Multimodal Vigilance Estimation Using Deep Learning

Wei Wu, *Member, IEEE*, Wei Sun, Q. M. Jonathan Wu, *Senior Member, IEEE*, Yimin Yang, *Senior Member, IEEE*, Hui Zhang, *Member, IEEE*, Wei-Long Zheng, *Member, IEEE*, and Bao-Liang Lu, *Senior Member, IEEE*

***Abstract*—The phenomenon of increasing accidents caused by reduced vigilance does exist. In the future, the high accuracy of vigilance estimation will play a significant role in public transportation safety. We propose a multimodal regression network that consists of multichannel deep autoencoders with subnetwork neurons (MCDAE$_{sn}$). After we define two thresholds of "0.35" and "0.70" from the percentage of eye closure, the output values are in the continuous range of 0–0.35, 0.36–0.70, and 0.71–1 representing the awake state, the tired state, and the drowsy state, respectively. To verify the efficiency of our strategy, we first applied the proposed approach to a single modality. Then, for the multimodality, since the complementary information between forehead electrooculography and electroencephalography features, we found the performance of the proposed approach using features fusion significantly improved, demonstrating the effectiveness and efficiency of our method.**

***Index Terms*—Deep learning, dimension reduction, electroencephalography (EEG), electrooculography (EOG), multimodal vigilance estimation.**

Manuscript received March 7, 2020; revised June 12, 2020; accepted September 3, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant U1813205, Grant 61971071, Grant 61673266, and Grant 61976135; in part by the Independent Research Project of State Key Laboratory of Advanced Design and Manufacturing for Vehicle Body under Grant 71765003; in part by the Hunan Key Laboratory of Intelligent Robot Technology in Electronic Manufacturing Open Foundation under Grant 2017TP1011 and Grant IRT2018009; in part by the Natural Science and Engineering Research Council (NSERC) of Canada, specifically the NSERC Discovery Grant Program and NSERC CREATE TrustCAV; in part by the National Key Research and Development Program of China under Grant 2018YFB1308200; in part by the Changsha Science and Technology Project under Grant kq1907087; in part by the Hunan Key Project of Research and Development Plan under Grant 2018GK2022, Special Funding for the construction of Innovative Provinces in Hunan under Grant 2020SK3007; in part by the National Key Research and Development Program of China under Grant 2017YFB1002501; in part by the SJTU Trans-med Awards Research under Grant WF540162605; in part by the Fundamental Research Funds for the Central Universities; in part by the 111 Project; and in part by the China Scholarship Council under Grant 201706130071. This article was recommended by Associate Editor H. A. Abbass. *(Corresponding authors: Wei Sun; Hui Zhang.)*

Wei Wu and Wei Sun are with the College of Electrical and Information Engineering, Hunan University, Changsha 410082, China, also with the State Key Laboratory of Advanced Design and Manufacturing for Vehicle Body, Hunan University, Changsha 410082, China, and also with the Hunan Key Laboratory of Intelligent Robot Technology in Electronic Manufacturing, Hunan University, Changsha 410082, China (e-mail: david-sun@126.com).

Q. M. Jonathan Wu is with the Department of Electrical and Computer Engineering, University of Windsor, Windsor, ON N9B 3P4, Canada.

Yimin Yang is with the Computer Science Department, Lakehead University, Thunder Bay, ON P7B 5E1, Canada.

Hui Zhang is with the College of Robot, Hunan University, Changsha 410082, China (e-mail: zhanghuihby@126.com).

Wei-Long Zheng is with the Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114 USA.

Bao-Liang Lu is with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China, and also with the Key Laboratory of Shanghai Education Commission for Intelligent Interaction and Cognitive Engineering, Shanghai Jiao Tong University, Shanghai 200240, China.

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

Digital Object Identifier 10.1109/TCYB.2020.3022647

I. Introduction

DIFFERENT groups of scientists use the term "vigilance" in different ways [1]. In psychiatry, vigilance can be specifically described as attention to a potentially dangerous condition, with hypervigilance that is a symptom of anxiety disorder [2]. In the area of cognitive neuroscience, vigilance refers to the ability to focus on a task within a lasting time [3]. In clinical neurophysiology, the vigilance level is related to refer to the corresponding arousal level on the spectrum of the sleep–wake [4]. In the above discussion, the most common definition is that vigilance represents sustained attention.

The Centers for Disease Control and Prevention (CDC) [5] reported that 4.2% of the 147 076 adult respondents from nearly 20 states or regions reported having had at least one drowsy drive in the past 30 days in 2009–2010. Reduced or complete loss of vigilance has been resulting in an increasing number of traffic accidents around the world [6], [7], meaning it should be taken seriously. Various methods of studying vigilance, including subjective methods [8], behavioral methods [9]–[11], and vehicle-based methods [12], have been proposed to cope with this problem. However, the primary limitations of those methods arise from ignoring the uniqueness of the individual driver and neglecting the personal biases involved, as well as the monotony of the simulated environment under experimental conditions.

Physiological signals or nonvisual features of drivers with healthy physical conditions have fewer false than visual features and can be used to predict drowsiness in a timely manner. In fact, the methods based on physiological signals that represent internal cognitive states have been gradually considered as an efficient means of assessing vigilance. Scientists have found a certain relationship between electrocardiography (ECG) and fatigue, including heart rate (HR) decrease and HR variability (HRV), changes during fatigue [13], [14], and a healthy

(a) (b)

Fig. 1. Simulation experiment environment. (a) LCD screen. (b) SMI-ETGW and the electrode channels.

subject with prolonged fatigue has the reduced respiration rate (RR) value [15]. The study of electromyography (EMG) shows that the frequency spectrum shifts to low frequencies and the amplitude of the EMG signal increases when muscles become fatigued [16], [17]. Electroencephalography (EEG) [18], [19] is generated from the potentials that are recorded from the rhythmic activity of the postsynaptic cortical neuron, which is synchronized by the complex interaction of a large number of cortical cells. Among various physiological indicators, EEG is considered to be the most important and reliable because it directly records the neurophysiological signal of the human brain.

In addition to EEG, electrooculography (EOG) evaluates various eye movements, which can provide valuable warning indications of drowsiness. EOG is another promising measure of assessing vigilance [20]. Unlike the traditional EOG (EOG_t)-based method, a simple but more robust method proposed by Zheng and Lu [21] utilizes the placement of wearable electrodes on the forehead area. The amplitude of forehead EOG (EOG_f) is significantly lower after extraction by a median filter, and suitable electrodes placement reduces the user's discomfort.

Huo *et al.* [22] and Wu *et al.* [23] have proposed the multimodal methods for estimating the level of vigilance and achieving better performance. For example, Huo *et al.* [22] used a fusion strategy that employs feature-level fusion (FLF) to detect fatigue levels, combined with a graph-regularized extreme learning machine (GELM). The average value of the correlation coefficient (COR)/root mean-square error (RMSE) greatly improved, moving to 0.81/0.07 using fusion signals, while the corresponding average COR/RMSE values were 0.73/0.09 for single modality.

We propose using a network of multichannel deep autoencoders with subnetwork neurons ($MCDAE_{sn}$) to obtain the optimal features, employing feature fusion to estimate vigilance. Here, we use four different EOG and EEG datasets from SEED-VIG. Our work contributes to the research on the topic as follows.

1) Compared to the other existing iterative deep-learning (DL)-based networks [24], instead of being randomly acquired, the hidden layers of the proposed autoencoder model are calculated by replacement technologies and include only four steps. Simultaneously, the proposed architecture aims for dimension reduction and signal reconstruction instead of relying on efficient classification applications, as do the other existing multilayer network models.
2) Unlike the other traditional multilayer networks [25], the proposed model consists of many hidden nodes, each of which can be considered as a layer of the network model and has capabilities of feature selection and representation learning. Simultaneously, the input data are randomly divided into five batches, each of which through processes of dimension reduction, subspace feature extraction, and subspace feature combination.
3) To quantify vigilance, the output values are a series of continuous value in the range of 0 to 1 corresponding to the percentage of eye closure (PERCLOS). Blink components, such as impulses from vertical EOG (EOG_v) feature and saccade components from horizontal EOG (EOG_h) feature that can be easily detected by the proposed algorithm, which is consistent with our previous conclusions [21].

II. Related Works

A. Description of the Dataset

Fig. 1(a) shows the data-collection apparatus, wherein the experimental vehicle is an engineless car in which the gas pedal and steering wheel are controlled by software. An LCD screen in front of the vehicle simulates a highway driving environment and is updated in real time. Subjects signed written informed consent before participating, and this research was approved by the local ethics committee. The data were gathered from 23 human subjects, including 11 men and 12 women with an average age of 23. All of the subjects were healthy, with normal hearing, visual acuity that was normal

or corrected to normal, no visible trauma to the head, no use of medication, and no addiction to alcohol or tobacco, and they were all given regular rest according to the timetable. To ensure that the subject was in a drowsy state while driving, the average time for the experiment was approximately 2 h, from 12:30 P.M. to 2:30 P.M., with the human subjects reaching their peak drowsiness at 1:30 P.M. [26]. Before a subject fell asleep during the experiment, no real-time warning feedback occurred. Fig. 1(b) displays the SensoMotoric Instruments eye-tracking glasses wireless (SMI-ETGW) used to calculate training label and all electrode channels used to collect EOG and EEG signals.

B. Previous Related Network Models

As an important algorithm in artificial intelligence (AI), neural network-based algorithms have been widely used for EEG signal processing. Our two extended neural network models are proposed for the vigilance estimation, such as deep autoencoder (DAE) [20] and multilayer autoencoder (MAE) [27]. We used EOG-based single-modal DAE to estimate vigilance and achieve an accuracy of 80%. Meanwhile, Yang *et al.* proposed MAE for image reconstruction and dimension reduction with the Moore–Penrose inverse matrix learning strategy. The features could be compressed by a single-hidden layer with extremely fast processing speed.

Autoencoders can only address single-type samples and can cause beneficial representations of the inputs; however, we argue that a better representation should also depend on the internal relationship between the input pairs. Thus, we proposed MCDAE$_{sn}$ that can handle multiple types of samples. The formula of output is

$$\mathbf{H}^j = S\left(\mathbf{a}_i^j, \mathbf{b}^j, \mathbf{x}\right), \quad i = \{1, 2, \ldots, d\}, \quad j = \{1, 2, \ldots, n\} \quad (1)$$

where j and i represent the jth subnetwork nodes and its ith hidden nodes, respectively.

Each subnetwork node is only connected to its adjacent, which can be considered as an independent system, improving learning efficiency effectively. In addition, the subnetwork neurons as subspace feature extractors significantly increase the generalization performance [28], as long as the generated subspace features could be mixed and combined in the late stages for classification.

