Faster Single Model Vigilance Detection Based on Deep Learning

Wei Wu, *Student Member, IEEE*, Wei Sun, Q. M. Jonathan Wu, *Senior Member, IEEE*, Cheng Zhang, Yimin Yang, *Senior Member, IEEE*, Hongshan Yu, and Bao-Liang Lu, *Senior Member, IEEE*

***Abstract*—Various reports have shown that the rate of road traffic accidents has increased due to reduced driver vigilance. Therefore, an accurate estimation of the driver's alertness status plays an important part. To estimate vigilance, we adopt a novel strategy that is a deep autoencoder with subnetwork nodes (DAE$_{SN}$). The proposed network model is designed not only for sparse representation but also for dimension reduction. Some hidden layers are not calculated by randomly acquired, but by replacement technologies. Unlike the traditional electrooculogram (EOG) signals, the forehead EOG (EOG$_F$) signals are collected through forehead electrodes that do not have to surround the eyes, which has a convenient and effective practical application. The root-mean-square error (RMSE) and correlation coefficient (COR) while separately using three EOG$_F$ features improved to 0.11/0.79, 0.10/0.83, and 0.11/0.80, respectively. Implemented in an experimental environment, percentage of eye closure over time is calculated in real time through SMI eye-tracking-glasses, up to 120 frames/s. In addition, the time to extract features from the raw signal and display the prediction is only 34 ms, that is the level of the driver's fatigue can be detected quickly. The experimental study shows that the proposed model for vigilance analysis has better robustness and learning capability.**

***Index Terms*—Deep learning (DL), dimension reduction, single model, vigilance detection.**

I. Introduction

SCIENTISTS pointed out that machines can learn and write code like a human. This probably was the embryonic definition of machine learning (ML) in the mid-twentieth century. The term ML was coined by Samuel [1] who indicated that ML gives computers the ability to "learn" from the data based on statistical techniques and progressively improve the performance for a specific task. The perception was perhaps the first computer neural network model based on neurosensory science [2] and simulated the way the human brain thinks. Since the neural network learning machine has a technical defect that can learn only a single concept, and as the limited memory and processing speed of the computer is not enough to solve any actual artificial intelligence problem, the development of ML has been slow.

Generally, an object's current behavior performance (e.g., service quality) is often correlated with its past performance records [3]–[5]. Based on this observation, more predictive methods based on deep learning (DL) algorithms are introduced to improve many practical applications. Gradually they entered into all kinds of commercial areas [6]–[11], i.e., image recognition, computer vision, robotics, etc. For example, Levine *et al.* [10] built a larger data set of more than 800 000 grasp attempts, which was collected by almost 14 robotic manipulators for two months. He then trained then grasp prediction models based on a deep convolutional neural network and the average grasp success rate was 84%. These deep networks increase the computational complexity and even lead to the curse of dimensionality [12] because the linear learning machines in the high-dimensional feature space and sparse data [13] need to obtain the explicit expression of the nonlinear mapping, which does not exist in a linear model. The strong generalization capability of ML arises from the optimal features of the data set, which is generated by human ingenuity and domain knowledge. The workloads that are significantly increased by processing high-dimensional data sets which should be reduced to extract the high-quality features required for ML algorithms [14]. Combining the dimension reduction [15] and feature extraction [16], [17] can be considered as a fast and effective way in emotion recognition.

Manuscript received October 17, 2019; accepted December 22, 2019. Date of publication December 31, 2019; date of current version September 9, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant U1813205, Grant 61573135, Grant 61973106, and Grant 61673266; in part by the Planned Science and Technology Project of Hunan Province under Grant 2016TP1023 and Grant 2017TP1011; in part by the Natural Sciences and Engineering Research Council of Canada Discovery Grant; in part by the Key Research and Development Project of Science and Technology Plan of Hunan Province under Grant 2018GK2021; in part by the Natural Science Foundation of Hunan Province under Grant 2017JJ3118; in part by the National Key Research and Development Program of China under Grant 2017YFB1002501; and in part by the Fundamental Research Funds for the Central Universities. *(Corresponding authors: Wei Sun; Cheng Zhang.)*

Wei Wu, Wei Sun, and Hongshan Yu are with the College of Electrical and Information Engineering, the State Key Laboratory of Advanced Design and Manufacturing for Vehicle Body, and the Hunan Key Laboratory of Intelligent Robot Technology in Electronic Manufacturing, Hunan University, Changsha 410082, China.

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

Cheng Zhang is with the College of Electrical and Information Engineering, Hunan University of Technology, Zhuzhou 412007, China.

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

Bao-Liang Lu is with the Department of Computer Science and Engineering and 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 figures in this article are available at https://doi.org/10.1109/TCDS.2019.2963073.

Digital Object Identifier 10.1109/TCDS.2019.2963073

According to Canadian police reports [18], reduced vigilance while driving is mainly a contribute factor for almost over 60 fatal transport accidents per month. In public transportation safety (PTS), therefore, vigilance state estimation of drivers' state has become a vital task. Typically, electroencephalogram (EEG) [19], [20], and EOG [21]–[23] are mainly used to estimate the level of vigilance. Compared to EEG with a lower signal-to-noise ratio [24], EOG makes more robust to noise since the amplitude of which has significantly increased. Unlike traditional EOG signals, which are collected by two pairs of electrodes channel and one reference channel and have no practical application [25], we now present the forehead EOG (EOG$_F$) is recorded by the new electrode placement with a suitable wearable device. It has been proven that the percentage of eye closure over time (PERCLOS) index can be considered as a good measure to estimate alertness in several variables of indicators [26]. Unlike traditional facial video technology [27], the PERCLOS is automatically calculated by SMI eye-tracking glasses with a high resolution of 120 Hz, which makes it suitable for real-time fatigue detection. As our previously reported [28], the EOG$_F$ has two characteristics of high signal-to-noise ratio and easy setup. In general, the methods, using EOG$_F$ that can be an objective and comprehensive reflection of the real physical state of the human, which gradually played a vital role to estimate vigilance.

In this article, we adopt a novel strategy that is a deep autoencoder with subnetwork nodes (DAE$_{SN}$). In particular, the contributions of this article are as follows.

1) Unlike traditional multilayer extreme learning machine (M-ELM) networks, where hidden nodes are calculated by randomly acquiring an encoding layer, the current features of DAE$_{SN}$ are obtained by replacing the previous decoding layer and simplifying more useful features for pattern recognition.
2) Unlike traditional multilayer autoencoder (MAE) approaches that only work for classification, DAE$_{SN}$ is adopted for data reconstruction, sparse representation, and dimension reduction. Meanwhile, our training speed can be several times or even dozens of times faster than other related methods.
3) Unlike traditional EOG signals collected through electrodes that surround the eyes, the EOG$_F$ collected by a convenient and practical way. Meanwhile, EOG$_F$ contains more important information in eye movements, including saccade, blink, and fixation component. Furthermore, PERCLOS is calculated by SMI eye-tracking glasses that have 120 frames/s, which can reflect eye movement in real time.

