Neurocomputing 283 (2018) 266–281

Contents lists available at ScienceDirect

Neurocomputing

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

Task-generic mental fatigue recognition based on neurophysiological signals and dynamical deep extreme learning machine

Zhong Yin[a,*], Jianhua Zhang[b]

[a] Engineering Research Center of Optical Instrument and System, Ministry of Education, Shanghai Key Lab of Modern Optical System, University of Shanghai for Science and Technology, Jungong Road 516, Yangpu District, Shanghai 200093, PR China
[b] Department of Automation, East China University of Science and Technology, Shanghai 200237, PR China

ARTICLE INFO

Article history:
Received 11 September 2017
Revised 26 December 2017
Accepted 30 December 2017

Communicated by Yudong Zhang

Keywords:
Mental fatigue
Human–machine system
Electroencephalography
Extreme learning machine
Deep learning

ABSTRACT

The electroencephalography (EEG) based machine-learning model for mental fatigue recognition can evaluate the reliability of the human operator performance. The task-generic model is particularly important since the time cost for preparing the task-specific training EEG dataset is avoid. This study develops a novel mental fatigue classifier, dynamical deep extreme learning machine (DD-ELM), to adapt the variation of the EEG feature distributions across two mental tasks. Different from the static deep learning approaches, DD-ELM iteratively updates the shallow weights at multiple time steps during the testing stage. The proposed method incorporates the both of the merits from the deep network for EEG feature abstraction and the ELM autoencoder for fast weight recompuation. The feasibility of the DD-ELM is validated by investigating EEG datasets recorded under two paradigms of AutoCAMS human–machine tasks. The accuracy comparison indicates the new classifier significantly outperforms several state-of-the-art mental fatigue estimators. By examining the CPU time, the computational burden of the DD-ELM is also acceptable for high-dimensional EEG features.

1. Introduction

In safety-critical working environments required human–machine collaboration, human operator is becoming a crucial factor that affects the task performance. To evaluate human–machine interaction efficiency, operator functional state (OFS) has been proposed to characterize human capability for prolonged monitoring and temporal strategy generation [1]. The abnormal OFS may cause instantaneous human performance degradation and increase the possibility of severe accidents in many public services such as driving, aviation, health care, and manufacturing [2]. According to well-documented studies [3–5], a group of variables, which are known as mental workload [6], situation awareness [7], and mental fatigue [8], quantify different aspects of OFS. In particular, mental fatigue is identified as a crucial dimension, which is defined by a cumulative process to the disinclination of the effort and drossiness.

Different from physical fatigue induced by the impaired muscular strength, mental fatigue is associated with the psychological responses indicated by the feelings of weariness and inhibition for continue performing the task [9]. Monotonous, long-hour working environments contribute to the gradual increase of the mental fatigue. Numerous studies reported it was a major human factor to the high risk of the operation errors [10–12]. The counter measure can be achieved by the adaptive automation theory, where a human-centered system is implemented to detect and regulate the mental fatigue by reallocating the functionalities between the operator and the machine [13].

The reliability of such adaptive systems is based on the success recognition of mental fatigue levels. The corresponding indicators mainly include secondary task performance, subjective ratings, and electrophysiological measurements [11]. Neurophysiological signals received increasing attention since they are closely related to the functions of responsible cortical or subcortical networks such as basal forebrain, thalamocortical neurons, and locus coerulus [9]. Among them, electroencephalography (EEG) generated by the summation of the postsynaptic potentials from a wide range of cortical neurons can be quite repeatable and predictive [14]. Besides, the EEG signals can be easily recorded via portable and wireless sensors owning to the progress of the biomedical engineering devices.

