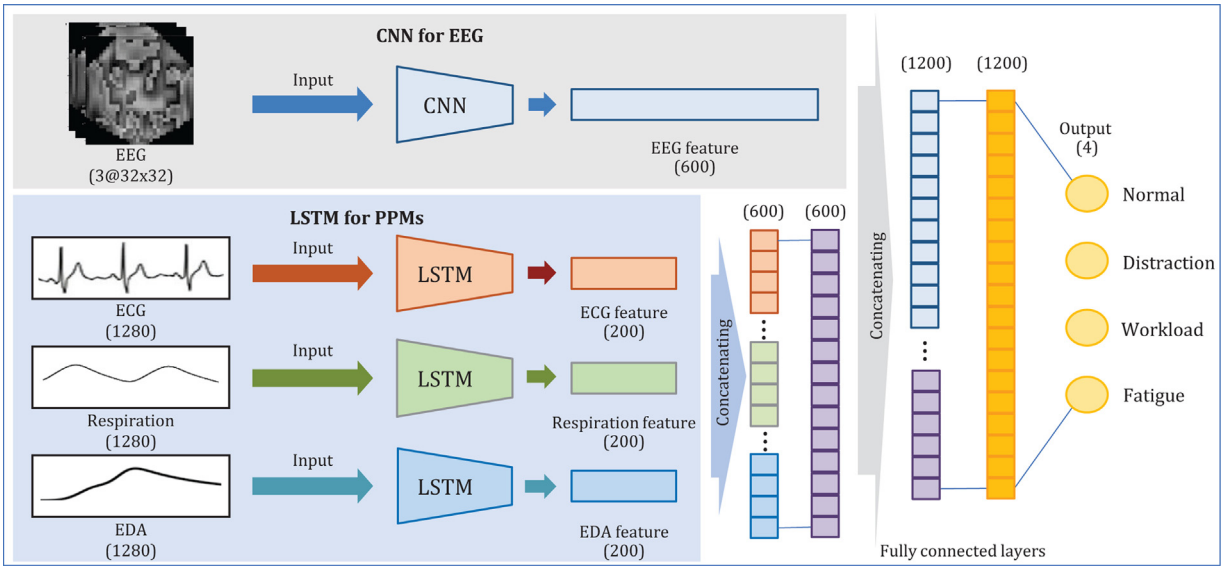

Fig. 7 – Overview of the proposed MDL network.

mon features across modalities [57]. Hence, we designed the MDL architecture to learn hierarchical features from the different types of biosignals. We utilized a CNN model for EEG signals and three LSTM models for PPMs. Further, the proposed network fused each feature from the sub-networks in a fully connected layer (Fig. 7). It is advantageous to reduce the influence of inherent properties of each modality. Details are described as follows.

3.3.1. *CNN for EEG*

We used a CNN architecture with three hidden layers (i.e., two convolutional layers and a fully connected layer). The convolutional layers used small receptive fields of size 3 × 3, and a stride of 1 unit with a rectified linear (ReLU) activation function. Each input of the convolutional layers was padded with 1 unit to preserve the spatial resolution after the convolution. Max pooling was performed over a 2 × 2 window with 2 strides. The fully connected layer outputs 600 units (see details in Table 2).

3.3.2. *LSTMs for PPMs*

Three LSTM models were employed for three PPMs, each of which is a single channel-based time series signal. LSTM has a memory ability, and partially removes the risk of vanishing gradients. LSTM is mainly comprised of three gates (a forget gate, an input gate and an output gate) and five parameters (i.e., forget gate's activation vector, input/update gate's activation vector, output gate's activation vector, hidden state vector, and cell state vector). This structure enables the LSTM unit to forget useless outdated information, and to update itself based on the new input [58]. This study applied an LSTM layer. After pre-processing, each PPMs raw signal (1 × 1280) was used as the input in each LSTM model, which has a single LSTM layer with 200 hidden units (see Table 3).

3.3.3. *Fully connected layers for classification*

To integrate the features of each modality, we concatenated the features of PPMs. The concatenated features were trained through a hidden fully connected layer consisting of 600 units. Then, the outputs of the layer were merged with the EEG features of 600 units into new 1200 units. The merged features were trained through the fully connected layer of 1200 units. Finally, the class was determined by the four output units, using the softmax classifier to indicate the normal, distraction, workload, and fatigue states.

3.3.4. *Training*

Training was carried out by optimizing the cross-entropy loss function. ReLU and hyper-tangent were respectively used as activation functions for CNN and LSTM. We used Adam as a gradient descent optimizer with a learning rate of 0.001, and the decay rate of the first and second moments were 0.9 and 0.99, respectively. The batch size was set to 20. The dropout

Table 2 – CNN configurations in the proposed network.