II. Methodology

A. Proposed Method

The proposed multilayered model with subnetwork nodes that subsumes autoencoder and regression networks. Fig. 1 depicts the entire process flow of the proposed model from the data preprocessing stage to the final regression stage where the blue circles represent the hidden nodes of the regression network, while the green circles denote the subnetwork nodes.

Algorithm 1 Proposed Algorithm

Part A: Subspace feature dimension reduction and extraction

Step (1) Original input data are transformed into a feature subspace through the random weight initialized first encoding layer.

Step (2) In the first decoding layer, the representational features are extracted from the hidden nodes in the subnetwork node. Thus, at this stage the feature dimension is equal to the total number of the hidden nodes in the subnetwork node. The optimal number of nodes is selected through empirical analysis.

Step (3) The output of the first decoding layer becomes the input to the second encoding layer. Sequentially, the feature H_e^1 of the 1st subnetwork node is obtained.

Step (4) Through parameter updating and adjustments based on reversible functions, the feature H_e^2 of the 2nd subnetwork node is computed.

Step (5) Finally, the features H_e of the subnetwork nodes are obtained by repeating the above steps several times. Thus, the high-dimensional input data is mapped into random subspace, and then converted to the low-dimensional feature.

Part B: Regression for vigilance estimation

Regression analysis was performed on the extracted low-dimensional feature set. The output is a continuous value between 0 and 1. By setting a double threshold of '0.35' and '0.70', three states can be derived, such as "drowsy state," "tired state," and "awake state." For example, the level between 0 to 0.35 represents the awake state.

Here, X is the EOG$_F$. The input data can reduce dimensionality and extract subspace features by the proposed model. The proposed algorithm shows the details.

In short, the accuracy of regression has a promising result with the above multilayered process instead of iterative back-propagation (BP)-based network training. Thus, it reduces a lot of computational overheads. Besides, the proposed approach utilizes two or three subnetwork nodes, only. It means that the training time is greatly reduced.

B. Autoencoders

Autoencoders that are special neural networks more and more widely used in the unsupervised learning. We set the initial input $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$ and the rebuilt output $\mathbf{H(x)} = (\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n)^T$. It uses the BP algorithm in unsupervised learning for training and the formula is

$$\mathbf{H(x)} = S(\mathbf{a}, \mathbf{b}, \mathbf{x})$$
$$\mathbf{J} = \frac{1}{2}\left|\left|\mathbf{H(x)} - \mathbf{x}\right|\right| \tag{1}$$

where $\mathbf{J}$ is the reconstruction error. The training goal of this model is to minimize the error $\mathbf{J}$ so that $\mathbf{H(x)}$ is close to $\mathbf{x}$.

According to our previous study, Yang *et al.* [16] proposed a double-layer autoencoder structure for image reconstruction. The formula of rebuilt output is

$$\mathbf{H(x)} = S(\mathbf{a}_i, \mathbf{b}_i, H(x_{i-1})). \tag{2}$$

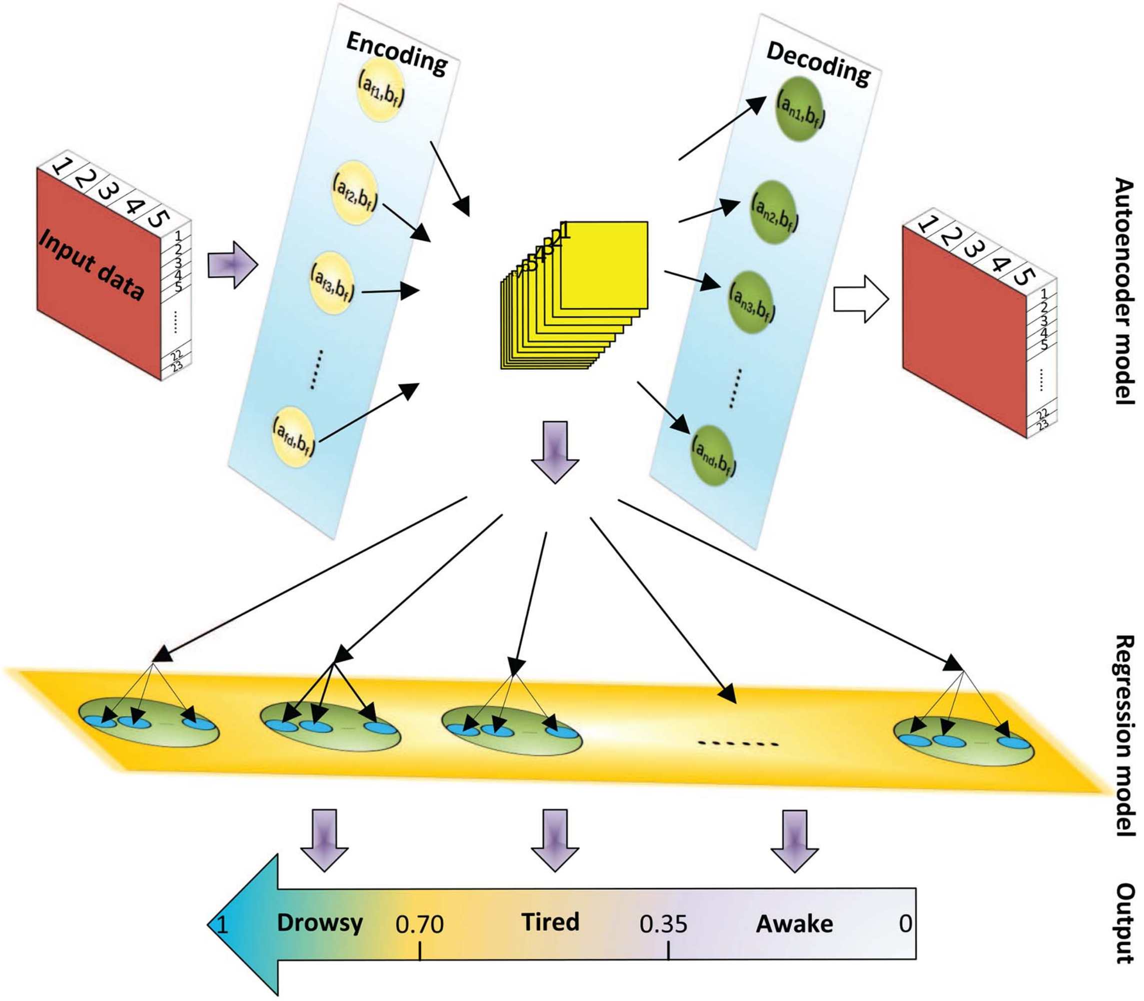


Fig. 1. Framework of the proposed method. Yellow and green dots in the autoencoder model represent subnetwork nodes. Green and blue dots in the regression model represent subnetwork nodes and hidden nodes, respectively.

Since its hidden nodes are calculated, the rebuilt output for image reconstruction is more closely related to the input data.