III. Methodology

A. Data Preprocessing

1) PERCLOS: In continuous vigilance assessment with a supervised machine learning paradigm, the chief challenge is how to quantitatively mark physiological signals that are collected from the sensors because it is theoretically difficult to accurately obtain the ground truth of the transformed physiological state. The association between eye movements and arousal is not causal. Instead, eye movement acts as a promising indicator of arousal states, which has been widely explored in previous studies. For example, in the neuroscience field, Wang *et al.* [29] proposed that spontaneous eyelid closures can serve as a proxy for vigilance and be jointly analyzed with functional magnetic resonance imaging (fMRI) [30], [31] to determine vigilance fluctuation. Spontaneous eyelid closure also served as a good marker of reduced responsiveness in sleep-deprived persons. In public driving safety fields, the PERCLOS is one of the most widely acknowledged vigilance indicators in [10] and [32]–[34]. The PERCLOS algorithm, introduced by Wierwille [32], has a mean that is the proportion of time for which eyes remain closed in a given unit of time. Grace and Davis [34] at Carnegie Mellon University repeatedly verified this finding, using a high-resolution camera for testing an eye closure over a specific value to judge drowsiness. Fig. 2 illustrates the measuring principle of the PERCLOS based on the eyes closed degree (ECD), and PERCLOS can be calculated as follows:

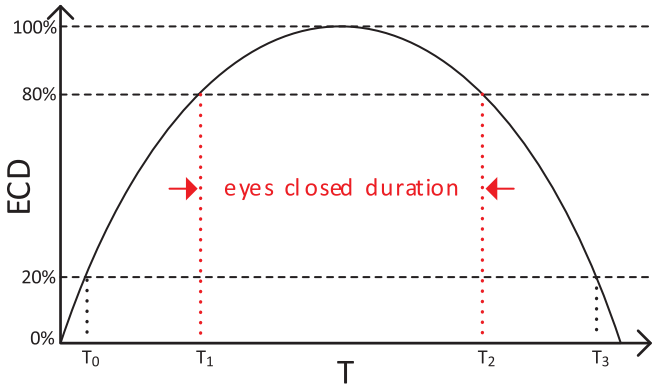


Fig. 2. Details of ECD.

$$V_P = \frac{T_2 - T_1}{T_3 - T_0}, \quad V_P \in [0, 1] \quad (2)$$

where V_P represents the value of PERCLOS; and $(T_2 - T_1)$ and $(T_3 - T_0)$ indicate the eyes closure duration and the duration time from 20% closed state to 20% open state, respectively.

We found that the method mentioned above only pays attention to two states—eyes closed and eyes open—instead of all the important eye movements that provide crucial information for estimating vigilance. Simultaneously, the performance of the method based on traditional facial videos [35] can easily be influenced by environmental factors, especially brightness and occlusion. Thus, we use an automatic continuous vigilance annotation method [36] that employs SMI-ETGW, offering up to 120-Hz high resolution. SMI-ETGW can more fully reflect eye movements, including blinks, saccades, and fixation components, and the PERCLOS training labels calculated by it can be regarded as an accurate and feasible ocular parameter for real-time testing fatigue. This approach can be used for dual tasks in both laboratory and real-world environments. The formula is as follows:

$$V_P = \frac{(T_2 - T_1) + T_s}{T_{\text{duration}}}, \quad V_P \in (0, 1)$$
$$T_{\text{duration}} = (T_2 - T_1) + T_b + T_s + T_f \quad (3)$$

where T_b, T_s, and T_f represent the time of blink, saccade, and fixation state, respectively.

2) Forehead EOG: The eye movements we have analyzed in this study were spontaneous, rather than intentional. The

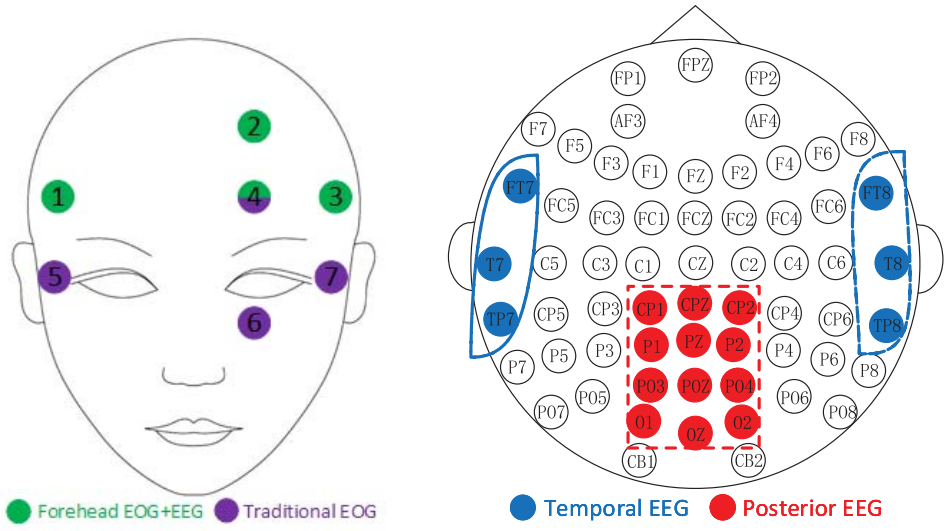


Fig. 3. EOG and EEG data collected from different electrode placements. (a) Forehead electrode positions. (b) Temporal EEG and posterior EEG were recorded from 6-channel electrode positions and 12-channel electrode positions, respectively.

TABLE I
ENCODE THE EOG FEATURE

Components	Sequence	Encode
Negative peak	Single	0
Positive peak	Single	1
Blink candidate	Three successive	010
Saccade candidate	Two successive	01 or 10

TABLE II
DIFFERENT EOG FEATURES

EOG	$EOG_{horizontal}$ by ICA	$EOG_{horizontal}$ by MIN
$EOG_{vertical}$ by ICA	EOG_f-ICA	EOG_h-MIN & EOG_v-ICA
$EOG_{vertical}$ by MIN	EOG_h-ICA & EOG_v-MIN	EOG_f-MIN

TABLE III
EOG FEATURES FORMAT

Features	Dimension	Sample	Data format
Saccade	13	885	
Fixation	10	885	36 × 885
Blink	13	885	

signals were recorded passively over a long time period, with the aim of finding the intrinsic associations between spontaneous eye movements and vigilance states. Since human–machine interfaces have been comprehensively researched, enough methods are available to extract the EOG signals. The traditional EOG signals without noise are obtained through traditional electrode placement around the eyes. In practical applications, many restrictions exist [37], especially in regard to the potential for obstructing one's normal sight or normal operation, intentional control of eye movements by the individual, and artifacts—potential shifts of the body surface caused by eyelid movement and retinal dipole movement. Although both vertical EOG (EOG_v) and horizontal EOG (EOG_h) exist in the traditional EOG signals, it is difficult to extract them directly. Compared with the methods of extracting EOG_v and EOG_h features, we used forehead electrodes placement to obtain the forehead EOG signals in the previous work [38], and we successfully separated the forehead vertical-EOG (EOG_{fv}) and horizontal-EOG (EOG_{fh}) signals in the raw EOG_f signals. Fig. 3(a) shows two types of electrode placements, which are represented by numbers 1–4 (green dots) and numbers 4–7 (purple dots), respectively, and the two techniques share the same electrode number 4. The electrodes of numbers 1, 3, and 4 have the same height. The EOG_{fh} and EOG_{fv} can be collected by the pairs of numbers 1 and 3 electrodes and numbers 2 and 4 electrodes, respectively.

EOG_{fv} and EOG_{fh} features can be obtained by the approaches of fast independent component analysis (FASTICA) [39] and the minus rule [21]. After extracting EOG_{fh} and EOG_{fv}, Mexican hat continuous wavelet transforms (MHWT) [40] are used, which is a peak detection strategy, using a scale of 8 to detect the peaks of the eye movements. If we define the Gaussian function as $\theta(x)$, we can derive the function of MHWT as follows:

$$\theta(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma}$$
$$\psi(x) = \frac{d^2\theta(x)}{dx^2} = \frac{2}{\sqrt{3}}\pi^{-1/4}\left(1 - x^2\right)e^{-\left(x^2/2\right)}. \tag{4}$$

Subsequent to the collection of these data, we encoded the eye movements, such as peaks, blink candidates, and saccade candidates in Table I. For example, a single blink contains three successive peaks: 1) a negative peak; 2) a positive peak; and 3) a negative peak. Thus, we encoded a blink as 010. We have presented the details of these data on the eye movements of the subjects in Tables II and III.

3) EEG: Eye movement alone is not enough to develop a robust vigilance estimation model. Eye movement can be

TABLE IV
DIFFERENT EEG FEATURES

EEG	Moving average (MA)	Linear dynamic system (LDS)
DE	DE-MA	DE-LDS
PSD	PSD-MA	PSD-LDS

intentionally controlled by subjects, which causes degraded performance in prediction. We therefore combined different modalities in our study by including brain signals. EEG is an electrical activity that is noninvasively recorded from the scalp, which cannot be intentionally controlled by subjects. Moreover, the changes in brain activities contribute to early warnings of reduced vigilance. Herein lies the motivation for our development of a multimodal machine learning algorithm combining both EOG and EEG to estimate vigilance. According to the traditional methods [41], [42], ocular artifacts (OAs), including eye movements and blinks, are always considered the most dominant type of contamination—especially for the signals that are collected from the frontal head, which produces higher—magnitude signals, allowing them to travel throughout the scalp, distorting and masking EEG signals. Compared with conventional approaches to removing EOG_f from forehead recording, we considered that EOG_f offers crucial information for the estimation of vigilance. We then used the FASTICA method to separately extract EOG_f and EEG_f signals from forehead electrodes recording.

The input matrix $u = [\text{No. 4; No. 2; } -\text{No. 1; No. 3}]$ is converted by the raw signals recorded by the front head channels (No. 1, No. 2, No. 3, and No. 4). Here, the four columns of the matrix u represent the data collected from channels No. 4, No. 2, No. 1, and No. 3, respectively. We obtained the unmixed matrix v after the decomposition. Then, the sum of independent components w was decomposed by the multichannel data. Thus, the pure EEG_f signals $\widetilde{u}$ can be extracted as follows:

$$\begin{aligned} w &= v * u \\ \widetilde{u} &= v^{-1} * \widetilde{w} \end{aligned} \tag{5}$$

where $\widetilde{w}$ is the matrix for activation waveforms w, of which the rows consist of EOG components set to zero; and v^{-1} represents the inverse matrix of v.