Type	Kernel/stride/padding	Output size
Input	–	3@32 × 32
Convolutional layer	3 × 3/1/1	32@32 × 32
Max pooling	2 × 2/2/–	32@16 × 16
Convolutional layer	3 × 3/1/1	64@16 × 16
Max pooling	2 × 2/2/–	64@8 × 8
Fully connected layer	–	600

Table 3 – LSTM configurations in the proposed network.

Type	Number of units	Output size
Input	–	1280
LSTM layer	200	200

probability was set to 0.5 in all fully connected layers. Our network parameters converged after 150 epochs; the iteration was selected by monitoring the models' performance. We did not use early stopping.

3.4. Evaluation and baseline methods

To validate the performance of the proposed method, we used a k-fold cross validation method, which is a resampling procedure used to evaluate models on a limited data sample. The advantage of cross validation is that it can increase the statistical reliability of the categorizer performance measurement [59,60]. Note that a value of $k = 10$ is commonly used. However, in general, k remains an unfixed parameter. In our study, we selected threefold cross validation. The reason is that we need to train deep learning model with sufficient data while increasing the statistical reliability. However, larger k value could increase training time of the proposed multimodal deep learning model. This study benchmarked the proposed method with several traditional machine learning methods and state-of the art methods. We used the Scikit-learn library for k-NN, random forest, logistic regression, support vector, and shrinkage linear discriminant analysis. The Pytorch library was used for LSTM and DCAEN. We compared the classification performances using EEG-alone, PPMs-alone and both EEG and PPMs. Here, we briefly present an explanation of the baseline methods. Input features of each method are summarized in Table 4.

k-nearest neighbor (k-NN): it classifies a new data by the majority of vote from its neighbors, with the object being assigned to the k nearest class [61]. In our study, we selected the number of neighbors in a cluster {1, 3, 5, 7, 10, 15} that yielded the best accuracy in the training phase.

Random forest (RF): RF is an ensemble classification method of multiple decision trees that depends on the random vector sampled independently and with the same distribution [62]. In this study, each input and outputs of all trees were computed. And the class with the majority of votes was selected. We selected the number of estimators within {5, 10, 50, 100, 1000, 2000, 3000} that yielded the highest accuracy in the training phase.

Logistic regression (LR): LR is a statistical technique to predict and classify the likelihood of an event using a linear combination of independent variables [63,64]. In our study, l_2-regularization was used to consider sparsity in the logistic regression model. The regularization parameter was selected within the range of $[10^{-2}, 10^{2}]$ in the training phase.

SVM: SVM finds the boundary with the largest margin for classification [65,66]. In our study, regularization penalty parameter C and inverse of RBF kernel's standard deviation (γ) were selected within parameter sets $C = \{0.01, 0.1, 1, 10, 100\}$, and $\gamma = \{0.01, 0.1, 1, 2, \ldots, 10\}$, respectively.

sLDA: LDA aims to find a boundary that can distinguish the categories after projecting the data on a specific axis [67]. The LDA separates the mean values of the categories from each other and reduces the variance. sLDA is designed to prevent a covariance matrix from becoming singular due to the small sample size for monitoring the pilot's mental workload using spectral power [8]. An automatic shrinkage using the Ledoit–Wolf lemma was adopted in this study.

LSTM: LSTM deals with the exploding and vanishing gradient problems that can be encountered when training traditional RNNs. A common LSTM model is composed of units, an input gate, an output gate and a forget gate [58]. In this study, we composed the network with a hidden LSTM layer and 200 hidden units.

DCAEN: A deep stacked contractive autoencoder network (DCAEN) with the softmax classifier was proposed to learn the EEG features that allow to recognize the mental fatigue status of pilots [12]. It established the network with three hidden layers. Following that study, the hidden layer nodes designed in the study were 800-400–50, and the contractive coefficient parameter λ was 0.3.

4. Results and discussion

4.1. Physiological signals analysis

In Fig. 8, we present the box plots for the features of PPMs (described in Table 4) of all trials corresponding to the mental states. From the top of the box plots, each row represents EEG, ECG, respiration, and EDA respectively. The central mark in each box represents the median, and the bottom and top edges of the box represent the first and a third of quartile (Q_1 and Q_3) for the range. The interquartile range (IQR) is calculated as $Q_3 - Q_1$. Note that the box-plots were averaged from all trials of all subjects. A more detailed investigation is needed to consider individual variation. In the EEG features, we calculated an averaged power from the six mid-line channels (Fz, FCz, Cz, CPz, Pz, and Oz) for each frequency band (delta, theta, alpha, beta, and gamma). For the delta and theta activity, the increase in the IQR of the distraction was observed. In all frequency bands, we commonly found the decreased median values in the workload state. Such a

Table 4 – Summary of input features and dimensions used in the baseline methods. The methods which are not mentioned used same input features with the proposed method.