Most autoencoders [29], [30] have shown their benefits in 2-D images or 3-D space. We tried to use it for 1-D EOG signals and hope gets a good performance. The rebuilt output can be expressed as

$$\mathbf{H}^j = S\left(\left[\mathbf{a}_1^j, \ldots, \mathbf{a}_d^j\right], \mathbf{b}^j, \mathbf{x}\right)$$
$$\mathbf{H}(\mathbf{x}) = \sum_{j=1}^{n} \mathbf{H}^j \tag{3}$$

where j represents the jth subnetwork nodes, and n represents total subnetwork nodes. d represents the number of hidden nodes.

Each subnetwork node is independent, and it contains hidden nodes as a separate system, and the connection is only between adjacent subnetwork nodes. Our network is more like a complete system (refer to the steps are shown in the next section). Thus, the proposed method can really shorten the training time and improve learning efficiency when compared to the iterative BP-based training procedures.

C. Dimensionality Reduction

Given M arbitrary distinct samples $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{M}$ ($\mathbf{x}_i \in \mathbb{R}^n, \mathbf{y}_i \in \mathbb{R}^m$). $\mathbb{R}$ represents sets of real numbers. $\mathbf{x}$ and $\mathbf{y}$ are the input data and output data, respectively. Table I defined all notations we used.

The autoencoder performs an unsupervised manner of feature extraction from the raw data. In the following, the raw input data would be converted to low-dimensional features, and several processes can be described as follows.

Step 1: Given any arbitrary different training samples from continuous systems $\{(\mathbf{x}_k, \mathbf{y}_k)\}_{k=1}^{M_a}$ ($\mathbf{x}_k \in \mathbb{R}^{n_a}$). Output data were reconstructed by autoencoder, which similar to the input data, here $\mathbf{x} = \mathbf{y}$. Then for the encoding layer of autoencoder, randomly generated input initial weights and bias can be expressed as

$$\mathbf{H}_e = S_a(\mathbf{a}_e \cdot \mathbf{x} + b_e)$$
$$(\mathbf{a}_e)^T \cdot \mathbf{a}_e = \mathbf{I}, (b_e)^T \cdot b_e = 1 \tag{4}$$

where $\mathbf{a}_e \in \mathbb{R}^{d_a \times n_a}$ and $b_e \in \mathbb{R}$ indicate orthogonal random weight and bias, respectively. $\mathbf{H}_e$ indicates current feature data.

Step 2: Given any continuous desired outputs $\mathbf{y}$ and the function $S_a = \sin(\cdot)$, the best parameters of the hidden layer

TABLE I
SYMBOLS DESCRIPTION

Model	Symbol	Property
Autoencoder	$\mathbf{a}_e$	input weights in encoding layer. $\mathbf{a}_e \in \mathbb{R}^{d_a \times m_a}$
	b_e	bias in encoding layer $b_e \in \mathbb{R}$.
	$(\mathbf{a}_e^j, b_e^j)$	the jth hidden node in encoding layer.
	$(\mathbf{a}_{ei}, b_e)$	the ith subnetwork node in encoding layer.
	$(\mathbf{a}_d, b_d)$	hidden nodes in decoding layer and $\mathbf{a}_d \in \mathbb{R}^{m_a \times d_a}$.
	$\mathbf{H}_e$	feature data generated by a encoding layer
	$\mathbf{H}_e^i$	feature data generated by the ith encoding layer
Regression	$\mathbf{a}_i$	the weight connecting the ith hidden node and the input nodes.
	b_i	the bias of the ith hidden node.
	$\hat{\mathbf{a}}_p^i$	input weight of the ith subnetwork node in entrance layer. $\hat{\mathbf{a}}_p^j \in \mathbb{R}^{d_r \times m_r}$
	$\hat{\mathbf{a}}_q^i$	input weight of the ith subnetwork node in exit layer. $\hat{\mathbf{a}}_i^q \in \mathbb{R}^{d_r \times n_r}$
	$\hat{b}_p^i$	bias of the ith subnetwork node in entrance layer. $\hat{b}_p^j \in \mathbb{R}$
	$(\mathbf{a}_{pi}^j, b_p^j)$	the ith hidden node of the jth subnetwork node.
	$\mathbf{H}_p^j$	feature data generated by jth subnetwork nodes.

$\{\hat{\mathbf{a}}_h, \hat{b}_h\}$ are obtained by the following formula:

$$\begin{aligned}\mathbf{a}_h &= S_a^{-1}(L_a(\mathbf{y})) \cdot (\mathbf{H}_e)^{-1} \quad, \mathbf{a}_h \in \mathbb{R}^{d_a \times m_a}\\ b_h &= \sqrt{\mathrm{MSE}\Big(\mathbf{a}_h \cdot \mathbf{H}_e - S_a^{-1}(L_a(\mathbf{y}))\Big)} \quad, L_a \in \mathbb{R}\\ S_a^{-1}(\cdot) &= \arcsin(\cdot)\end{aligned} \tag{5}$$

where $\mathbf{H}^{-1} = \mathbf{H}^T([C/\mathbf{I}] + \mathbf{H}\mathbf{H}^T)^{-1}$ and C is a regularization value ($C > 0$). MSE is the mean-squared-error. $L(\cdot)$ represents the normalized function.

Step 3: Update $\mathbf{a}_e, b_e$ by

$$\begin{aligned}\mathbf{a}_e &= (\mathbf{a}_h)^T\\ b_e &= \sqrt{\mathrm{MSE}(\mathbf{a}_e \cdot \mathbf{x} - \mathbf{y})} \quad, b_e \in \mathbb{R}\end{aligned} \tag{6}$$

and update the feature data $\mathbf{H}_e = S_a(\mathbf{a}_e \cdot \mathbf{x} + b_e)$.

Step 4: Repeat steps 2 and 3 $l - 1$ times. The feature data $\mathbf{H}_e = S_a(\mathbf{x}, \mathbf{a}_e, b_e)$.

D. Regression Model

According to our previous study [31], the proposed bidirectional extreme learning machine (B-ELM) with the two-layer network has less computational workloads than other deep networks and it excels other models in terms of processing time and accuracy in regression problems. However, the two-layer network has hundreds of hidden nodes, and these hidden nodes are connected to each other, greatly affecting the speed of the network model. It is because the proposed subnetwork node has a direct connection between adjacent subnetwork nodes, only. Thus, the test accuracy and learning efficiency are improved by selecting a small number of nodes and further extracting features. The steps were described as follows.

Step 1: Given $\{(\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^{M_r}, \mathbf{x}_i \in \mathbb{R}^{m_r}$ arbitrary distinct training samples in the entrance layer from a continuous system, the weight ($\hat{\mathbf{a}}_p^k$), and the bias ($\hat{b}_p^k$) obtained by orthogonal random. The initial subspace features of subnetwork neuron $\mathbf{H}_P^k$ are

$$\begin{aligned}\mathbf{H}_p^k &= S_r\Big(\hat{\mathbf{a}}_p^k, \hat{b}_p^k, \mathbf{x}\Big)\\ \Big(\hat{\mathbf{a}}_p^k\Big)^T &\cdot \hat{\mathbf{a}}_p^k = \mathbf{I}\\ \Big(\hat{b}_p^k\Big)^T &\cdot \hat{b}_p^k = \mathbf{1}\end{aligned} \tag{7}$$

where the initial value $k = 1$.