The temporal and posterior electrode channel is a well-established area, which is used for detecting changes in the drowsiness state [43]. Temporal-EEG (EEG_t) and posterior-EEG (EEG_p) were recorded by 6-channel electrode placements and 12-channel electrode placements, respectively. Fig. 3(b) displays the two electrode placements. For the process of denoising EEG data, we used a bandpass filter with 1–75 Hz to remove noise and artifacts manually and downsampled with a sampling frequency of 200 Hz to improve computational efficiency. We then used short-term Fourier transforms with an 8-s nonoverlapping Hanning window to calculate the differential entropy (DE) feature, which is considered one of the most efficient EEG features for estimating vigilance. We define $[x, x \sim G(\mu, \sigma^2)]$ representing a random time series that follows the Gaussian distribution, and $f(x \mid \mu, \sigma^2)$ representing its probability density function. The formula can be expressed as follows:

$$f\left(x \mid \mu, \sigma^2\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}((x-\mu)/\sigma)^2}. \tag{6}$$

Then, we obtained the DE features $h(x \mid \mu, \sigma^2)$ as follows:

$$\begin{aligned} h\left(x \mid \mu, \sigma^2\right) &= -\int_x f\left(x \mid \mu, \sigma^2\right) log f\left(x \mid \mu, \sigma^2\right) dx \\ &= -\int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}((x-\mu)/\sigma)^2} \\ &\quad \times \log \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}((x-\mu)/\sigma)^2} dx \\ &= \frac{1}{2}\log\left(2\pi e\sigma^2\right). \end{aligned} \tag{7}$$

In addition, we extracted DE, and power spectral density (PSD) features are extracted by the total frequency bands with 2-Hz frequency resolution (2 Hz) between 1 and 50 Hz and five frequency bands (5Bands), respectively. After using the moving average (MA) and the linear dynamic system (LDS) filtering, we listed the EEG features used, and their formats are in Tables IV and V, respectively. For example, in Table IV, PSD-MA represented PSD features extracted by raw data prior to the use of the MA separation strategy.

TABLE V
EEG FEATURES FORMAT

Features	Dimension	Sample	Data format
EEG_{f2}	100	885	100 × 885
EEG_{f5}	20	885	20 × 885
EEG_{t2}	150	885	150 × 885
EEG_{t5}	30	885	30 × 885
EEG_{p2}	275	885	275 × 885
EEG_{p5}	55	885	55 × 885

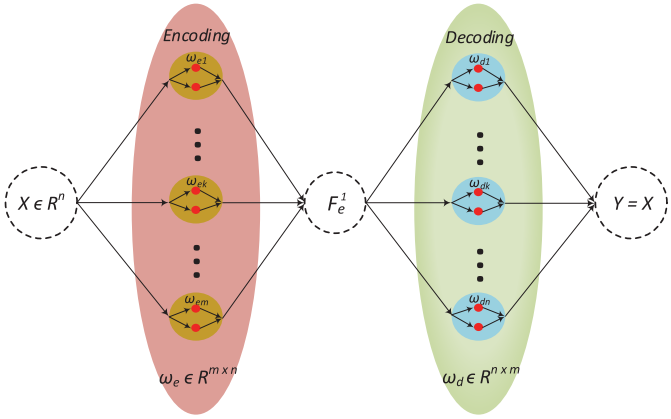


Fig. 4. Double-layer network in the proposed framework. The m-dimensional features F_e are obtained by mapping n-dimensional input data X.

B. Network Model

According to the advantage of physiological signals, we know that the EOG_f has two properties: 1) a high ratio of signal to noise and 2) ease of setting up. Meanwhile, EEG can fully record neurophysiological signals about vigilance.

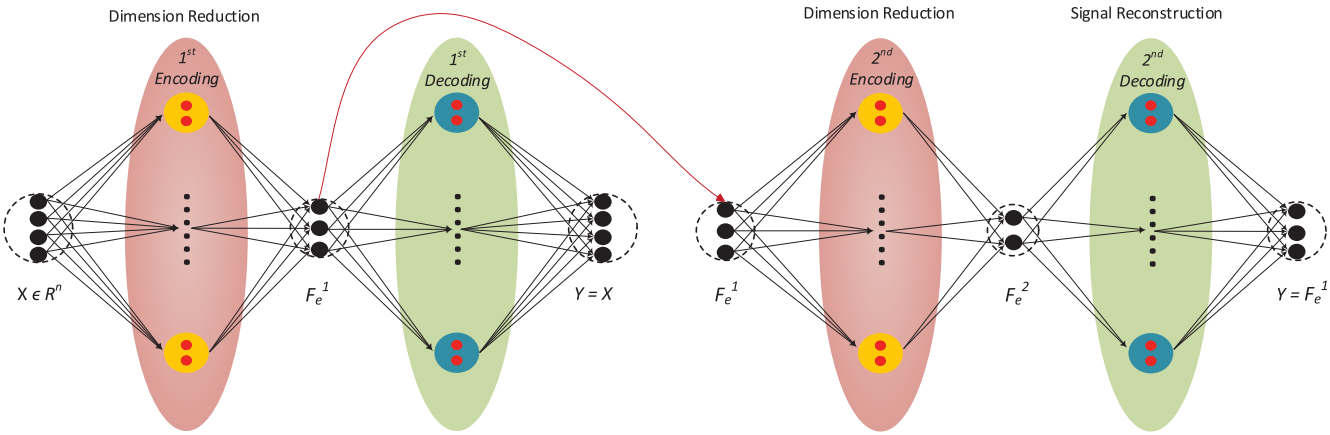


Fig. 5. Multilayer structure for dimension reduction and signal reconstruction.

EOG and EEG signals have complementary characteristics. We proposed a network MCDAE$_{sn}$ to test the accuracy of vigilance. The related notations are defined in Table VI.

1) Subspace Feature Dimension Reduction and Signal Reconstruction: A double-layer network in the proposed framework is described in Fig. 4. The low-dimensional features F_e are obtained by mapping high-dimensional input data X. The weights of the decoding layer are used as the input data of the next encoding layer and all parameters used by the encoding layer are updated (see Fig. 5). The raw input data were divided into five batches, each through the processes of dimension reduction and signal reconstruction. The details can be described as follows.

Step 1: Given $\mathbb{N}$ arbitrary distinct training samples from a continuous system $\{(\mathbf{X}_n, \mathbf{Y}_n)_1^{\mathbb{N}}, \mathbf{X} \in \mathbb{R}^{n_1}, \mathbf{Y} \in (0, 1]\}$. Output data were similar to the input data, which were reconstructed by an autoencoder. Here $\mathbf{X} = \mathbf{Y}$, we then used the activation function $\phi_1(x) = \sin(x)$ in the encoding layer. The initial general input weights $(\mathbf{W}_{ej}, \mathbf{W}_{ej} \in \mathbb{R}^{n_1 \times m})$ and biases $(b_{ej}, b_{ej} \in \mathbb{R})$ were orthogonal random generation, as follows:

$$\begin{aligned} \mathbf{F}_{e1} &= \sum_{i=1}^{d} \mathbf{F}_{e1}^{i} = \phi_1\left(\mathbf{W}_{e1}^{i}, \mathbf{X}, b_{e1}^{i}\right) \\ &= \sum_{i=1}^{d} \sin\left(\mathbf{W}_{e1}^{i} \cdot \mathbf{X} + b_{e1}^{i}\right) \\ \mathbf{W}_{ej}' \cdot \mathbf{W}_{ej} &= \mathbf{I}, \ b_{ej}' \cdot b_{ej} = 1 \end{aligned} \tag{8}$$

where $\mathbf{F}_{e1}$ is current feature data.

Step 2: The inverse of the activation function is $\phi_1^{-1}(x) = \arcsin(x)$. The optimal parameters for the jth decoding layer $\{(\mathbf{W}_{dj}, b_{dj}), \mathbf{W}_{dj} \in \mathbb{R}^{n_1 \times m}, b_{dj} \in \mathbb{R}\}$ are obtained by

$$\begin{aligned} \mathbf{W}_{d1} &= \phi_1^{-1}(\mathbf{Y}) \cdot \mathbf{F}_{e1}^{\dagger} = \arcsin(\mathbf{Y}) \cdot \mathbf{F}_{e1}^{\dagger} \\ b_{d1} &= \mathbf{RMSE}(\mathbf{W}_{d1} \cdot \mathbf{F}_{e1} - \arcsin(\mathbf{Y})) \end{aligned} \tag{9}$$

where RMSE represents the root-mean-squared error that is the standard deviation (SD) of the residuals.

According to the ridge regression (RR) technique, the diagonal elements of the matrix $\mathbf{F}'\mathbf{F}$ or $\mathbf{F}\mathbf{F}'$ should contain the shrinkage ridge parameter $(K > 0)$ in a multiple regression analysis. The inverse function of Moore–Penrose $\mathbf{F}^{\dagger}$ can be expressed as follows:

$$\begin{cases} \mathbf{F}^{\dagger} = \left(\mathbf{F}'\mathbf{F}\right)^{-1}\mathbf{F}'\mathbf{F}^{-1} = \mathbf{F}'\left(K/I + \mathbf{F}\mathbf{F}'\right)^{-1} \\ \quad \left(\text{if } \left(\mathbf{F}'\mathbf{F}\right) \text{ is nonsingular}\right) \\ \mathbf{F}^{\dagger} = \mathbf{F}'\left(\mathbf{F}'\mathbf{F}\right)^{-1} = \left(K/I + \mathbf{F}\mathbf{F}'\right)^{-1}\mathbf{F}' \\ \quad \left(\text{if } \left(\mathbf{F}'\mathbf{F}\right) \text{ is singular}\right). \end{cases} \tag{10}$$

Step 3: Set $j = j + 1$, update $\mathbf{W}_{ej}$, b_{ej}, and $\mathbf{F}_{ej}$ by

$$\begin{aligned} \mathbf{W}_{ej} &= \mathbf{W}_{dj}' \\ b_{ej} &= b_{dj} \\ \mathbf{F}_{ej} &= \phi_1\left(\mathbf{W}_{ej} \cdot \mathbf{X} + b_{ej}\right). \end{aligned} \tag{11}$$

Step 4: We repeat steps 2 and 3 $(n-1)$ times, obtaining the parameters of $(\mathbf{W}_e, b_e)$, $(\mathbf{W}_d, b_d)$, and the feature $\mathbf{F}_e$.

2) Subspace Feature Extraction: The initial feature of the jth (the initial index $j = 1$) subnetwork neurons in a subspace feature extraction layer is obtained from step 4. We found that the initial feature is $\mathbf{F}_{2j} = \mathbf{F}_e$.