Method	Modality	Input features	Dimension
K-NN, RF, LR, SVM, sLDA	EEG	PSD of 60 channels at three bands	60 × 3
	ECG	MHR, SDNN, LF, HF, LFHF	1 × 5
	Respiration	SDAbd, SDThor, DRFAbd, DRFThor	1 × 4
	EDA	MEDA, SDEDA, EDASymp	1 × 3
LSTM	EEG	Single trial of 60 channels (channel × sample)	60 × 250
DCAEN	EEG	Flattened input image (3@32 × 32)	3072

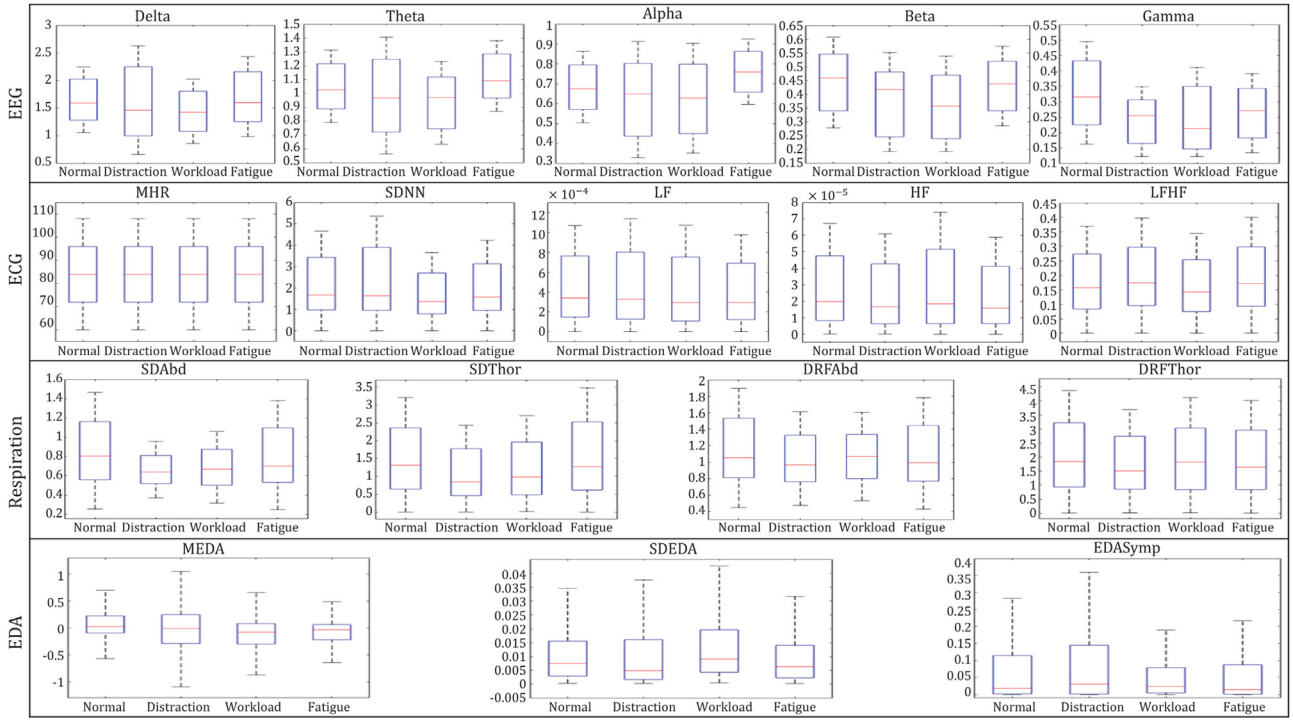

Fig. 8 – Effect of various mental states on the biosignals using the box-plot for all participants' trials. The center mark in each box represents the median, and the bottom and top edges of the box represent the 25th and the 75th percentile.

decrease had also been reported in the previous study [17]. We also found an increase in the median values of the theta and alpha and a decrease in the median value of the beta in the fatigue state; it is a similar result to [16]. In the gamma activity, the median values in all states tend to be decreased compared to the normal state. In the ECG features, we observed a decreased IQR of the SDNN in the workload state. However, in overall, we could not observe the evident changes according to transitions of mental states. In the respiration, the median values and IQR in the distraction and workload states were decreased in the SDAbd and SDThor. In the EDA features, the IQR of the MEDA and EDASymp in distraction state was increased in comparison with the others.