Step 2: Given the S_r activation function of the exit layer for any continuous desired outputs $\mathbf{y}$, the subspace features $(\hat{\mathbf{a}}_q^k, \hat{b}_q^k)$ are obtained by

$$\begin{aligned}\hat{\mathbf{a}}_q^k &= S_r^{-1}(L_r(\mathbf{y})) \cdot \Big(S_r\Big(\hat{\mathbf{a}}_p^k, \hat{b}_p^k, \mathbf{x}\Big)\Big)^{-1}\\ \hat{b}_q^k &= \sqrt{\mathrm{MSE}\Big(\hat{\mathbf{a}}_q \cdot S_r\Big(\hat{\mathbf{a}}_p^k, \hat{b}_p^k, \mathbf{x}\Big) - S_r^{-1}(L(\mathbf{y}))\Big)}\end{aligned} \tag{8}$$

where $\mathbf{H}^{-1} = \mathbf{H}^T(\frac{U}{\mathbf{I}} + \mathbf{H}\mathbf{H}^T)^{-1}$, U represents a regularization value ($U > 0$), $\hat{\mathbf{a}}_q^k \in \mathbb{R}^{d \times n}$, and $\hat{b}_m^k \in \mathbb{R}$. $L(\cdot)$ represents the normalized function.

Step 3: Update $\mathbf{e}_k$, $\hat{\mathbf{a}}_p^k$, and $\hat{b}_p^k$ as

$$\begin{aligned}\mathbf{e}_k &= \mathbf{y} - L_r^{-1} S_r\Big(\mathbf{H}_p^k, \hat{\mathbf{a}}_q^k, \hat{b}_q^k\Big)\\ \hat{\mathbf{a}}_p^k &= S_r^{-1}\Big(L_r\Big(\mathbf{P}_{k-1} + \mathbf{H}_P^k\Big)\Big) \cdot \mathbf{x}^{-1}\\ \hat{b}_p^k &= \sqrt{\mathrm{MSE}\Big(\hat{\mathbf{a}}_p^k \cdot \mathbf{x} - \mathbf{P}_{k-1}\Big)}\end{aligned} \tag{9}$$

where the output error $\mathbf{e}_k$ feedback the data $[\mathbf{P}_k = S_r^{-1}(L_r(\mathbf{e}_k)) \cdot (\hat{\mathbf{a}}_q^k)^{-1}, \hat{\mathbf{a}}_p^k \in \mathbb{R}^{m \times d}, \hat{b}_p^k \in \mathbb{R}]$. The k and $(k+1)$th subspace features can be obtained are $[\mathbf{H}_p^k = S_r(\mathbf{x}, \hat{\mathbf{a}}_p^k, \hat{b}_p^k)]$ and $[\mathbf{H}_p^{k+1} = S_r(\mathbf{x}, \hat{\mathbf{a}}_p^{k+1}, \hat{b}_p^{k+1})]$, respectively.

Step 4: Repeat steps 2 and 3 $l - 1$ times, the final subspace features $\{\mathbf{H}_p^1, \ldots, \mathbf{H}_p^l\}$ can be obtained. l is the number of subnetwork nodes.

E. Vigilance Estimation

The autoencoder mixed neurons encode essential brain signals and generate a stable feature representation of complex driver vigilance status. The specific content of vigilance estimation described as follows.

For the training samples, Moore–Penrose generalized inverse can be described as $X^T([V/I] + XX^T)^{-1} = X^{-1}$, the equation $\lim_{k\to\infty} ||t - L_r^{-1}(S_r(\hat{\mathbf{a}}_p^1 \cdot X + \hat{b}_p^1)) \cdot \hat{\beta}_p^1 + \cdots + L_r^{-1}(S_r(\hat{\mathbf{a}}_p^k \cdot X + \hat{b}_p^k)) \cdot \hat{\beta}_p^k|| \equiv 0$ holds when

$$\begin{aligned}\hat{\mathbf{a}}_p^k &= S^{-1}(L(e_{n-1})) \cdot X^T\left(\frac{V}{I} + XX^T\right)^{-1}, \hat{\mathbf{a}}_p^k \in \mathbb{R}^{m_r \times n_r}\\ \hat{b}_p^k &= \sum \frac{\Big(\hat{\mathbf{a}}_p^k \cdot X - S^{-1}(L(e_{m-1}))\Big)}{N}, \hat{b}_p^k \in \mathbb{R}\\ S^{-1} &= -\log\left(\frac{1}{x} - 1\right)\\ \beta_p^k &= \frac{\Big(e_{m-1}, L^{-1}\Big(q\Big(\hat{\mathbf{a}}_m^k \cdot X + \hat{b}_m^k\Big)\Big)\Big)}{\Big\|L^{-1}\Big(q\Big(\hat{\mathbf{a}}_m^k \cdot X + \hat{b}_m^k\Big)\Big)\Big\|^2}\end{aligned} \tag{10}$$

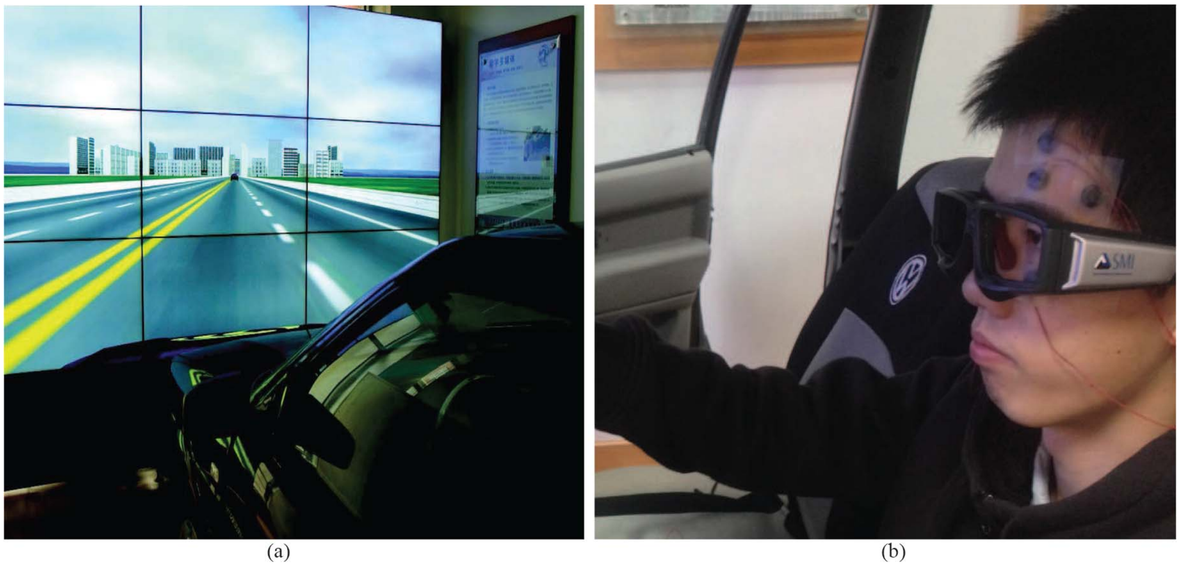


Fig. 2. Experimental setup. (a) Experimental environment. (b) Forehead electrode placement.

where β_p^k represents the output weight, and N represents the total number of the samples.