Step 5: Given the $\phi_2(x) = 1/(1 + \exp^{-x})$ and $\phi_3(x) = 1/(1 + \exp^{-x})$ activation functions of the entrance layer and

TABLE VI
SYMBOLS USED FOR THE PROPOSED METHOD

Symbol	Property
k	$k = e$, $k = d$, $k = 2$, and $k = 3$ represent encoding layer, decoding layer, subspace feature extraction layer and subspace feature combination layer, respectively.
$(\mathbf{W}_{ki}^{j}, b_{ki}^{j})$	is the ith hidden neuron's feature of the jth encoding layer $(k = e)$ or decoding layer $(k = d)$; is the jth subnetwork neuron $(k = 2$ or $k = 3)$ in kth layer. $(\mathbf{W}_{kj}^{i}, \mathbf{W}_{ki}^{j} \in \mathbb{R}^{D \times m_k})$ and $(b_{ki}^{j}, b_{kj}^{i} \in \mathbb{R})$ represent its weight and bias, respectively.
$\mathbf{F}_{ki}^{j}$	feature data generated by the i^{th} hidden neuron of the jth encoding layer $(k = e)$ or decoding layer $(k = d)$; by the jth subnetwork neuron $(k = 2$ or $k = 3)$ in kth layer.
$\mathbf{F}^{\dagger}$	is the Moore-Penrose inverse of the matrix $\mathbf{F}$.
m_k	m_k represents input data dimension of kth layer.
$\phi_k(x)$	$\phi_e(x) = \phi_d(x) = sine$, $\phi_2(x) = sigmoid$, and $\phi_3(x) = sigmoid$ represent activation functions of the dimension reduction layer, subspace feature extraction layer and subspace feature combination layer, respectively.

exit layer, respectively, we obtained the features of the jth subnetwork neurons $\{(\mathbf{W}_{3j}, b_{3j}), \mathbf{W}_3 \in \mathbb{R}^{n_3 \times m}, b_3 \in \mathbb{R}\}$ by

$$\begin{aligned}\mathbf{W}_{3j} &= \phi_3^{-1}(L(\mathbf{Y})) \cdot \mathbf{F}_{2j}^{\dagger} \\ &= -(\log(1/L(\mathbf{Y}) - 1))\Big/\Big(1 + \exp^{-(\mathbf{W}_{2j}, \mathbf{X}, b_{2j})}\Big)^{\dagger} \\ b_{3j} &= \mathbf{RMSE}\Big(\mathbf{W}_{2j} \cdot \Big(1 + \exp^{-(\mathbf{W}_{2j}, \mathbf{X}, b_{2j})}\Big) + \log(1/L(\mathbf{Y}) - 1)\Big).\end{aligned} \tag{12}$$

Step 6: We updated $\mathbf{e}_j$, $\mathbf{W}_{2j}$, and b_{2j} as follows:

$$\begin{aligned}\mathbf{e}_j &= \mathbf{Y} - L^{-1} \cdot \phi_3\Big(1 + \exp^{-(\mathbf{W}_{2j}, \mathbf{X}, b_{2j})}\Big) \\ \mathbf{W}_{2j} &= \phi_3^{-1}\big(L(\mathbf{P}_{j-1} + \mathbf{F}_{2j})\big) \cdot \mathbf{X}^{-1} \\ b_{2j} &= \mathbf{RMSE}\big(\mathbf{W}_{3j} \cdot \mathbf{X} - \mathbf{P}_{j-1}\big)\end{aligned} \tag{13}$$

where $\mathbf{e}_j$ feedback the data $\{\mathbf{P}_j = \phi_3^{-1}(L(\mathbf{e}_j)) \cdot (\mathbf{W}_{2j})^{-1}, \mathbf{P}_0 = 0\}$. L and L^{-1} represent the normalized function and its reverse function, respectively.

Step 7: For set $j = j + 1$, we can determine the jth subspace features $(\mathbf{W}_{2j}, b_{2j})$ and the $(j+1)$th subspace features $(\mathbf{W}_{2(j+1)}, b_{2(j+1)})$ to be

$$\begin{aligned}\mathbf{F}_{2j} &= \phi_2\big(\mathbf{X}, \mathbf{W}_{2j}, b_{2j}\big) \\ \mathbf{F}_{2(j+1)} &= \phi_2\big(\mathbf{X}, \mathbf{W}_{2(j+1)}, b_{2(j+1)}\big).\end{aligned} \tag{14}$$

Step 8: We repeated steps 5–7 $(n-1)$ times to obtain the subspace features $\{\mathbf{F}_{21}, \ldots, \mathbf{F}_{2n}\}$.

3) Subspace Feature Fusion: Dong *et al.* [44] demonstrated that if the feature contains correct information, early fusion can be considered as a robust strategy over late fusion by an uncomplicated union of different features into one super vector. We considered two pooling to reduce estimation variance error and bias: average pooling [45] and max pooling, especially max pooling, which is employed in many currently popular models of convolutional neural networks (CNNs), including GoogLeNet [46], VGGNet [47], and AlexNet [48]. Max pooling is also widely used to reduce dimension and feature combination in all types of physiological signals [23], [49], [50]. For the multimodality approach, two different types of input data of EOG $\{\mathbf{F}_{21}^{\mathrm{EOG}}, \ldots, \mathbf{F}_{2l}^{\mathrm{EOG}}, \mathbf{F}_{2}^{\mathrm{EOG}} \in \mathbb{R}^{n_2 \times m}\}$ and EEG $\{\mathbf{F}_{21}^{\mathrm{EEG}}, \ldots, \mathbf{F}_{2l}^{\mathrm{EEG}}, \ \mathbf{F}_{2}^{\mathrm{EEG}} \in \mathbb{R}^{n_2 \times m}\}$ were obtained from the subspace feature extraction layer. The feature vectors of EEG and eye movements are directly concatenated into a larger feature vector as the inputs. Then, we found feature fusion to be

$$\hat{\mathbf{F}} = g\Big(\mathbf{F}^{\mathrm{EOG}}, \mathbf{F}^{\mathrm{EEG}}\Big)\begin{cases}\mathrm{mean}_{n_2 \times m}\big(\mathbf{F}^{\mathrm{EOG}}, \mathbf{F}^{\mathrm{EEG}}\big) \\ \mathrm{max}_{n_2 \times m}\big(\mathbf{F}^{\mathrm{EOG}}, \mathbf{F}^{\mathrm{EEG}}\big)\end{cases} \tag{15}$$

where g is a combination operator.

According to our previous studies [51], [52], given distinct $\mathbb{N}$ samples $\{(X_t, Y_t)_{t=1}^{t=N}, X \in \mathbb{R}^n, Y \in \mathbb{R}^m\}$, if the following conditions are met:

$$\begin{aligned}\mathbf{W}_2 &= \phi_2^{-1}(L(e_{m-1})) \cdot X'\big(K/I + XX'\big)^{-1} \\ &= -\log(1/L(e_{m-1}) - 1) \cdot X'\big(K/I + XX'\big)^{-1} \\ b_2 &= \sum\Big(\mathbf{W}_2 \cdot X - \phi_2^{-1}(L(e_{n-1}))\Big)\Big/N \\ &= \sum(\mathbf{W}_2 X + \log(1/L(e_{n-1}) - 1))\big/N\end{aligned}$$

$$\mathbf{W}_3 = \Big(e_{n-1}, L^{-1}(\mathbf{F}_2)\Big)\Big/\Big\|L^{-1}(\mathbf{F}_2)\Big\| \tag{16}$$

the equation $\lim_{j\to\infty} ||Y - (L^{-1}(\phi_2(\mathbf{W}_{21} \cdot X + b_{21})) \cdot \mathbf{W}_{31} + \cdots + L^{-1}(\phi_2(\mathbf{W}_{2j} \cdot X + b_{2j})) \cdot \mathbf{W}_{3j})|| \equiv 0$ holds. Both the input and output weights of the proposed method are shown to have the smallest norm among all the least-squares methods.

4) Feature Combination: Since the data have been divided into five batches, we obtained the feature of the 1st batch $\{\mathbf{F}_3^1\}$ by performing steps 1–8 one time.

Step 9: We repeated steps 1–8 $(5-1)$ times to obtain the entire subspace feature $\{\mathbf{F}_3^1 + \mathbf{F}_3^2 + \mathbf{F}_3^3 + \mathbf{F}_3^4 + \mathbf{F}_3^5\}$.

C. Regression for Vigilance Estimation

Yang and Wu [28] indicated that mixed neurons play a vital role in the coding and functioning of our brains. By recasting subspace features into the mapping space, relevant brain signals can be extracted by these features while generating complex and stable behaviors. This process, from dimension reduction and signal reconstruction to feature fusion, as illustrated in Fig. 6, shows the learning structures and dimensions related to the above-mentioned biological evidence. We used such a model to process signals recognition. The entire data of one experiment were separated into five sessions for evaluation, and after fusing all sessions' features, we used five-fold cross-validation to evaluate the performance. The value of the continuous output data y in the range of 0–0.35, 0.36–0.70, and 0.71–1 indicates the awake state, the fatigue state, and the drowsy state, respectively.

The mean root-mean-square error (RMSE_m) and the mean correlation coefficient (COR_m) are used to quantitatively assess the extent of vigilance like quantitative testing of alcohol in the blood. RMSE_m and COR_m usually reflect the squared error and linear relationship between the observed and predicted values, respectively. The range of the COR value is $[-1, 1]$, where -1, 0, and 1 represent the most disagreement, lack of linear relationship, and the most agreement, respectively. The formulas are

$$\begin{aligned}\mathrm{RMSE}(x, y) &= \sqrt{\sum_{t=1}^{n}(x_t - y_t)^2\Big/n} \\ \mathrm{COR}(x, y) &= \frac{\sum_{t=1}^{n}(x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum_{t=1}^{n}(x_t - \bar{x})^2 \sum_{t=1}^{n}(y_t - \bar{y})^2}}\end{aligned} \tag{17}$$

where $x = (x_1, x_2, \ldots, x_n)^T$ and $y = (y_1, y_2, \ldots, y_n)^T$ represent the observed values and the predicted values, respectively, while $\bar{x}$ and $\bar{y}$ represent the average value of x and y, respectively. In short, the lower the RMSE value, and the higher the COR value, the higher the accuracy of the predicted regression.

Analysis of variance (ANOVA) [53] is not only used to study the statistical models and their associated estimation procedures between groups but also within a group. We used ANOVA to assess the statistical significance of the final experimental results. According to Fisher's F statistic [54], the observed F-value can be calculated with the original data; the empirical frequency distribution of a new F-value—that is, F^*-value—can be obtained through the labels permuted randomly, which belong to a particular group. Thus, the P-value

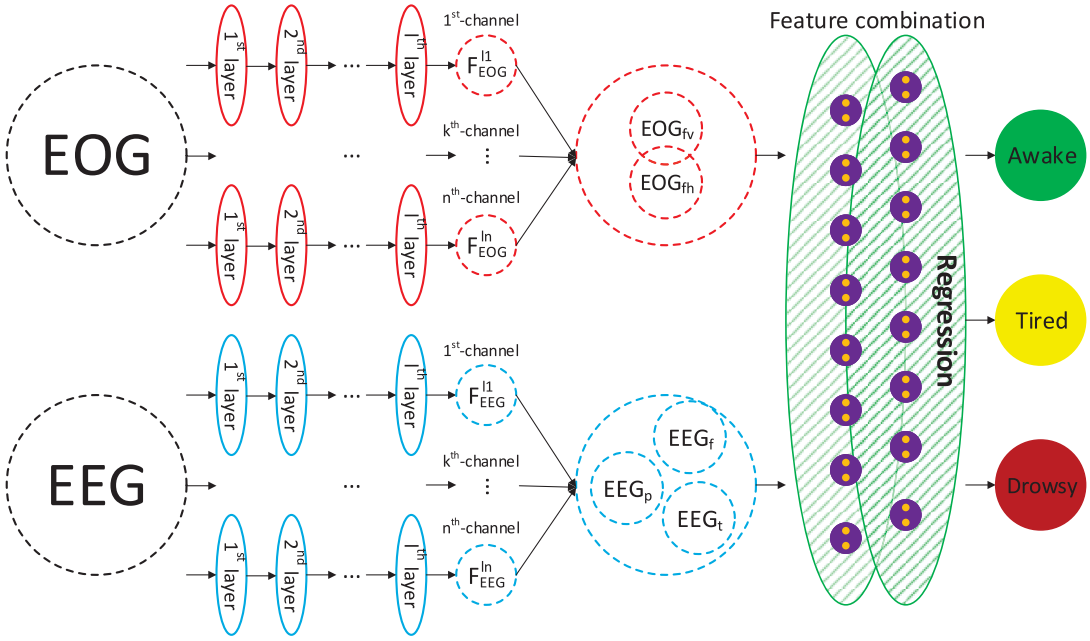


Fig. 6. Proposed method with n-channel autoencoder network and each channel comprising an l-layer structure.