4.2. *Classification performance with baseline methods*

Table 5 shows the averaged classification accuracy of the proposed and baseline methods in the three conditions (based on EEG-alone, based on PPMs-alone, and based on both EEG and PPMs) with the threefold cross validation. Classification results based on EEG-alone were higher than on PPMs-alone in all methods. When using EEG-alone, the sLDA classifier showed the highest performance ($78.2 \pm 5.7\%$). When using PPMs-alone, the proposed and the LSTM methods showed the highest accuracy of $72.5 \pm 8.4\%$; they have same LSTM network for PPMs. We conjecture that the high performance was because it can integrate and learn each PPM feature simulta-

Table 5 – Classification performance of the pilots' four mental states with baseline methods based on EEG-alone, PPMs-alone, and both EEG and PPMs with a threefold cross validation.

Methods	Modalities		
	EEG-alone (Acc.)	PPMs-alone (Acc.)	Both EEG and PPMs (Acc.)
k-NN	$69.7 \pm 7.7\%$	$69.0 \pm 8.4\%$	$76.1 \pm 6.4\%$
RF	$71.7 \pm 8.2\%$	$70.8 \pm 8.8\%$	$81.0 \pm 7.8\%$
LR	$76.7 \pm 7.2\%$	$70.4 \pm 8.4\%$	$80.0 \pm 5.6\%$
SVM	$77.6 \pm 6.1\%$	$70.2 \pm 7.6\%$	$80.5 \pm 5.0\%$
LSTM	$65.9 \pm 8.2\%$	$72.5 \pm 8.4\%$	$76.5 \pm 7.2\%$
sLDA [8]	$78.2 \pm 5.7\%$	$71.1 \pm 10.1\%$	$81.4 \pm 4.5\%$
DCAEN [12]	$67.5 \pm 8.2\%$	$54.7 \pm 8.2\%$	$55.6 \pm 6.5\%$
Proposed method	$77.7 \pm 6.1\%$	$72.5 \pm 8.4\%$	$85.2 \pm 4.3\%$

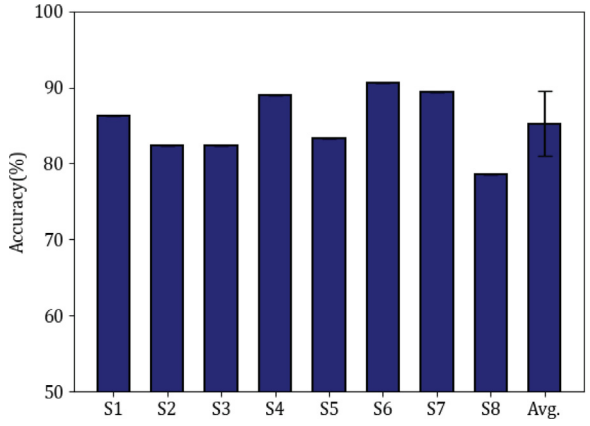

Fig. 9 – Overall performance of the proposed method.

neously through the shared layer. Furthermore, the classification results using both EEG and PPMs showed higher performance than that of using EEG-alone or PPMs-alone in most methods except DCAEN in EEG-alone. These results suggest that the use of multimodal biosignals could be effective in improving the performance of pilots' mental state classification. However, the conventional machine learning methods did not seem to increase the classification accuracy significantly, while the accuracies of deep learning models were largely increased including the proposed method. These results imply that the deep learning method is suitable for multimodal learning because it's ability to extract the high-level representation.

We found that the accuracy accomplished by our model based on multimodal biosignals (85.2 ± 4.3%) was higher than that of other models. A t-test was performed between the classification accuracy of the proposed model and that of the baseline methods, respectively. The statistical analysis showed that the accuracy of the proposed methods was the significantly increased compared with other methods ($p < 0.05$). Moreover, the proposed method was stably effective for all subjects except subject 2 and 8, as evidenced by the small standard deviations. Our framework can capture the information robustly and effectively from the multimodal biosignals, due to its ability to fuse the spatial-spectral features from EEG and temporal features from PPMs. Fig. 9 shows the average individual performances of the proposed method of threefold cross validation for eight subjects. The average accuracy achieved 85.2 ± 4.3%. The best performing subject (S6) accomplished an accuracy of 90.6 ± 5.1%, while subject (S8) recorded the lowest accuracy (78.5 ± 6.7%).