The output data are obtained by subspace feature extraction. The ranges of the continuous output data y "0−0.35," "0.36–0.70," and "0.71–1" indicate awake state, tired state, and drowsy state, respectively.

Then, the final quantitative analysis of the vigilance level is computed based on the root-mean-square error (RMSE) and the mean Pearson product moment correlation coefficient (COR). Generally, RMSE indicates the standard deviation between the observed values and predicted values, the formula of which is

$$\mathrm{RMSE}(x, y) = \sqrt{\frac{\sum_{t=1}^{n}(x_t - y_t)^2}{n}} \tag{11}$$

where $x = (x_1, x_2, \ldots, x_n)^T$ and $y = (y_1, y_2, \ldots, y_n)^T$ represent observed values and predicted values, respectively.

The parameter of COR can reach the relationship between the observed values and predicted values, the formula of which is

$$\mathrm{COR}(x, y) = \frac{\sum_{t=1}^{n}(x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum_{t=1}^{n}(x_t - \bar{x})^2 \sum_{t=1}^{n}(y_t - \bar{y})^2}} \tag{12}$$

where $\bar{x}$ and $\bar{y}$ indicate the mean of x and the mean of y, respectively. In general, the accuracy of the regression algorithm increases for lower values of the RMSE and higher values of the COR.

III. Experiments

A. Experimental Environment Setting

We used two different data sets from SEED-VIG. 23 subjects participated without the influence of all kinds of drugs, whose average age is almost 23 years. As we can see Fig. 2, there is an experimental vehicle without an engine system, the movement of which is controlled by software. Meanwhile, we use a large LCD screen contain a four-lane highway scene to simulate the real environment that is updated in real time. All participants did not receive any type of feedback while driving, even if they were asleep.

B. Alertness Annotations

So far, among the various ways to obtain alertness annotations, lane-departure is a popular method [32], [33]. Lin *et al.* [32] proposed the lane-departure events task that was introduced by an 8–12 s interval time, recording vehicle trajectory and the time of the lane-departure event, and defined the response time (RT) to reflect subjects' vigilance and arousal state. However, because it is based on the behavior of the subject, it cannot be considered feasible for dual tasks in the real world.

There is a most widely accepted way to get alertness annotations named the PERCLOS in [27], [28], and [34]. It is also proven that the PERCLOS index calculated by the eye-tracking-glasses-based approach [28] is more accurate than the facial video-based approach [27]. Thus, we use the SensoMotoric Instruments eye-tracking-glasses (SMI-ETG) with a window up to 120-Hz sampling rate, which provides a more accurate real-time reflection of eye movements, including blinks, glances, and fixed components. The formula is

$$\mathrm{PERCLOS} = \frac{\mathrm{blink} + \mathrm{CLOS}}{\mathrm{blink} + \mathrm{saccade} + \mathrm{fixation} + \mathrm{CLOS}} \tag{13}$$

where "CLOS" indicates the duration of the closed eye.

C. EOG Processing and Extraction

There are sufficient EOG signal extraction methods that have been comprehensively researched. The noise-free EOG signals can be obtained directly through placing electrodes near our eyes, but there are plenty of limitations in practical applications that do exist, notably, including the interferences with the subject's sight, other artifacts, intentional behavior of the individual, etc. Compared to the traditional method [25] that is difficultly extracted forehead vertical-EOG EOG_{FV} and

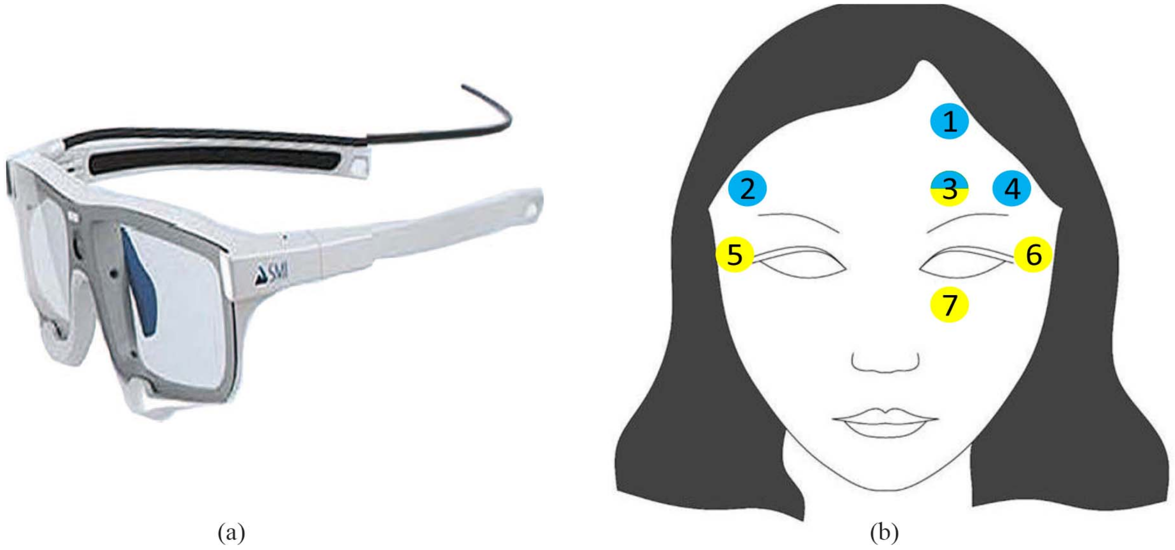


Fig. 3. PERCLOS and forehead electrodes position. (a) Eye tracking glasses. (b) EOG$_F$ and traditional EOG are collected by different electrodes placement.

forehead horizontal-EOG (EOG$_{FH}$) features from EOG difficultly, we designed a novel electrode placement to obtain the EOG$_F$ signals. The EOG$_{FH}$ and EOG$_{FV}$ can be separated effectively from the mixed EOG$_F$ signals [22]. Simultaneously, EOG$_{FH}$ and EOG$_{FV}$ contain eye movements, such as saccade, fixation, and blink. There are some differences in the way EOG$_F$ and traditional EOG are obtained by the electrode placements. As we can see Fig. 3(b), the blue dots (1, 2, 3, and 4) and the yellow dots (3, 5, 6, and 7) represent the forehead and traditional electrode positions, respectively.