TABLE VII
PARAMETER SETTINGS FOR THE PROPOSED METHOD

Methods	Parameters
ELM	Grid search in $2^{[-10,\cdots,10]}$; 1000 hidden neurons are used.
B-ELM	Grid search in $2^{[-10,\cdots,10]}$; 1000 hidden neurons are used.
SVR	Grid search in $2^{[-10,\cdots,10]}$; Linear Kernel.
CCRF	Regularization hyper-parameters: $\alpha_k = 10^{[0,1,2]}$; Vertex features $K_1 = \{10,20,30\}$; $K_2 = 1$; The sequence length $n = 7$; If i^{th} and j^{th} neurons are adjacent, $S^{(K)} = 1$; otherwise, $S^{(K)} = 0$.
CCNF	Regularization hyper-parameters: $\beta_k = 10^{[-3,-2,-1,0]}$; Vertex features $K_1 = \{10,20,30\}$; $K_2 = 1$; The sequence length $n = 7$; If i^{th} and j^{th} neurons are adjacent, $S^{(K)} = 1$; otherwise, $S^{(K)} = 0$.
DNN_{SN}	$C_1 = 2^{[-10,\cdots,10]}$; $C_2 = 2^{[-10,10]}$; Three subnetwork neurons are used, each of which contains 500 hidden neurons.
$Ours_{ap}$	For single modality: $C_1 = 2^{[-10,\cdots,10]}$ and $C_2 = 2^{[-10,\cdots,10]}$; For multi-modality: $C_1 = 2^{[-10,\cdots,0]}$ and $C_2 = 2^{[-10,\cdots,0]}$; Three subnetwork neurons are used, each of which contains 400 hidden neurons.
$Ours_{mp}$	For single modality: $C_1 = 2^{[-10,\cdots,10]}$ and $C_2 = 2^{[-10,\cdots,10]}$; For multi-modality: $C_1 = 2^{[-10,\cdots,0]}$ and $C_2 = 2^{[-10,\cdots,0]}$; Three subnetwork neurons are used, each of which contains 400 hidden neurons.

from the F statistic based upon F^* is the probability of the true null hypothesis (H_0), which is calculated as the proportion of the F^* that is greater than or equal to F, as follows:

$$P = \frac{\text{Numbers}(F^* \geq F_{\text{observed}})}{\text{Total Numbers}(F^*)}. \tag{18}$$

IV. EXPERIMENTAL VERIFICATION

We tested all of the algorithms outlined in this section with MATLAB 2019a with 64-GB memory. The valuation of parameters can be tuned in every step of the experiment, and Table VII shows the details.

A. Single Modality

1) Using Forehead EOG: We compared the regression models that are commonly utilized with EOG features for vigilance estimation: ICA_f, $MINUS_f$, ICA_{fv}-MIN_{fh}, ELM [55], bidirectional-ELM (B-ELM) [56], DNN_{SN} [23], and the proposed method. The three different features of EOG_{fh}-ICA, EOG_{fv}-ICA, EOG_{fh}-MINUS, and EOG_{fv}-MINUS were extracted by the MINUS and ICA separation strategies. We then obtained three types of EOG features, including ICA_f, ICA_{fv}-MIN_{fh}, and $MINUS_f$.

The performance of these regression models on three types of EOG_f features is shown in Fig. 7, including the mean RMSE/COR ($RMSE_m$/COR_m) and $RMSE_\sigma$/COR_σ. Here, σ represents the SD. The ICA and MINUS methods have been shown to have the advantage of regressing high-dimensional features using the big training dataset. The mean RMSE/COR of the ICA_f, ICA_{fv}-MIN_{fh}, and $MINUS_f$ is 0.16/0.48, 0.12/0.78, and 0.13/0.72, respectively. The blink and saccade components can be easily detected by the ICA_{fv}-MIN_{fh} separation method from the EOG_f signal, which shows a better performance, and we use its performance as the benchmark. The ELM model is frequently used in regression and has an effective and trustworthy performance. ELM is inherently a single-layer feedforward neural network, meaning it can recognize multiple EOG features. The $RMSE_m$/COR_m of the ELM using EOG_f features is improved to 0.13/0.67, 0.13/0.72, and 0.13/0.73, respectively. DNN_{SN} is another strong learning method that improves the overall performance of a series of regressors. We observe that the $RMSE_m$/COR_m is greatly improved to 0.12/0.72, 0.11/0.79, and 0.11/0.78, respectively.

In addition, the performance of our single-modality algorithm with EOG_f notably improved to 0.11/0.79 ($p < 0.01/p < 0.01$, ANOVA), 0.10/0.83 ($p < 0.01/p < 0.01$, ANOVA), and 0.10/0.80 ($p < 0.01/p < 0.01$, ANOVA),

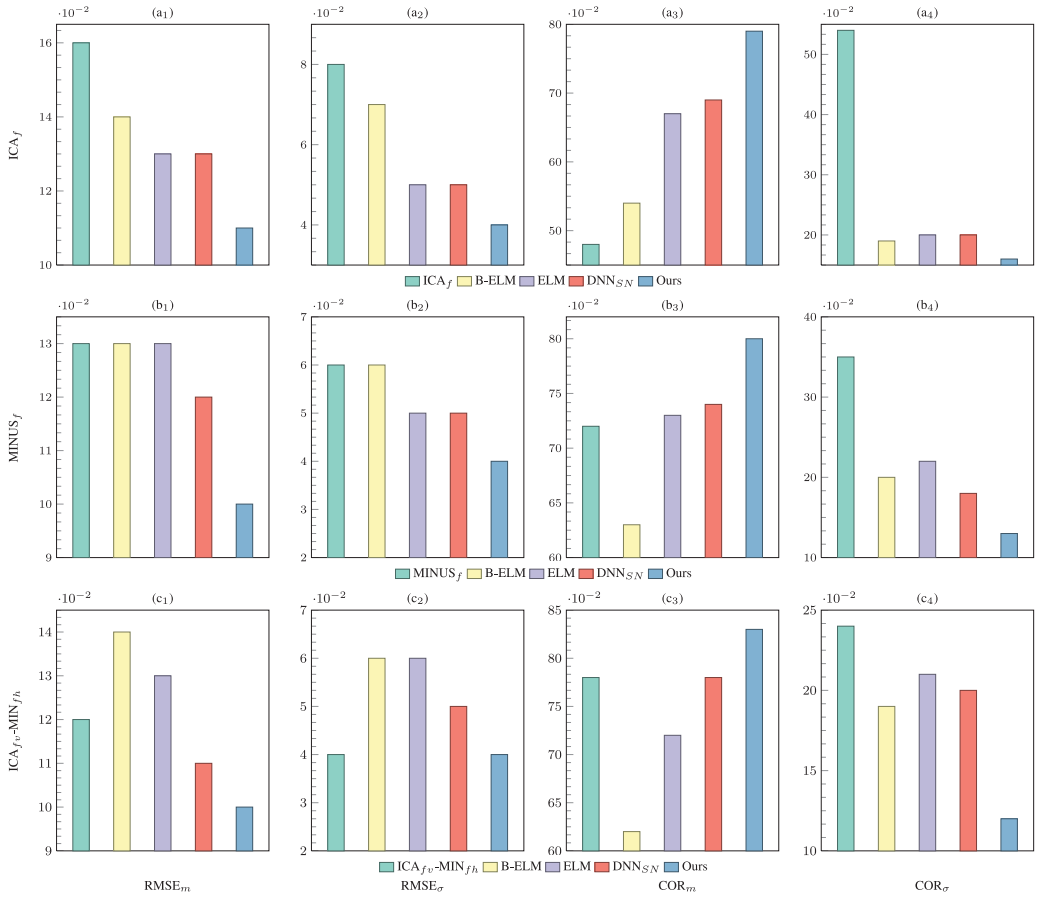


Fig. 7. Performance using different EOG features. σ represents its STD.

TABLE VIII
EXPERIMENTAL RESULTS FOR SINGLE MODALITY USING FOREHEAD EOG. THE BEST RESULTS ARE BOLDED

	$\overline{RMSE_m}$	$\overline{RMSE_\sigma}$	$\overline{COR_m}$	$\overline{COR_\sigma}$
ICA$_f$	0.16	0.08	0.48	0.54
MINUS$_f$	0.13	0.06	0.63	0.20
B-ELM	0.14	0.06	0.60	0.19
ELM	0.13	0.05	0.71	0.21
DNN$_{SN}$	0.12	0.05	0.74	0.19
Ours	**0.10**	**0.04**	**0.81**	**0.14**

TABLE IX
EXPERIMENTAL RESULTS FOR SINGLE MODALITY USING EEG. THE BEST RESULTS ARE BOLDED. (a) EEG WITH 2-HZ FREQUENCY RESOLUTION. (b) EEG WITH FIVE FREQUENCY BANDS

(a)

	$\overline{RMSE_m}$	$\overline{RMSE_\sigma}$	$\overline{COR_m}$	$\overline{COR_\sigma}$
ICA	0.14	0.04	0.67	0.23
B-ELM	0.15	0.05	0.62	0.19
ELM	0.14	0.05	0.65	0.23
DNN$_{SN}$	0.13	0.04	0.68	0.18
Ours	**0.11**	**0.03**	**0.76**	**0.17**

(b)

	$\overline{RMSE_m}$	$\overline{RMSE_\sigma}$	$\overline{COR_m}$	$\overline{COR_\sigma}$
ICA	0.15	0.05	0.63	0.21
B-ELM	0.16	0.05	0.57	0.18
ELM	0.14	0.05	0.66	0.27
DNN$_{SN}$	0.14	0.05	0.70	0.19
Ours	**0.11**	**0.04**	**0.77**	**0.16**

respectively. The mean performances of all compared single-modal methods using EOG are listed in Table VIII, and the proposed method for obtaining the best mean performance of RMSE$_m$/COR$_m$ is 0.10/0.81, which far outperformed other single methods. Our strategy should be seen as a good technique for detecting blink, glances, and fixation components of vigilance.