4.3. *Confusion matrix*

We also evaluated the degree of confusion in the different classifiers. The confusion matrix for 3-fold cross validation results using the *k*-NN classifier is shown in Fig. 10(a); the RF classifier was used in (b), LR in (c), SVM in (d), LSTM in (e), sLDA in (f), DCAEN in (g), and the proposed method in (h). The values of the diagonal elements represent the proportion of correctly predicted classes. Compared to the distraction and the normal states, the fatigue and the workload states were identified with a relatively higher accuracy in all methods. In addition, for the normal state, the models' accuracy did not surpass 65%; the highest accuracy for the normal state was reached with the proposed method (62.8%). On average, the proposed method outperforms others. We conjecture that the low accuracies in the identification of the normal state are because we did not determine the normal state by an independent procedure. Note that we merged the data corresponding to the normal state data from each paradigm.

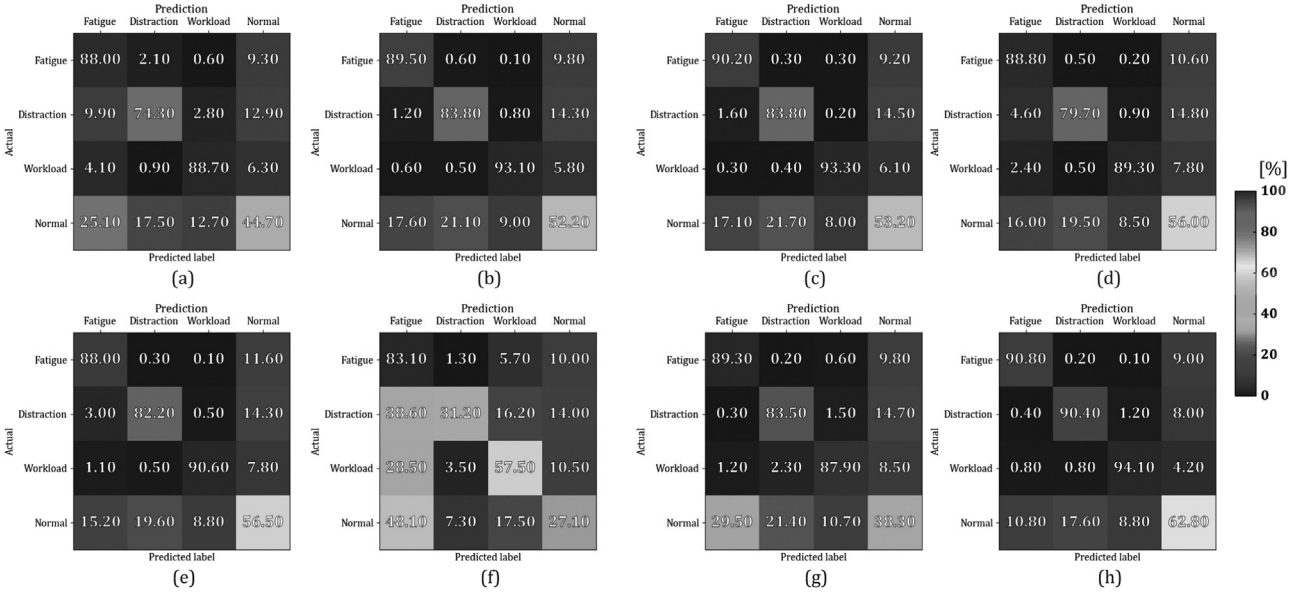

Fig. 10 – Confusion matrices for the accuracy with the baseline method and the proposed method.

5. Conclusion

In this study, we investigated the feasibility of classification of the pilot's various mental states (i.e., distraction, workload, fatigue, and normal) based on multimodal sensors (EEG, ECG, respiration, and EDA) and the proposed MDL method. We collected data corresponding to the four mental states from subjects in a simulated flight environment. And the proposed MDL network successfully enhanced the decoding accuracy compared with several existing methods by fusing the information extracted from the different biosignals. We also showed the changes of physiological measures influenced by the pilot's diversified mental states.

Although our experimental paradigm for data acquisition is similar to that of a previous study [53], future work should consider acquiring the randomized sequence of each mental state. Also, as it is not easy for the human nervous system to switch to different mental states quickly, and consequently, more sophisticated experimental protocols are should be designed. In this study, we focused a subject-dependent classification problem. To increase the convenience, usefulness, and general purpose of the BCI system, comprehensive investigation of the inter and intra-subject variability is needed to obtain more insight, which is required to design transfer-learning frameworks. For this reason, building an optimal transfer model is an important issue and needs to be further investigated. In addition, the optimization of the architecture of the MDL should be investigated. Further, investigation is warranted to test the proposed method in the actual flight environment.