1.1. Related works

The relationship between the spectral EEG features and the mental fatigue was revealed in early studies. Grandjean indicated EEG theta (4–7 Hz) rhythm reflected the decrease of the alertness

* Corresponding author.
E-mail address: yinzhong@usst.edu.cn (Z. Yin).

https://doi.org/10.1016/j.neucom.2017.12.062
0925-2312/

during drowsiness and the fatigue [15]. Okogbaa et al. reported the energy of the alpha rhythm (8–13 Hz) increased when the impairment of the sustained attention arose [16]. Markand suggested the alpha rhythm might temporally attenuate and reappear for a few minutes in high fatigue state [17]. Lal and Craig found that the spatial distribution variation of the occipital and parietal alpha wave was the useful indicator [18]. Makeig and Jung reported both of the alpha and theta rhythms were related to the variation of the fatigue level [19]. In recent studies, EEG features of higher sensitivity were abstracted from the conventional spectral power. Simon et al. proposed that the alpha spindle was a more objective clue, which measured the short bursts of the alpha waves [20]. Papadelis et al. employed the alpha relative band ratio in the central and parietal as the mental fatigue indicator [21].

Since various EEG features have been extensively analyzed in the literature, the pattern classifier constructed via machine learning principles were used to further extract and fuse the useful information hidden in those multidimensional variables. Such classifier derives a predictive model for instantaneous mental fatigue estimation. Among these works, the shallow structure based pattern classifiers are employed. Aleksandra et al. extracted cross-spectral density features from EEG and employed learning vector quantization based artificial neural network (ANN) for mental fatigue recognition [22]. Kiymik et al. combined discrete wavelet transformation and ANN classifier to improve the classification performance [23]. Yildiz et al. computed Shannon entropy features from EEG and built an adaptive neural fuzzy inference system to distinguish binary mental fatigue levels [24]. Mervyn et al. used support vector machine (SVM) model to process EEG power generated by fast Fourier transformation [25] and improved the binary fatigue classification rate to 99.3%.

The applications of deep learning methods on neuroimaging datasets have received attention [26] in recent reported works. The deep learning approaches are applied to abstract the high-level EEG features and to build more robust mental fatigue estimators [27]. Different from shallow classifiers such as SVM and single-hidden-layer ANN, deep learning model generates more compact feature abstractions via multilayered, hierarchical networks [28]. In the newest studies, Hajinoroozi et al. developed a channel-wise deep convolutional neural networks (CNN) to predict the driver mental fatigue based on EEG signals [27]. They generalized the conventional CNN model structure by introducing restricted Boltzmann machine between each two hidden layers. In our previous work, we proposed switching deep belief networks (DBN) to cope with the individual differences in EEG signals and constructed a cross-subject mental fatigue classifier [29]. These works exploited the merits of the deep learning in intelligent feature engineering, where the time and frequency domain EEG features were automatically combined to more salient mental fatigue indicators.

1.2. Motivation of the present study

According to the brief literature review, we note that most of the existing studies on mental fatigue recognition were task specific. That is, the classifier was trained and validated via the EEG datasets recorded from exactly the same human–machine task. However, the pattern of the mental fatigue stimuli for EEG signals cannot be always the same in real-world environment, e.g. the speed of the driver fatigue accumulation could vary on different traffic status. Therefore, it is crucial to improve the generalization capability of a task-generic, mental-fatigue classifier constructed by the EEG data from a typical task and tested on different one. Such cross-task OFS recognition issue on the dimension of mental workload has been investigated via shallow learning approaches [30,31]. These works showed the huge volume of historical EEG data could be transferred and adaptively used for predicting novel mental stimuli via proper machine learning approaches.

Since the task-generic mental fatigue recognition received much less attention in the literature, this study attempts to address this issue by employing the deep classifier. The critical motivation behind is that task-generic classifier facilitates sufficient and transferable EEG data in the training stage, which has the capability to prevent the overfitting from the deep network with numerous parameters for tuning. It can be thus expected to maximally exploit the advantage of the deep model for high efficiency of feature abstraction or feature denoising.

To achieve the transferability of the EEG data across different fatigue stimuli, the hierarchical extreme learning machine (H-ELM) has been employed as the basis for the task-generic classifier [32]. Different from the stacked autoencoder (SAE) and DBN based deep learning primitives, H-ELM is generalized from ELM, a high-speed learning machine proposed by Huang et al., which does not require the gradient-based optimization for both of the pre-training and fine-tuning on network weights [33,34]. Such mechanism significantly reduces the time cost for training and further facilitates the real-time weight updating to tackle the distribution variation of the novel EEG data. Inspired by this, we specifically design a dynamical deep extreme learning machine (DD-ELM) for coping with the task-generic mental fatigue recognition issue. The essential of the DD-ELM is to dynamically recompute part of the weights of the deep ELM by using the unsupervised transfer learning principle. The values of the weights at current time step is determined based on the historical EEG feature abstractions. We examine the performance of the DD-ELM by two well-controlled experimental paradigms via Auto CAMS platform with different mental fatigue stimuli.