2) Using EEG: The performance of these regressors using six EEG signals—(EEG$_{f2}$, EEG$_{f5}$, EEG$_{t2}$, EEG$_{t5}$, EEG$_{p2}$, and EEG$_{p5}$)—is shown in Figs. 8 and 9. The mean performances of all compared single-modal methods using EEG are listed in Table IX. Here, f, t, and p represent forehead, temporal, and posterior, respectively. For example, EEG$_{f5}$ represents EEG signals gathered from the forehead site of the brain, which are separated from the five frequency bands method. After using two approaches—MA and LDS filtering—four different features of each EEG signal were extracted: DE-MA, DE-LDS, PSD-MA, and PSD-LDS. We found that the DE feature has reliably recognized EEG patterns between low and high-frequency energy due to the comparison regression models, which include ICA, ELM, B-ELM, DNN$_{\text{SN}}$,

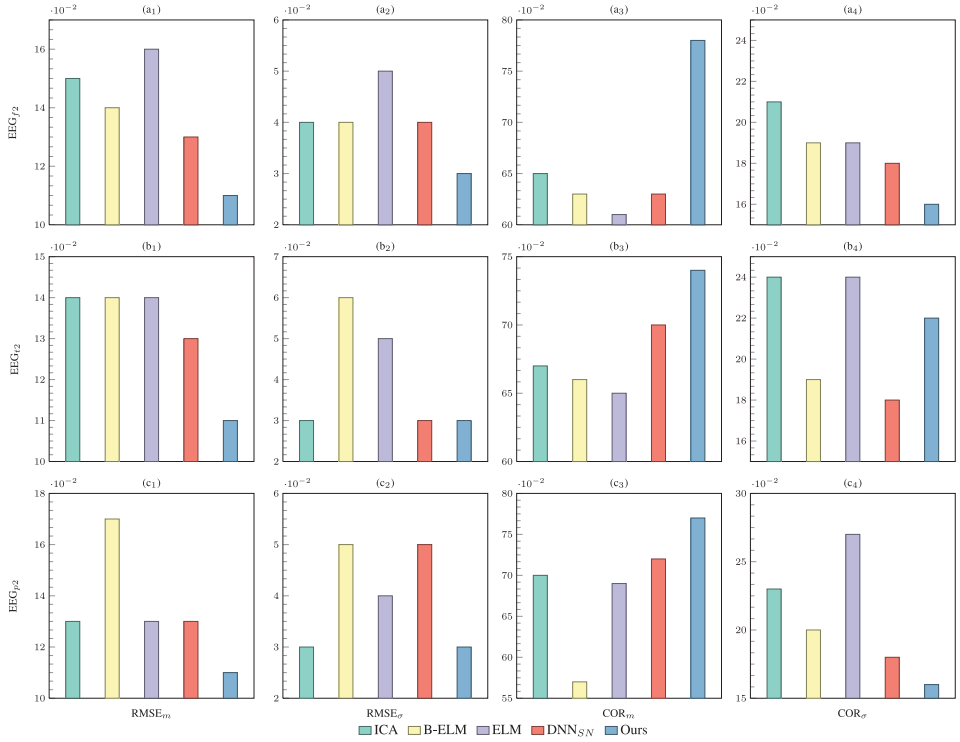


Fig. 8. Performance using three sites EEG with a 2-Hz frequency resolution. (a_1, a_2, a_3, a_4), (b_1, b_2, b_3, b_4), and (c_1, c_2, c_3, c_4) represent brain sites of the forehead, temporal, and posterior, respectively. σ represents its STD.

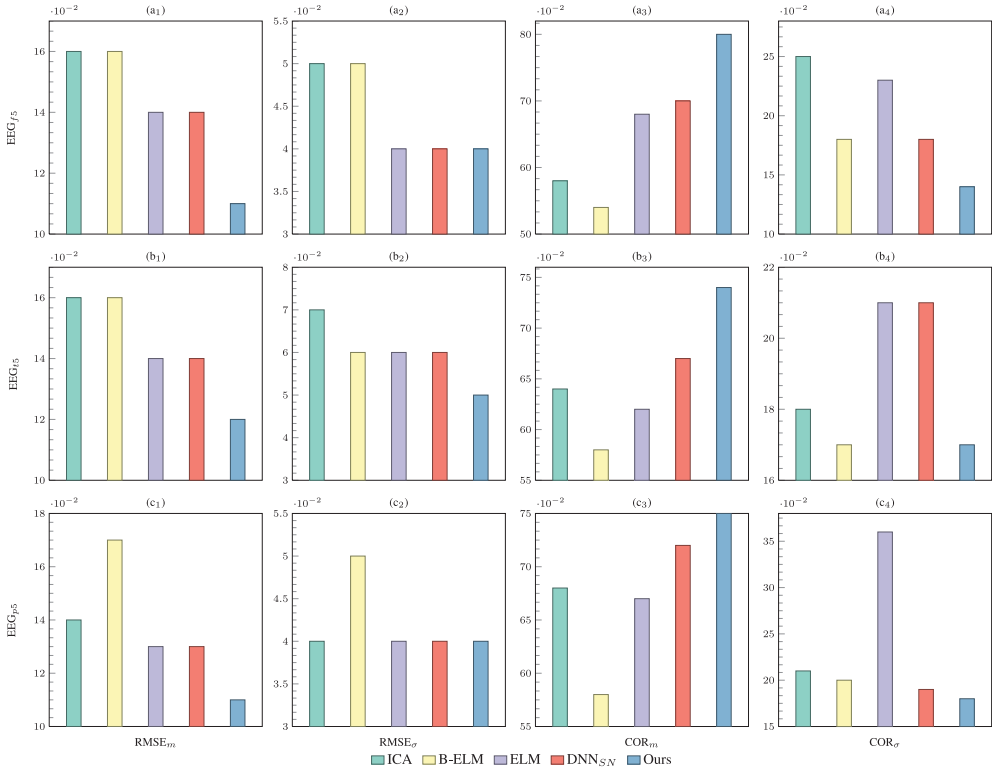


Fig. 9. Performance using three sites EEG with five frequency bands. (a_1, a_2, a_3, a_4), (b_1, b_2, b_3, b_4), and (c_1, c_2, c_3, c_4) represent brain sites of the forehead, temporal, and posterior, respectively. σ represents its STD.

and the proposed method; using the DE feature therefore appears to have an effect consistent with that found in previous studies [57].

We observed that ICA regressor EEG-based vigilance estimation has a promising performance and the better RMSE$_m$/COR$_m$ of EEG$_f$, EEG$_t$, and EEG$_p$ which are

TABLE X
EXPERIMENTAL RESULTS FOR MULTIMODALITY. THE BEST RESULTS ARE BOLDED

Methods	ELM	B-ELM	SVR	CCRF
RMSE_m	0.11	0.12	0.10	0.10
COR_m	0.78	0.70	0.83	0.84
Methods	CCNF	DNN_{SN}	Ours_{ap}	Ours_{mp}
RMSE_m	0.09	0.09	**0.08**	**0.08**
COR_m	0.85	0.85	0.88	**0.89**

0.15/0.65, 0.14/0.67, and 0.13/0.70, respectively. We used its performance as the benchmark. Similarly, the ELM regressor has performed well on each of the EEG features. We noticed that the DNN_{SN} approach has a remarkable performance, and the $\text{RMSE}_m/\text{COR}_m$ of EEG_f, EEG_t, and EEG_p is 0.14/0.70, 0.13/0.70, and 0.13/0.72, respectively. Once the proposed method adds the multichannel autoencoder network to the regressor model, the performance significantly improves, and the $\text{RMSE}_m/\text{COR}_m$ using every type of EEG feature is 0.11/0.80 ($p < 0.01/p < 0.01$, ANOVA), 0.11/0.74 ($p < 0.01/p < 0.01$, ANOVA), and 0.11/0.77 ($p < 0.01/p < 0.01$, ANOVA), respectively. Not only did we demonstrate that the posterior site contains more crucial information for vigilance assessment, consistent with our previous studies [58]; we also found that the forehead site contains key information as well. It is worth noting that the performance we obtained using the feature extracted with the 5-bands method performed as well as the 2 Hz method in our strategy. Thus, we verified the effectiveness of our single-modality strategy for vigilance estimation using different EEG features.

In short, whether using EOG or EEG, the proposed method achieves the best results and far outperforms other methods.

B. Multimodality

We used the complementarity characteristic between EOG and EEG signals to test various multimodal regression methods with features fusion to assess levels of vigilance: ELM, B-ELM, autoencoder-ELM (AE-ELM) [52], support vector regression (SVR) [59], continuous conditional random field (CCRF) [60], continuous conditional neural field (CCNF) [61], DNN_{SN}, and the proposed method with two pooling types: 1) Ours_{ap} and 2) Ours_{mp}. Table X shows all experimental results and Ours_{ap} and Ours_{mp} are the proposed method, with average pooling and max pooling, respectively.

The mean RMSE/COR of ELM was significantly improved to 0.11/0.78, for which performance is much better than for its single modality. SVR achieves the COR value of 0.83, which shows that it is a popular and robust regression method in machine learning.

In addition, we can also observe that the $\text{RMSE}_m/\text{COR}_m$ of the CCNF and CCRF with temporal dependency is 0.10/0.84 and 0.09/0.85, respectively, marking a great improvement in performance, which proves its ability to predict continuous vigilance levels. The convolution parameter errors produce the deviation in the mean estimates, which can be reduced by early fusion with the max pooling used in the DNN_{SN} model.

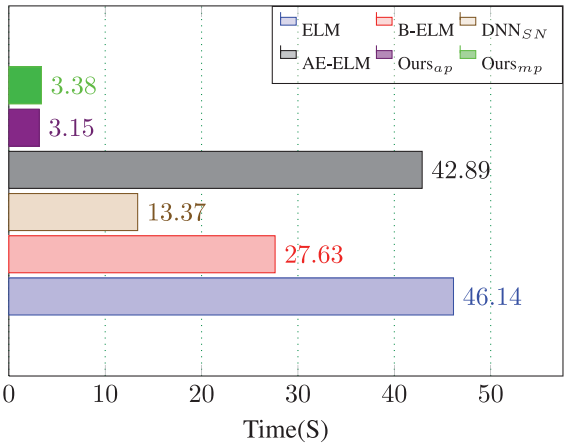


Fig. 10. Computational time analysis.

The performance of the DNN_{SN} model obviously improved to 0.09/0.85 and demonstrated its effectiveness.

Compared to other fusion strategies, the proposed method uses input features extracted by the calculated decoding layer of the multichannel autoencoder model and uses the subnetwork neurons of multilayer regression models to extract subspace features and fusion subspace features, based on early fusion, which could allow it to obtain a nearly 12% boost. The performance of the proposed multimodality method with average pooling and max pooling is 0.08/0.88 ($p < 0.01/p < 0.01$, ANOVA) and 0.08/0.89 ($p < 0.01/p < 0.01$, ANOVA), respectively. Furthermore, from training the original signal to displaying detection, the proposed method only takes 3 s to achieve good performance (see Fig. 10), which is nearly ten times faster than other methods. The robustness of the proposed method is proven through the lowest RMSE_m, the highest COR_m, and the lowest time cost, obtained with two pooling types in subspace feature combination. All experimental results of the compared multimodality algorithms perform better than the single-modality method, and the proposed multimodal method performs the best of all.