Authors' contribution

Soo-Yeon Han: conceptualization, data curation, investigation, methodology, validation, visualization, writing original draft. No-Sang Kwak: conceptualization, formal analysis, methodology, validation, writing original draft. Taegeun Oh: investigation, validation. Seong-Whan Lee: conceptualization, methodology, supervision, validation, writing original draft.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgments

This work was supported by the 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).

REFERENCES

[1] Hankins TC, Wilson GF. A comparison of heart rate, eye activity, EEG and subjective measures of pilot mental workload during flight. Aviat Space Environ Med 1998;69(4):360–7.
[2] Boksem MA, Tops M. Mental fatigue: costs and benefits. Brain Res Rev 2008;59(1):125–39.
[3] Yen J-R, Hsu C-C, Yang H, Ho H. An investigation of fatigue issues on different flight operations. J Air Transp Manag 2009;15(5):236–40.
[4] Wiegmann DA, Shappell SA. A human error approach to aviation accident analysis: the human factors analysis and classification system. Routledge; 2003.
[5] Sonnleitner A, Treder MS, Simon M, Willmann S, Ewald A, Buchner A, et al. EEG alpha spindles and prolonged brake reaction times during auditory distraction in an on-road driving study. Accid Anal Prev 2014;62:110–8.
[6] Borghini G, Astolfi L, Vecchiato G, Mattia D, Babiloni F. Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. Neurosci Biobehav Rev 2014;44:58–75.
[7] Chaudhuri A, Routray A. Driver fatigue detection through chaotic entropy analysis of cortical sources obtained from scalp EEG signals. IEEE Trans Intell Transp Syst 2019.
[8] Dehais F, Duprès A, Blum S, Drougard N, Scannella S, Roy RN, et al. Monitoring pilot's mental workload using ERPs and spectral power with a six-dry-electrode EEG system in real flight conditions. Sensors 2019;19(6):13–24.
[9] Bashivan P, Rish I, Yeasin M, Codella N. Learning representations from EEG with deep recurrent-convolutional neural networks; 2015;1–15, arXiv:1511.06448.
[10] Jiao Z, Gao X, Wang Y, Li J, Xu H. Deep convolutional neural networks for mental load classification based on EEG data. Pattern Recogn 2018;76:582–95.
[11] Zhang P, Wang X, Zhang W, Chen J. Learning spatial-spectral-temporal EEG features with recurrent 3D convolutional neural networks for cross-task mental workload assessment. IEEE Trans Neural Syst Rehabil Eng 2018;27(1):31–42.
[12] Wu EQ, Peng X, Zhang CZ, Lin J, Sheng RS. Pilots' fatigue status recognition using deep contractive autoencoder network. IEEE Trans Instrum Meas 2019.
[13] Gao Z, Wang X, Yang Y, Mu C, Cai Q, Dang W, et al. EEG-based spatio-temporal convolutional neural network for driver fatigue evaluation. IEEE Trans Neural Netw Learn Syst 2019.
[14] Patel M, Lal SK, Kavanagh D, Rossiter P. Applying neural network analysis on heart rate variability data to assess driver fatigue. Expert Syst Appl 2011;38(6):7235–42.
[15] Hogervorst MA, Brouwer A-M, Van Erp JB. Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental workload. Front Neurosci 2014;8(322):1–14.
[16] Ahn S, Nguyen T, Jang H, Kim JG, Jun SC. Exploring neuro-physiological correlates of drivers' mental fatigue caused by sleep deprivation using simultaneous EEG, ECG, and fNIRS data. Front Human Neurosci 2016;10(219):1–14.
[17] Liu Y, Ayaz H, Shewokis PA. Multisubject ''learning'' for mental workload classification using concurrent EEG, fNIRS, and physiological measures. Front Human Neurosci 2017;11(389):1–14.
[18] Zhang P, Wang X, Chen J, You W. Feature weight driven interactive mutual information modeling for heterogeneous bio-signal fusion to estimate mental workload. Sensors 2017;17(10):1–24.
[19] Lee M-H, Fazli S, Mehnert J, Lee S-W. Subject-dependent classification for robust idle state detection using multi-modal neuroimaging and data-fusion techniques in BCI. Pattern Recogn 2015;48(8):2725–37.
[20] Suk H-I, Lee S-W. A novel Bayesian framework for discriminative feature extraction in brain–computer