We used two different approaches of fast independent component analysis (FASTICA) [35] and the minus (MIN) rule [36] to extract EOG$_{FV}$ and EOG$_{FH}$. For the ICA method, the approximation of EOG$_{FV}$ and EOG$_{FH}$ can be obtained from the electrodes No. 3 and No. 4 and No. 1 and No. 2, respectively. For the minus rule, the approximation of EOG$_{FV}$ and EOG$_{FH}$ can be obtained from the subtractions of electrodes No. 2 and No. 4 and No. 1 and No. 3, respectively.

After preprocessing, eye movements include saccade and blink components can be detected for EOG$_F$ by the wavelet transform algorithm. The wavelet coefficients with a scale of 8 are computed by the Mexican hat wavelet [37], which defined as follows:

$$\psi(x) = \frac{2}{\sqrt{3}}\pi^{-1/4}\left(1-x^2\right)e^{-(x^2/2)}. \quad (14)$$

D. Compared Methods and Experimental Results

The eye movements can be detected by the peak detection method on the wavelet coefficients. We encoded the peaks, such as negative peaks and positive peaks after using the thresholds to the continuous wavelet coefficients. For example, a blink has three successive peaks of negative, positive, and negative. Thus, a blinking movement can be encoded as “010.” All eye movements of the statistical parameters are listed in Table II.

ICA$_{FH}$, ICA$_{FV}$, MINUS$_{FH}$, and MINUS$_{FV}$ features are extracted from the EOG$_F$ signal using ICA and MIN separation approaches. Besides that, we compare the proposed study to several recent state-of-the-art signal recognition methods, such as extreme learning machine (ELM), B-ELM, and double-layered neural network with subnetwork nodes (DNN$_{SN}$). We will then introduce the above-mentioned methods shortly and the parameter settings are listed in Table III.

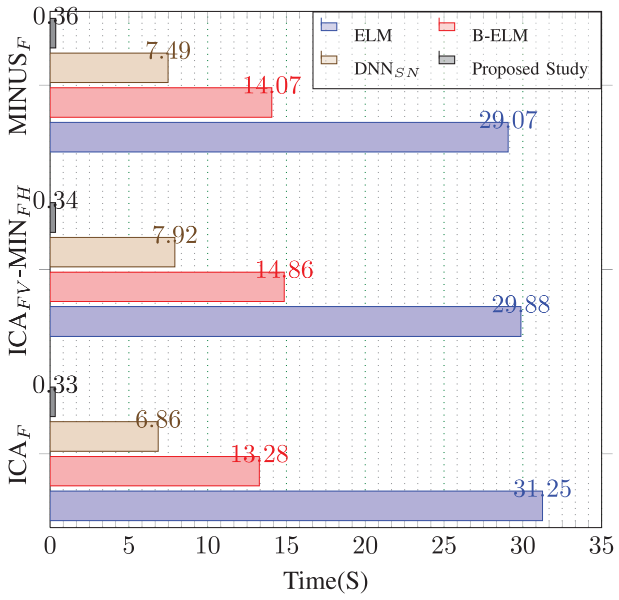


Fig. 4. Time to extract features from the raw signal and display the detections.

TABLE II
ENCODE THE EOG FEATURE

Components	Sequence	Encode
Negative peak	one	0
Positive peak	one	1
Blink candidate	three successive	010
Saccade candidate	two successive	01 or 10

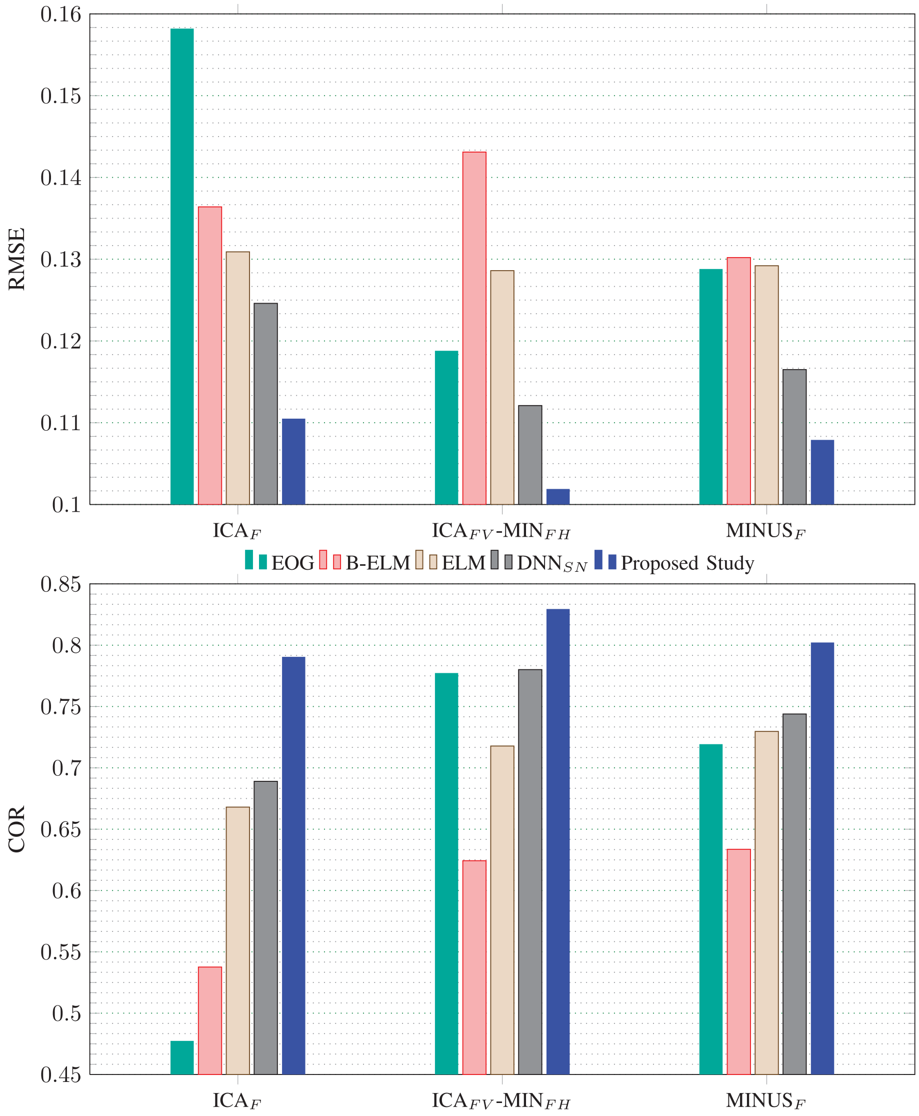


Fig. 5. Performance of single modality.

TABLE III
SYMBOLS USED 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.
DNN$_{SN}$	$C_1 = 2^{[-10,\cdots,10]}$, and $C_2 = 2^{[-10,\cdots,10]}$; Three subnetwork neurons are used, each of which contains 500 hidden neurons.
Proposed study	$C = 2^{[-5,\cdots,5]}$, and $U = 2^{[-5,\cdots,5]}$, $V = 2^{[-5,\cdots,5]}$; $l = 3$, and each subnetwork node contains 400 hidden neurons.