The application of the ELM is very promising for the data-driven modeling because of its universal approximation capability. The related theory was proofed by Huang et al. [35]. In detail, the ELM network is a universal approximator when users simply choose randomly neuron parameters in the hidden layer where the activation function can be any type of bounded non-constant piecewise continuous functions. It ensures the proposed mental fatigue classifier has the capability to fit the complex distribution of the psycho-physiological data. Moreover, the classification capability of ELM with high generalizability has been also shown in [34], where the maximal margin optimization has been introduced with the regularization term for the output weight vector of the ELM model. To this end, the ELM shares the merits of the SVM classifier since both of the two methods minimizes the norm of the weight vector to control the over-fitting.

In particular, the ELM approach has been shown as an efficient tool to analyze the neuroimaging data in recent studies. In [36], Lu et al. proposed a hybrid model that combines the 2D discrete wavelet transformation and the ELM optimized by the bat algorithm to classify the magnetic resonance data. In their work, the ELM was used as an auto-detector for pathological diseases of the brain functionality. By incorporating the 10-fold cross validation, 98.33% overall classification accuracy was achieved on 132 instances. In [37], Zhang et al. developed a framework for smart pathological brain detection by applying a synthetic minority oversampling, wavelet packet Tsallis entropy, extreme learning machine, and Jaya algorithm. Their work showed competitive classification rate of $99.57 \pm 0.57\%$. Both of the two works strongly supported the classification capability of the ELM model.

The rest of the paper is organized as follows. The paradigms of two different human–machine tasks and the derived EEG datasets are described in Section 2. In Section 3, the DD-ELM algorithms are presented in detail. Section 4 summarizes the task-generic classification results across DD-ELM and several state-of-the-art fatigue estimators. The merits and the limitations of the DD-ELM are