interfaces. IEEE Trans Pattern Anal Mach Intell 2012;35(2):286–99.
[21] Ding X, Lee S-W. Changes of functional and effective connectivity in smoking replenishment on deprived heavy smokers: a resting-state FMRI study. PLOS ONE 2013;8(3):1–12.
[22] Zhu X, Suk H-I, Lee S-W, Shen D. Canonical feature selection for joint regression and multi-class identification in Alzheimer's disease diagnosis. Brain Imaging Behav 2016;10(3):818–28.
[23] Chen X, Zhang H, Zhang L, Shen C, Lee S-W, Shen D. Extraction of dynamic functional connectivity from brain grey matter and white matter for MCI classification. Human Brain Mapp 2017;38(10):5019–34.
[24] Stehlin SA, Nguyen XP, Niemz MH. EEG with a reduced number of electrodes: where to detect and how to improve visually, auditory and somatosensory evoked potentials. Biocybern Biomed Eng 2018;38(3):700–7.
[25] Kwak N-S, Lee S-W. Error correction regression framework for enhancing the decoding accuracies of ear-EEG brain–computer interfaces. IEEE Trans Cybern 2019;1–14.
[26] Kam T-E, Suk H-I, Lee S-W. Non-homogeneous spatial filter optimization for electroencephalogram EEG-based motor imagery classification. Neurocomputing 2013;108:58–68.
[27] Kim I-H, Kim J-W, Haufe S, Lee S-W. Detection of braking intention in diverse situations during simulated driving based on EEG feature combination. J Neural Eng 2014;12(1):1–12.
[28] Yeom S-K, Fazli S, Müller K-R, Lee S-W. An efficient ERP-based brain-computer interface using random set presentation and face familiarity. PLOS ONE 2014;9(11):1–13.
[29] Suk H-I, Lee S-W, Shen D, Initiative ADN. Deep sparse multi-task learning for feature selection in Alzheimer's disease diagnosis. Brain Struct Funct 2016;221(5):2569–87.
[30] Kwak N-S, Müller K-R, Lee S-W. A lower limb exoskeleton control system based on steady state visual evoked potentials. J Neural Eng 2015;12(5):1–14.
[31] Górecka J, Walerjan P. Artifacts extraction from EEG data using the infomax approach. Biocybern Biomed Eng 2011;31(4):59–74.
[32] Lee M-H, Kwon O-Y, Kim Y-J, Kim H-K, Lee Y-E, Williamson J, et al. EEG dataset and OpenBMI toolbox for three BCI paradigms: an investigation into BCI illiteracy. GigaScience 2019;8(5):1–16.
[33] Lotfan S, Shahyad S, Khosrowabadi R, Mohammadi A, Hatef B. Support vector machine classification of brain states exposed to social stress test using EEG-based brain network measures. Biocybern Biomed Eng 2019;39(1):199–213.
[34] Lal SK, Craig A, Boord P, Kirkup L, Nguyen H. Development of an algorithm for an EEG-based driver fatigue countermeasure. J Saf Res 2003;34(3):321–8.
[35] Jap BT, Lal S, Fischer P, Bekiaris E. Using EEG spectral components to assess algorithms for detecting fatigue. Expert Syst Appl 2009;36(2):2352–9.
[36] Kar S, Bhagat M, Routray A. EEG signal analysis for the assessment and quantification of driver's fatigue. Transp Res Part F: Traff Psychol Behav 2010;13(5):297–306.
[37] Trejo LJ, Kubitz K, Rosipal R, Kochavi RL, Montgomery LD. EEG-based estimation and classification of mental fatigue. Psychology 2015;6(5):572–89.
[38] Rogado E, Garcia JL, Barea R, Bergasa LM, López E. Driver fatigue detection system. 2008 IEEE International Conference on Robotics and Biomimetics 2009;1105–10.
[39] Jung S-J, Shin H-S, Chung W-Y. Driver fatigue and drowsiness monitoring system with embedded electrocardiogram sensor on steering wheel. IET Intell Transp Syst 2014;8(1):43–50.
[40] Hajinoroozi M, Mao Z, Jung T-P, Lin C-T, Huang Y. EEG-based prediction of driver's cognitive performance by deep convolutional neural network. Signal Process: Image Commun 2016;47:549–55.
[41] Yang G, Lin Y, Bhattacharya P. A driver fatigue recognition model based on information fusion and dynamic Bayesian network. Inform Sci 2010;180(10):1942–54.
[42] Shin H-S, Jung S-J, Kim J-J, Chung W-Y. Real time car driver's condition monitoring system. SENSORS 2010 IEEE 2010;951–4.