1) *ELM:* Huang *et al.* [38] proposed ELM which not only achieves smaller training error but also has a smaller output weight norm. Meanwhile, ELM works with an extensive feature mapping without iterative tuning, which only requires lower computational complexity and it can be applied in classification and regression applications directly.

2) *B-ELM:* Compared to ELM, Yang *et al.* [31] proposed that B-ELM can select the optimal size of the single-hidden layer feedforward network (SLFN) to set the optimum number of hidden nodes, thereby reducing training time and improving learning efficiency greatly.

3) *DNN$_{SN}$:* According to our previous study [39], the proposed DNN$_{SN}$ model can be directly applied to all physiological signals of single modality and multimodality, and each subnetwork neuron has the ability to feature learning and feature selection. As a result, the accuracy of the regression prediction is significantly improved.

All experimental results are listed in Table IV. ELM was originally proposed for "generalized" SLFN, which provides good generalization ability. The ELM-based method works reasonably and achieves a good RMSE/COR of 0.13/0.67, 0.13/0.72, and 0.13/0.73, respectively. Through its features extracted and combined from hierarchical network layers, the DNN$_{SN}$-based method showed good results of 0.13/0.68, 0.11/0.78, and 0.12/0.74, respectively, which can be also

TABLE IV
EXPERIMENTAL RESULTS OF COMPARED METHODS. THE BEST RESULTS ARE BOLDED

Methods	RMSE-Mean	RMSE-STD	COR-Mean	COR-STD	Time(S)
ICA$_F$	0.1582	0.0844	0.4774	0.5381	–
ELM	0.1309	0.0486	0.6680	0.1957	31.25
B-ELM	0.1364	0.0683	0.5376	0.1906	13.28
DNN$_{SN}$	0.1246	0.0540	0.6890	0.2041	6.862
Proposed Study	**0.1105**	**0.0444**	**0.7905**	**0.1609**	**0.3281**
ICA$_{FV}$-MIN$_{FH}$	0.1188	0.0391	0.7773	0.2352	–
ELM	0.1286	0.0557	0.7178	0.2100	29.88
B-ELM	0.1431	0.0584	0.6243	0.1904	14.86
DNN$_{SN}$	0.1121	0.0540	0.7801	0.2041	7.918
Proposed Study	**0.1019**	0.0413	**0.8295**	**0.1217**	**0.3438**
MINUS$_F$	0.1288	0.0588	0.7193	0.3492	–
ELM	0.1292	0.0489	0.7297	0.2225	29.07
B-ELM	0.1302	0.0596	0.6336	0.1984	14.07
DNN$_{SN}$	0.1165	0.0508	0.7439	0.1819	7.488
Proposed Study	**0.1079**	**0.0378**	**0.8022**	**0.1306**	**0.3594**

considered as an effective way. In the autoencoder model, using the reduced feature dimension and extracted feature, the time taken from the original signal to the display detection is approximately 34 ms, which is far lower than other comparison methods (see Fig. 4), and the driver's alert level can be monitored fast. Furthermore, due to the subspace feature is further extracted by subnetwork nodes, the performance of the proposed method improved to 0.11/0.79, 0.10/0.83, and 0.11/0.80, respectively. Meanwhile, the proposed model also obtains the lowest RMSE and the highest COR. Compared to the results of other regression models using every EOG feature in Fig. 5, the profit of our strategies is apparent. This reflects that the impulses components from EOG$_{FV}$ and saccade components from EOG$_{FH}$ can be detected easily and accurately.

IV. CONCLUSION

This article proposed a novel multilayer network structure that includes an autoencoder layer for dimension reduction and regression for single-model vigilance estimation. Compared with other single model methods, the proposed method achieves higher learning accuracy. It demonstrated that our experimental algorithm has a better performance in detecting important eye components from EOG$_F$, including saccade, blink, and fixation. In addition, the total training and test time is much lower than other comparison methods. The driver's alertness can be monitored fast by the proposed method, which also surmounts other state-of-the-art single model techniques.

Because of research funding and time constraints, all subjects are recruited from undergraduate and postgraduate students of college campuses, thus the average age is about 23 years old. It is noticed that the EEG signal differs from person to person. Therefore, the experimental results may change if the age of the subject increases. As for how much age will impact on this model should be concluded from actual experimental data. We planned to address this issue in our future works. We should consider a larger age range to arrive at a more reliable conclusion. Simultaneously, EEG is one of the commonly used physiological signals in the field of affective computing. We should combine EEG and EOG in our future research and apply the multimodal emotion recognition applications based on DL algorithms.

All subjects signed written informed consent. All subjects gave written informed consent before participation. The study was approved by the local ethics committee.

REFERENCES

[1] A. L. Samuel, "Some studies in machine learning using the game of checkers," *IBM J. Res. Develop.*, vol. 3, no. 3, pp. 210–229, Jul. 1959.

[2] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain," *Psychol. Rev.*, vol. 65, no. 6, pp. 65–386, 1958.

[3] L. Qi, W. Dou, W. Wang, G. Li, H. Yu, and S. Wan, "Dynamic mobile crowdsourcing selection for electricity load forecasting," *IEEE Access*, vol. 6, pp. 46926–46937, 2018.

[4] L. Qi *et al.*, "An exception handling approach for privacy-preserving service recommendation failure in a cloud environment," *Sensors*, vol. 18, no. 7, 2018, Art. no. 2037.

[5] W. Gong, L. Qi, and Y. Xu, "Privacy-aware multidimensional mobile service quality prediction and recommendation in distributed fog environment," *Wireless Commun. Mobile Comput.*, vol. 2018, Mar. 2018, Art. no. 3075849.

[6] G. Hinton *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, Nov. 2012.

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778.

[8] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[9] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10 000 classes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 1891–1898.

[10] S. Levine, P. Pastor, A. Krizhevsky, J. Ibarz, and D. Quillen, "Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection," *Int. J. Robot. Res.*, vol. 37, nos. 4–5, pp. 421–436, 2018.

[11] H. Greenspan, B. van Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1153–1159, May 2016.

[12] J. Rust, "Using randomization to break the curse of dimensionality," *Econometrica J. Econometr. Soc.*, vol. 65, no. 3, pp. 487–516, 1997.

[13] L. Qi, X. Xu, W. Dou, J. Yu, Z. Zhou, and X. Zhang, "Time-aware IoE service recommendation on sparse data," *Mobile Inf. Syst.*, vol. 2016, Nov. 2016, Art. no. 4397061.

[14] L. Qi, R. Wang, C. Hu, S. Li, Q. He, and X. Xu, "Time-aware distributed service recommendation with privacy-preservation," *Inf. Sci.*, vol. 480, pp. 354–364, Apr. 2019.