V. CONCLUSION

In this article, we proposed a novel multilayer network structure for vigilance estimation, MAE-MELM_{sn}, which is composed of the multichannel autoencoders with subnetwork neurons for dimensionality reduction and signal reconstruction. Moreover, compared with other methods of feature selection, the training of our system achieves higher learning accuracy. Simultaneously, the higher efficiency of decoding the brain signals can better identify the specific relationship between the brain activity and cognitive state, while providing evidence and support to aid in decoding brain states and understanding information processing mechanisms [62]. We then used mixed features with complementary characteristics. The proposed multimodality method shows strong performance in identifying the vigilance states of our brain activity and proves to work better than other state-of-the-art single modality and multimodality approaches.

Although eye movement is a promising indicator of arousal states, eye movement alone is not enough to develop a

robust vigilance estimation model. Eye movement can be intentionally controlled by subjects, which causes degraded performance in prediction. Recently, the robustness of intelligent systems based on machine learning has drawn great attention [63]. Intentional eye movements could be considered as adversarial examples in comparison with spontaneous eye movements. However, we can leverage additional information from EEG to differentiate spontaneous and intentional eye movements. The changes in brain activities contribute to early warnings of reduced vigilance. How to improve the robustness of multimodal vigilance estimation systems need further systematic investigation.

We noticed that the physiological signal varies from person to person. If we obtain more experimental samples and select a larger age range, we can also verify the effectiveness of this model, undoubtedly providing more convincing results. Due to research funding and time constraints, however, all subjects were students recruited from the university campus, with a relatively narrow age range. Using experimental data, we could work to better understand the relationship between age and this model. This is the focus of our future work. Meanwhile, we would like to propose an efficient general method of converting the tabular signals into 2-D shape signals. By doing so, the CNNs, such as ResNet and DenseNet, could be directly combined to improve performance.

REFERENCES

[1] B. S. Oken, M. C. Salinsky, and S. Elsas, "Vigilance, alertness, or sustained attention: Physiological basis and measurement," *Clin. Neurophysiol.*, vol. 117, no. 9, pp. 1885–1901, 2006.

[2] J. Kim, T. Nakamura, H. Kikuchi, and Y. Yamamoto, "Psychobehavioral validity of self-reported symptoms based on spontaneous physical activity," in *Proc. IEEE 37th Ann. Int. Conf. Eng. Med. Biol. Soc. (EMBC)*, Milan, Italy, 2015, pp. 4021–4024.

[3] R. Parasuraman, J. S. Warm, and J. E. See, "Brain systems of vigilance," in *The Attentive Brain*. Cambridge, MA, USA: MIT Press, 1998, pp. 221–256.

[4] M. Steriade, "Coherent oscillations and short-term plasticity in corticothalamic networks," *Trends Neurosci.*, vol. 22, no. 8, pp. 337–345, 1999.

[5] Centers for Disease Control and Prevention (CDC), "Drowsy driving—19 states and the district of Columbia, 2009–2010," *Morbidity Mortality Weekly Rep.*, vol. 61, nos. 51–52, pp. 1033–1037, 2013.

[6] A. G. Wheaton, R. A. Shults, D. P. Chapman, E. S. Ford, and J. B. Croft, "Drowsy driving and risk behaviors—10 states and puerto rico, 2011–2012," *Morbidity Mortality Weekly Rep.*, vol. 63, no. 26, pp. 557–562, 2014.

[7] A. H. Goodwin, L. Thomas, B. Kirley, W. Hall, N. P. O'Brien, and K. Hill, "Countermeasures that work: A highway safety countermeasure guide for state highway safety offices: 2015," Nat. Highway Traffic Safety Admin. (NHTSA), Washington, DC, USA, Rep. DOT HS 812 202, 2015.

[8] G. Sikander and S. Anwar, "Driver fatigue detection systems: A review," *IEEE Trans. Intell. Transp.*, vol. 20, no. 6, pp. 2339–2352, Jun. 2019.

[9] R. Gupta, K. Aman, N. Shiva, and Y. Singh, "An improved fatigue detection system based on behavioral characteristics of driver," in *Proc. IEEE Int. Conf. Intell. Transp. Eng. (ICITE)*, Singapore, 2017, pp. 227–230.

[10] B. Mandal, L. Li, G. S. Wang, and J. Lin, "Towards detection of bus driver fatigue based on robust visual analysis of eye state," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 3, pp. 545–557, Mar. 2017.

[11] Y. Qiao, K. Zeng, L. Xu, and X. Yin, "A smartphone-based driver fatigue detection using fusion of multiple real-time facial features," in *Proc. IEEE Consum. Commun. Netw. Conf. (CCNC)*, Las Vegas, NV, USA, 2016, pp. 230–235.

[12] Z. Wang, R. Zheng, T. Kaizuka, K. Shimono, and K. Nakano, "The effect of a haptic guidance steering system on fatigue-related driver behavior," *IEEE Trans. Human-Mach. Syst.*, vol. 47, no. 5, pp. 741–748, Oct. 2017.

[13] I. C. Jeong, D. H. Lee, S. W. Park, J. I. Ko, and H. R. Yoon, "Automobile driver's stress index provision system that utilizes electrocardiogram," in *Proc. IEEE Intell. Veh. Symp. (ITSS)*, Istanbul, Turkey, 2007, pp. 652–656.

[14] Q. Wu, Y. Zhao, and X. Bi, "Driving fatigue classified analysis based on ECG signal," in *Proc. IEEE 5th Int. Symp. Comput. Intell. Design (ISCID)*, vol. 2. Hangzhou, China, 2012, pp. 544–547.

[15] J. Boyle, N. Bidargaddi, A. Sarela, and M. Karunanithi, "Automatic detection of respiration rate from ambulatory single-lead ECG," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 6, pp. 890–896, Nov. 2009.

[16] M. Knaflitz and F. Molinari, "Assessment of muscle fatigue during biking," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 1, pp. 17–23, Mar. 2003.

[17] L. Boon-Leng, L. Dae-Seok, and L. Boon-Giin, "Mobile-based wearable-type of driver fatigue detection by GSR and EMG," in *Proc. IEEE Region 10 Conf. (TENCON)*, Macao, China, 2015, pp. 1–4.

[18] J. Li, S. Qiu, Y.-Y. Shen, C.-L. Liu, and H. He, "Multisource transfer learning for cross-subject EEG emotion recognition," *IEEE Trans. Cybern.*, vol. 50, no. 7, pp. 3281–3293, Jul. 2020.

[19] W.-L. Zheng, W. Liu, Y. Lu, B.-L. Lu, and A. Cichocki, "EmotionMeter: A multimodal framework for recognizing human emotions," *IEEE Trans. Cybern.*, vol. 49, no. 3, pp. 1110–1122, Mar. 2019.

[20] W. Wu *et al.*, "Faster single model vigilance detection based on deep learning," *IEEE Trans. Cogn. Develop. Syst.*, early access, Dec. 31, 2019, doi: 10.1109/TCDS.2019.2963073.

[21] W.-L. Zheng and B.-L. Lu, "A multimodal approach to estimating vigilance using EEG and forehead EOG," *J. Neural. Eng.*, vol. 14, no. 2, 2017, Art. no. 026017.

[22] X.-Q. Huo, W.-L. Zheng, and B.-L. Lu, "Driving fatigue detection with fusion of EEG and forehead EOG," in *Proc. IEEE Int. Joint Conf. Neural Netw. (IJCNN)*, Vancouver, BC, Canada, 2016, pp. 897–904.

[23] W. Wu *et al.*, "A regression method with subnetwork neurons for vigilance estimation using EOG and EEG," *IEEE Trans. Cogn. Develp. Syst.*, early access, Dec. 21, 2018, doi: 10.1109/TCDS.2018.2889223.

[24] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[25] H. Liu, J. Qin, F. Sun, and D. Guo, "Extreme kernel sparse learning for tactile object recognition," *IEEE Trans Cybern.*, vol. 47, no. 12, pp. 4509–4520, Dec. 2017.

[26] M. Ferrara and L. De Gennaro, "How much sleep do we need?" *Sleep Med. Rev.*, vol. 5, no. 2, pp. 155–179, 2001.

[27] Y. Yang, Q. J. Wu, and Y. Wang, "Autoencoder with invertible functions for dimension reduction and image reconstruction," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 48, no. 7, pp. 1065–1079, Jul. 2018.

[28] Y. Yang and Q. J. Wu, "Features combined from hundreds of midlayers: Hierarchical networks with subnetwork nodes," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 11, pp. 3313–3325, Nov. 2019.

[29] C. Wang, J. L. Ong, A. Patanaik, J. Zhou, and M. W. Chee, "Spontaneous eyelid closures link vigilance fluctuation with fmri dynamic connectivity states," *Proc. Nat. Acad. Sci.*, vol. 113, no. 34, pp. 9653–9658, 2016.

[30] J. Han, C. Chen, L. Shao, X. Hu, J. Han, and T. Liu, "Learning computational models of video memorability from fMRI brain imaging," *IEEE Trans. Cybern.*, vol. 45, no. 8, pp. 1692–1703, Aug. 2015.

[31] G. Deshpande, P. Wang, D. Rangaprakash, and B. Wilamowski, "Fully connected cascade artificial neural network architecture for attention deficit hyperactivity disorder classification from functional magnetic resonance imaging data," *IEEE Trans. Cybern.*, vol. 45, no. 12, pp. 2668–2679, Dec. 2015.

[32] W. Wierwille, "Historical perspective on slow eyelid closure: Whence PERCLOS," in *Proc. Ocular Meas. Driver Alertness Techn. Conf.*, 1999, pp. 31–52.

[33] J.-F. Xie, M. Xie, and W. Zhu, "Driver fatigue detection based on head gesture and PERCLOS," in *Proc. IEEE Int. Conf. Wavelet Active Media Technol. Inf. Process. (ICWAMTIP)*, Chengdu, China, 2012, pp. 128–131.

[34] R. Grace and R. K. Davis, "Apparatus and method of monitoring a subject's eyes using two different wavelengths of light," U.S. Patent 6 082 858, Jul. 2000.

[35] Y.-K. Wang, T.-P. Jung, and C.-T. Lin, "EEG-based attention tracking during distracted driving," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 23, no. 6, pp. 1085–1094, Nov. 2015.

[36] X.-Y. Gao, Y.-F. Zhang, W.-L. Zheng, and B.-L. Lu, "Evaluating driving fatigue detection algorithms using eye tracking glasses," in *Proc. 7th Int. IEEE/EMBS Conf. Neural Eng.*, Montpellier, France, 2015, pp. 767–770.

[37] A. A. Hayawi and J. Waleed, "Driver's drowsiness monitoring and alarming auto-system based on EOG signals," in *Proc. IEEE 2nd Int. Conf. Eng. Technol. Appl. (IICETA)*, Al-Najef, Iraq, 2019, pp. 214–218.

[38] H.-Y. Cai, J.-X. Ma, L.-C. Shi, and B.-L. Lu, "A novel method for EOG features extraction from the forehead," in *Proc. Ann. IEEE Int. Conf. Eng. Med. Biol. Soc. (EMBC)*, Boston, MA, USA, 2011, pp. 3075–3078.