[43] Khushaba RN, Kodagoda S, Lal S, Dissanayake G. Driver drowsiness classification using fuzzy wavelet-packet-based feature-extraction algorithm. IEEE Trans Biomed Eng 2010;58(1):121–31.
[44] Rigas G, Goletsis Y, Bougia P, Fotiadis DI. Towards driver's state recognition on real driving conditions. Int J Vehic Technol 2011;2011:1–14.
[45] Lee B-G, Lee B-L, Chung W-Y. Mobile healthcare for automatic driving sleep-onset detection using wavelet-based EEG and respiration signals. Sensors 2014;14(10):17915–36.
[46] Fu R, Wang H, Zhao W. Dynamic driver fatigue detection using hidden Markov model in real driving condition. Expert Syst Appl 2016;63:397–411.
[47] Sahayadhas A, Sundaraj K, Murugappan M, Palaniappan R. A physiological measures-based method for detecting inattention in drivers using machine learning approach. Biocybern Biomed Eng 2015;35(3):198–205.
[48] Malik M. Heart rate variability: standards of measurement, physiological interpretation, and clinical use: Task Force of the European Society of Cardiology and the North American Society for Pacing and Electrophysiology. Ann Noninvas Electrocardiol 1996;1(2):151–81.
[49] Skibniewski FW, Dziuda L, Baran PM, Krej MK, Guzowski S, Piotrowski MA, et al. Preliminary results of the LF/HF ratio as an indicator for estimating difficulty level of flight tasks. Aerosp Med Human Perform 2015;86(6):518–23.
[50] Greco A, Valenza G, Bicchi A, Bianchi M, Scilingo EP. Assessment of muscle fatigue during isometric contraction using autonomic nervous system correlates. Biomed Signal Process Control 2019;51:42–9.
[51] Bulling A, Ward JA, Gellersen H, Troster G. Eye movement analysis for activity recognition using electrooculography. IEEE Trans Pattern Anal Mach Intell 2010;33(4):741–53.
[52] Çakir MP, Vural M, Koç SÖ, Toktas A. Real-time monitoring of cognitive workload of airline pilots in a flight simulator with fNIR optical brain imaging technology. International Conference on Augmented Cognition 2016;147–58.
[53] Choi M, Koo G, Seo M, Kim SW. Wearable device-based system to monitor a driver's stress, fatigue, and drowsiness. IEEE Trans Instrum Meas 2017;67(3):634–45.
[54] Hyvärinen A, Oja E. Independent component analysis: algorithms and applications. Neural Netw 2000;13(4–5):411–30.
[55] Cochran WT, Cooley JW, Favin DL, Helms HD, Kaenel RA, Lang WW, et al. What is the fast Fourier transform? Proc IEEE 1967;55(10):1664–74.
[56] Bulthoff HH, Lee S-W, Poggio T, Wallraven C. Biologically motivated computer vision. Springer-Verlag; 2003.
[57] Hu D, Wang C, Nie F, Li X. Dense multimodal fusion for hierarchically joint representation. Proceedings of the International Conference on Acoustics Speech and Signal Processing (ICASSP) 2019;3941–5.
[58] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation 1997;9(8):1735–80.
[59] Lemm S, Blankertz B, Dickhaus T, Müller K-R. Introduction to machine learning for brain imaging. NeuroImage 2011;56(2):387–99.
[60] Kwak N-S, Müller K-R, Lee S-W. A convolutional neural network for steady state visual evoked potential classification under ambulatory environment. PLOS ONE 2017;12(2):1–20.

[61] Cover TM, Hart PE. Nearest neighbor pattern classification. IEEE Trans Inform Theory 1967;13(1):21–7.
[62] Breiman L. Random forests. Mach Learn 2001;45(1):5–32.
[63] Cox DR. The regression analysis of binary sequences. J R Stat Soc: Ser B Methodol 1958;20(2):215–32.
[64] Defazio A, Bach F, Lacoste-Julien S. SAGA: a fast incremental gradient method with support for non-strongly convex composite objectives. Advances in Neural Information Processing Systems 2014;1646–54.
[65] Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995;20(3):273–97.
[66] Chang C-C, Lin C-J. Libsvm: A library for support vector machines. ACM Trans Intell Syst Technol 2011;2(3):1–27.
[67] Mika S, Ratsch G, Weston J, Schölkopf B, Müller K-R. Fisher discriminant analysis with kernels. Neural Networks for Signal Processing IX: Proceedings of the 1999 IEEE Signal Processing Society Workshop 1999;41–8.