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

[16] Y. Yang, Q. M. 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.

[17] Y. Yang and Q. M. 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.

[18] P. Thiffault, *Addressing Human Factors in the Motor Carrier Industry in Canada*. Ottawa, ON, Canada: Can. Council Motor Transport Administrators, 2011.

[19] 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.

[20] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neur. Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.

[21] Y.-F. Zhang, X.-Y. Gao, J.-Y. Zhu, W.-L. Zheng, and B.-L. Lu, "A novel approach to driving fatigue detection using forehead EOG," in *Proc. 7th Int. IEEE/EMBS Conf. Neur. Eng.*, 2015, pp. 707–710.

[22] 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. Neur. Eng.*, 2015, pp. 767–770.

[23] W.-L. Zheng *et al.*, "Vigilance estimation using a wearable EOG device in real driving environment," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 1, pp. 170–184, Jan. 2020.

[24] C.-H. Chuang, Y.-P. Lin, L.-W. Ko, T.-P. Jung, and C.-T. Lin, "Automatic design for independent component analysis based brain-computer interfacing," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, 2013, pp. 2180–2183.

[25] A. Bulling, J. A. Ward, H. Gellersen, and G. Tröster, "Robust recognition of reading activity in transit using wearable electrooculography," in *Proc. Int. Conf. Pervasive Comput.*, 2008, pp. 19–37.

[26] D. Sommer and M. Golz, "Evaluation of perclos based current fatigue monitoring technologies," in *Proc. Int. Conf. IEEE Eng. Med. Biol.*, 2010, pp. 4456–4459.

[27] Y. Dong, Z. Hu, K. Uchimura, and N. Murayama, "Driver inattention monitoring system for intelligent vehicles: A review," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 2, pp. 596–614, Jun. 2011.

[28] J.-X. Ma, L.-C. Shi, and B.-L. Lu, "An EOG-based vigilance estimation method applied for driver fatigue detection," *Neurosci. Biomed. Eng.*, vol. 2, no. 1, pp. 41–51, 2014.

[29] C. Hong, J. Yu, J. Wan, D. Tao, and M. Wang, "Multimodal deep autoencoder for human pose recovery," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5659–5670, Dec. 2015.

[30] Z. Zhu, X. Wang, S. Bai, C. Yao, and X. Bai, "Deep learning representation using autoencoder for 3D shape retrieval," *Neurocomputing*, vol. 204, pp. 41–50, Sep. 2016.

[31] 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.

[32] C.-T. Lin, K.-C. Huang, C.-H. Chuang, L.-W. Ko, and T.-P. Jung, "Can arousing feedback rectify lapses in driving? prediction from EEG power spectra," *J. Neur. Eng.*, vol. 10, no. 5, 2013, Art. no. 056024.

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

[34] D. F. Dinges and R. Grace, "Perclos: A valid psychophysiological measure of alertness as assessed by psychomotor vigilance," U.S. Dept. of Transp., Federal Highw. Admin., Washington, DC, USA, Rep. FHWA-MCRT-98–006, 1998.

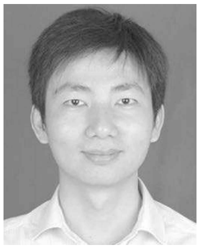
[35] C.-T. Lin *et al.*, "Adaptive EEG-based alertness estimation system by using ICA-based fuzzy neural networks," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 53, no. 11, pp. 2469–2476, Nov. 2006.

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

[37] Z. Masood, K. Majeed, R. Samar, and M. A. Z. Raja, "Design of mexican hat wavelet neural networks for solving Bratu type nonlinear systems," *Neurocomputing*, vol. 221, pp. 1–14, Jan. 2017.

[38] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 2, pp. 513–529, Apr. 2012.

[39] W. Wu *et al.*, "A regression method with subnetwork neurons for vigilance estimation using EOG and EEG," *IEEE Trans. Cognitive Dev. Syst.*, to be published.

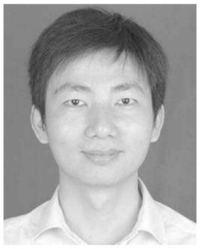


Wei Wu (Student Member, IEEE) received the M.S. degree in power electronics and power drives from Hunan University of Technology, Zhuzhou, China, in 2011, and the first Ph.D. degree from the Department of Electrical and Computer Engineering, University of Windsor, Windsor, ON, Canada, in 2019. He is currently pursuing the second Ph.D. degree in electrical engineering with the College of Electrical and Information Engineering, Hunan University, Changsha, China.

He was a Lecturer with the College of Electrical and Information Engineering, Hunan University of Technology from April 2008 to November 2016. He was awarded a scholarship under the State Scholarship Fund by the China Scholarship Council for two years. His research focuses on robotics, artificial neural networks, and affective computing.

Wei Sun received the B.S., M.S., and Ph.D. degrees from the Department of Control Science and Engineering, 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 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.



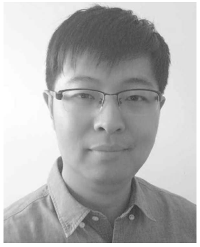
Q. M. Jonathan Wu (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the University of Wales, Swansea, U.K., in 1990.

He was affiliated with the National Research Council of Canada for ten years beginning in 1995, 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 is a Visiting Professor with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China. He has published 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 is an Associate Editor of the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS and *Cognitive Computation*. He has served on technical program committees and international advisory committees for many prestigious conferences.

Cheng Zhang received the B.S. and M.S. degrees in computer science from Central South University, Changsha, China, in 2004 and 2012, respectively.

In 2004, she joined the College of Electrical and Information Engineering, Hunan University of Technology, Zhuzhou, China, where she is currently working as a Lecturer with the College of Electrical and Information Engineering. Her research interests include artificial neural networks and robotics.



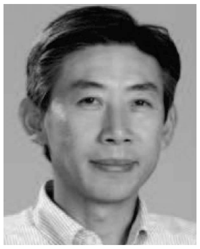
Yimin Yang (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the College of Electrical and Information Engineering, Hunan University, Changsha, China, in 2013.

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 Department of Computer Science, Lakehead University, Thunder Bay, ON, Canada. His research interests are 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 has been serving as a reviewer for international journals of his research field, a guest editor of multiple journals, and a program committee member of some international conferences.

Hongshan Yu received the B.S., M.S., and Ph.D. degrees from Hunan University, Changsha, China, in 2001, 2004, and 2007, respectively.

He is a Professor with the College of Electrical and Information Engineering, Hunan University. His research interests include mobile robot navigation and computer vision.



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, Xian, 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, Qingdao, China, 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 Professor with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China, where he has been an Adjunct Professor with the Laboratory for Computational Biology, Shanghai Center for Systems Biomedicine since 2005. His current research interests include brain-like computing, neural network, machine learning, computer vision, bioinformatics, brain–computer interface, and affective computing.

Dr. Lu 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 DEVELOPMENTAL SYSTEMS and *Neural Networks*, and a Board Member of APNNA.