[39] C.-T. Lin *et al.*, "Adaptive EEG-based alertness estimation system by using ICA-based fuzzy neural networks," *IEEE Trans. Circuits-I.*, vol. 53, no. 11, pp. 2469–2476, Nov. 2006.

[40] A. Bulling, J. A. Ward, H. Gellersen, and G. Tröster, "Eye movement analysis for activity recognition using electrooculography," *IEEE Trans. Pattern Anal. Mech. Intell.*, vol. 33, no. 4, pp. 741–753, Apr. 2011.

[41] X. Li, C. Guan, H. Zhang, and K. K. Ang, "Discriminative ocular artifact correction for feature learning in EEG analysis," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 8, pp. 1906–1913, Aug. 2017.

[42] R. Mahajan and B. I. Morshed, "Unsupervised eye blink artifact denoising of EEG data with modified multiscale sample entropy, Kurtosis, and wavelet-ICA," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 1, pp. 158–165, Jan. 2015.

[43] R. N. Khushaba, S. Kodagoda, S. Lal, and G. Dissanayake, "Driver drowsiness classification using fuzzy wavelet-packet-based feature-extraction algorithm," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 1, pp. 121–131, Jan. 2011.

[44] Y. Dong, S. Gao, K. Tao, J. Liu, and H. Wang, "Performance evaluation of early and late fusion methods for generic semantics indexing," *Pattern Anal. Appl.*, vol. 17, no. 1, pp. 37–50, 2014.

[45] A. Kasagi, T. Tabaru, and H. Tamura, "Fast algorithm using summed area tables with unified layer performing convolution and average pooling," in *Proc. IEEE 27th Int. Workshop Mach. Learn. Signal Process. (MLSP)*, Tokyo, Japan, 2017, pp. 1–6.

[46] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)* Boston, MA, USA, 2015, pp. 1–9.

[47] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014. [Online]. Available: arXiv:1409.1556.

[48] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances Neural Information Processing Systems (NIPS)*. Red Hook, NY, USA: Curran, 2012, pp. 1097–1105.

[49] Y. Yang, Q. J. Wu, W.-L. Zheng, and B.-L. Lu, "EEG-based emotion recognition using hierarchical network with subnetwork nodes," *IEEE Trans. Cogn. Develop. Syst.*, vol. 10, no. 2, pp. 408–419, Jun. 2018.

[50] H. P. Martinez, Y. Bengio, and G. N. Yannakakis, "Learning deep physiological models of affect," *IEEE Comput. Intell. Mag.*, vol. 8, no. 2, pp. 20–33, May 2013.

[51] Y. Yang and Q. J. Wu, "Extreme learning machine with subnetwork hidden nodes for regression and classification," *IEEE Trans. Cybern.*, vol. 46, no. 12, pp. 2885–2898, Dec. 2016.

[52] Y. Yang and Q. J. Wu, "Multilayer extreme learning machine with subnetwork nodes for representation learning," *IEEE Trans. Cybern.*, vol. 46, no. 11, pp. 2570–2583, Nov. 2016.

[53] T. D'haene, R. Pintelon, J. Schoukens, and E. Van Gheem, "Variance analysis of frequency response function measurements using periodic excitations," *IEEE Trans. Instrum. Meas.*, vol. 54, no. 4, pp. 1452–1456, Aug. 2005.

[54] F. Hassainia, V. Medina, J. Stauder, L. Mottron, and P. Robaey, "The use of F-statistic mapping as a complementary tool to t-statistic mapping in group comparisons," in *Proc. IEEE 17th Int. Conf. Eng. Med. Biol. Soc. (EMBC)*, vol. 2. Montreal, QC, Canada, 1995, pp. 1013–1014.

[55] Z. Huang, Y. Yu, J. Gu, and H. Liu, "An efficient method for traffic sign recognition based on extreme learning machine," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 920–933, Apr. 2017.

[56] Y. Yang, Y. Wang, and X. Yuan, "Bidirectional extreme learning machine for regression problem and its learning effectiveness," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 9, pp. 1498–1505, Sep. 2012.

[57] R.-N. Duan, J.-Y. Zhu, and B.-L. Lu, "Differential entropy feature for EEG-based emotion classification," in *Proc. IEEE 6th Int. IEEE/EMBS Conf. Neural Eng. (NER)*, San Diego, CA, USA, 2013, pp. 81–84.

[58] L.-C. Shi and B.-L. Lu, "EEG-based vigilance estimation using extreme learning machines," *Neurocomputing*, vol. 102, pp. 135–143, Feb. 2013.

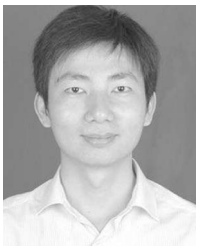
[59] Y. Tian, Z. Qi, X. Ju, Y. Shi, and X. Liu, "Nonparallel support vector machines for pattern classification," *IEEE Trans. Cybern.*, vol. 44, no. 7, pp. 1067–1079, Jul. 2014.

[60] T. Baltrušaitis, N. Banda, and P. Robinson, "Dimensional affect recognition using continuous conditional random fields," in *Proc. 10th IEEE Int. Conf. Workshops Autom. Face Gesture Recognit. (FG)*, Shanghai, China, 2013, pp. 1–8.

[61] V. Imbrasaitė, T. Baltrušaitis, and P. Robinson, "CCNF for continuous emotion tracking in music: Comparison with CCRF and relative feature representation," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Chengdu, China, 2014, pp. 1–6.

[62] J.-D. Haynes and G. Rees, "Decoding mental states from brain activity in humans," *Nat. Rev. Neurosci.*, vol. 7, no. 7, pp. 523–534, 2006.

[63] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," 2014. [Online]. Available: arXiv:1412.6572.



Wei Wu (Member, IEEE) received the M.S. degree in electrical engineering from the Hunan University of Technology, Zhuzhou, China, in 2011. He is currently pursuing the Ph.D. degree in electrical engineering with the College of Electrical and Information Engineering, Hunan University, Changsha, China.

He was with the College of Electrical and Information Engineering, Hunan University of Technology from 2004 to 2008 and from 2011 to 2016. From 2017 to 2019, he awarded a State Scholarship by the China Scholarship Council and was a joint Ph.D. student with the Department of Electrical and Computer Engineering, University of Windsor, Windsor, ON, Canada. His research focuses on affective computing, artificial neural networks, and robotics.

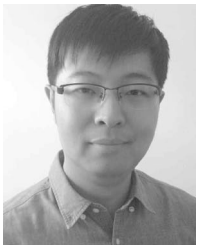
Wei Sun received the B.S., M.S., and Ph.D. degrees from the in control science and engineering from Hunan University, Changsha, China, in 1996, 1999, and 2003, respectively.

He is currently working as a Professor with the College of Electrical and Information Engineering, Hunan University, where he is also the Director of the Hunan Provincial Key Laboratory of Intelligent Robot Technology in Electronic Manufacturing. His areas of interests are computer vision and robotics, neural networks, and intelligent control.

Q. M. Jonathan Wu (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the University of Wales, Swansea, U.K., in 1990.

In 1995, he joined the National Research Council of Canada, Vancouver, BC, Canada, where he became a Senior Research Officer and a Group Leader. He is currently a Professor with the Department of Electrical and Computer Engineering, University of Windsor, Windsor, ON, Canada. He has authored or coauthored more than 300 peer-reviewed papers in computer vision, image processing, intelligent systems, robotics, and integrated microsystems. His current research interests include 3-D computer vision, active video object tracking and extraction, interactive multimedia, sensor analysis and fusion, and visual sensor networks.

Prof. Wu holds the Tier 1 Canada Research Chair in automotive sensors and information systems. He was an Associate Editor of the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS. He is an Associate Editor of the IEEE TRANSACTION ON CYBERNETICS, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, and *Cognitive Computation*. He served on technical program committees and international advisory committees for many prestigious conferences.



Yimin Yang (Senior Member, IEEE) received the Ph.D. degree in pattern recognition and intelligent systems from the College of Electrical and Information Engineering, Hunan University, Changsha, China, in 2013.

From 2014 to 2018, he was a Postdoctoral Fellow with the University of Windsor, Windsor, ON, Canada. He is currently an Assistant Professor with the Computer Science Department, Lakehead University, Thunder Bay, ON, Canada. His current research interests include artificial neural networks, signal processing, and robotics.

Dr. Yang was a recipient of the Outstanding Ph.D. Thesis Award of Hunan Province and the Outstanding Ph.D. Thesis Award Nominations of Chinese Association of Automation, China, in 2014 and 2015, respectively. He is a Program Committee Member of some international conferences. He is an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY and *Neurocomputing*. He has been serving as a reviewer for many international journals of his research field and a Guest Editor of multiple journals.

Hui Zhang (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in pattern recognition and intelligent systems from Hunan University, Changsha, China, in 2004, 2007, and 2012, respectively.

He is currently a Professor with the College of Robot, Hunan University, where he is also the Deputy Director of the National Engineering Laboratory of Robot Visual Perception and Control Technology. His research interests include machine vision, deep learning, and defect detection.

Wei-Long Zheng (Member, IEEE) received the bachelor's degree in information engineering from the Department of Electronic and Information Engineering, South China University of Technology, Guangzhou, China, in 2012, and the Ph.D. degree in computer science from the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China, in 2018.

He is a Research Fellow with the Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA. His research focuses on affective computing, brain–computer interaction, machine learning, and pattern recognition.

Dr. Zheng received the IEEE TRANSACTIONS ON AUTONOMOUS MENTAL DEVELOPMENT Outstanding Paper Award from the IEEE Computational Intelligence Society in 2018.

Bao-Liang Lu (Senior Member, IEEE) received the B.S. degree in instrument and control engineering from the Qingdao University of Science and Technology, Qingdao, China, in 1982, the M.S. degree in computer science and technology from Northwestern Polytechnical University, Xi'an, China, in 1989, and the Dr.Eng. degree in electrical engineering from Kyoto University, Kyoto, Japan, in 1994.

He was with the Qingdao University of Science and Technology from 1982 to 1986. From 1994 to 1999, he was a Frontier Researcher with the Bio-Mimetic Control Research Center, Institute of Physical and Chemical Research (RIKEN), Nagoya, Japan, and a Research Scientist with the RIKEN Brain Science Institute, Wako, Japan, from 1999 to 2002. Since 2002, he has been a Full Professor with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China. His current research interests include brain-like computing, neural networks, machine learning, brain–computer interaction, and affective computing.

Prof. Lu received the IEEE TRANSACTIONS ON AUTONOMOUS MENTAL DEVELOPMENT Outstanding Paper Award from the IEEE Computational Intelligence Society in 2018. He is currently a Board Member of the Asia Pacific Neural Network Society (APNNS, previously APNNA) and the Steering Committee member of the IEEE TRANSACTIONS ON AFFECTIVE COMPUTING. He was the President of the Asia Pacific Neural Network Assembly (APNNA) and the General Chair of the 18th International Conference on Neural Information Processing in 2011. He is currently an Associate Editor of the IEEE TRANSACTIONS ON COGNITIVE AND DEVELOPMENT SYSTEMS.