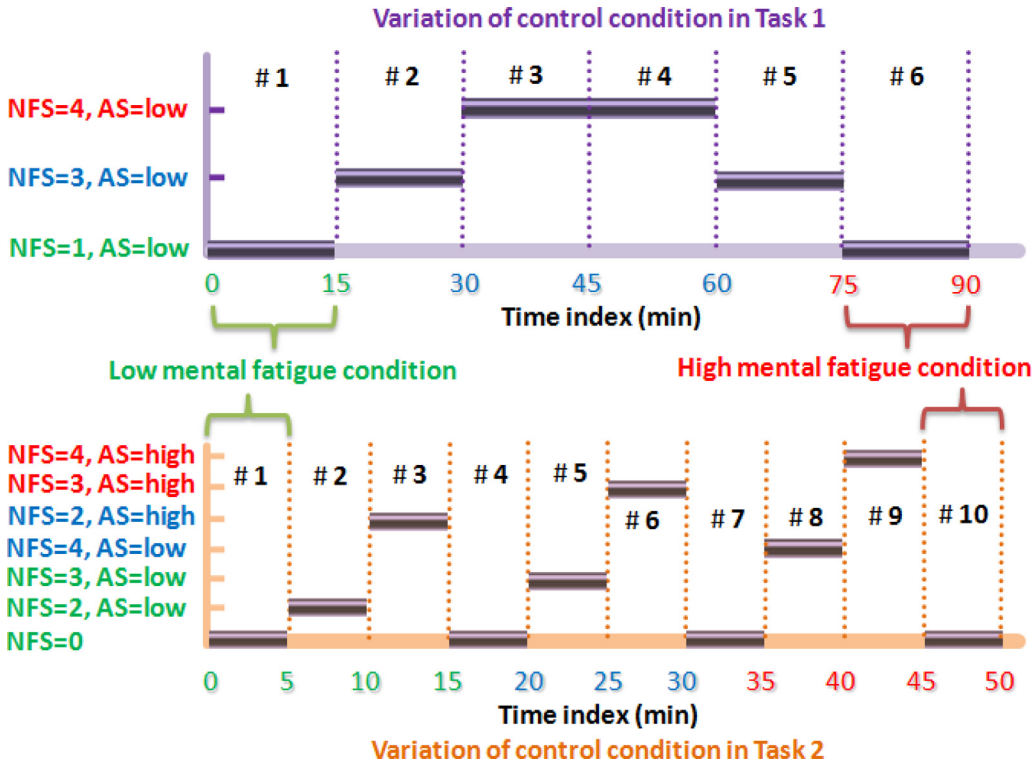

Fig. 1. Change of the control condition in a session of Tasks 1 and 2. The binary levels of the mental fatigue are defined by the first and the last conditions, respectively.

discussed in Section 5. Finally, Section 6 concludes the contribution of the study.

2. Experimental settings and EEG databases

2.1. Experimental paradigms of two mental tasks

In this study, we use two EEG databases built in our previous experiments [38,39] (denoted as Tasks 1 and 2, respectively), where automation-enhanced cabin air management system (AutoCAMS) was employed to facilitate real-time human–machine interaction [40]. AutoCAMS provided a safety-critical process control tasks in a micro-world, where the operator was required to monitor and regulate four subsystems of the air quality in a space cabin. For each subsystem, the actuator sensitivity (*AS*) could be set to "low" or "high" indicating that an easy or difficult mode for manual operation was activated. By predefining the number of failed subsystems (*NFS*) and *AS* modes, different levels of the task demand were programmed, which resulted in different speed for mental fatigue accumulation.

Eight and six healthy volunteers (21–24 years, all male, right-handed) participated Tasks 1 and 2, respectively. The participants of all tasks were chosen from on-campus postgraduate students with informed consent. Before the experiment began, they were trained with AutoCAMS operation skills for more than 10 h to reduce learning effect. Each participant was instructed to perform two sessions of operations with the same settings of task condition as shown in Fig. 1.

In a session of Task 1, the mode of the AS was fixed to "low" while the value of the *NFS* varied across six 15-min control conditions. Low (#1 and #6), medium (#2 and #5), and high (#3 and #4) task-demand conditions were predefined by setting *NFS* as 1, 3, and 4, respectively. In a session of Task 2, the AS model can be switched between "low" and "high" in ten 5-min control conditions. Different from Task 1, the baseline conditions #1, #4, #7, and #10 were introduced without manual operations for participants. In low-*AS* conditions #2, #5, and #8, the *NFS* was set to 2, 3, and 4, respectively. The same paradigm of *NFS* was also applied in high-*AS* conditions #3, #6, and #9. In summary, Task 1 followed a cyclical paradigm of the task demand and the mental fatigue can be recovered in the end of the session. On the other hand, Task 2 employed a monotonously increased task complexity while the baseline conditions could help to reduce the fatigue accumulation. Based on these two different paradigms, the low and high mental fatigue states are defined according to the first and the last condition of each session.

2.2. Acquisition and preprocessing of EEG signals

When participants were performing the task session based on the paradigm defined in Task 1 or 2, EEG signals were simultaneously recorded via 11 and 15 channels according to the international 10–20 system, respectively. To ensure extracting exactly the same EEG features, we selected the shared eleven channels, F3, F4, Fz, C3, C4, Cz, P3, P4, Pz, O1, and O2, across two tasks with the sampling frequency of 500 Hz. The details of the data acquisition can be found in our previous works [38,39].

The raw EEG signals were preprocessed by using a 4-order Butterworth filter via the band-pass, cutoff frequencies of 0.5 and 40 Hz. We further applied independent component analysis to remove the ocular and scalp-muscular noise from the filtered signals. For each session, the EEG in the first and the last control conditions were selected and split into 2-s, non-overlapping segments. For Task 1, 450 EEG segments were prepared for each 15-min control condition. For Task 2, the first 10-s signals in each condition were removed because of the severe muscular noise [39]. Then, 145 EEG segments were prepared for each 290-s control condition.

For all segments, the feature selection was carried out as shown in Fig. 2. For a channel data of one segment, the average power in

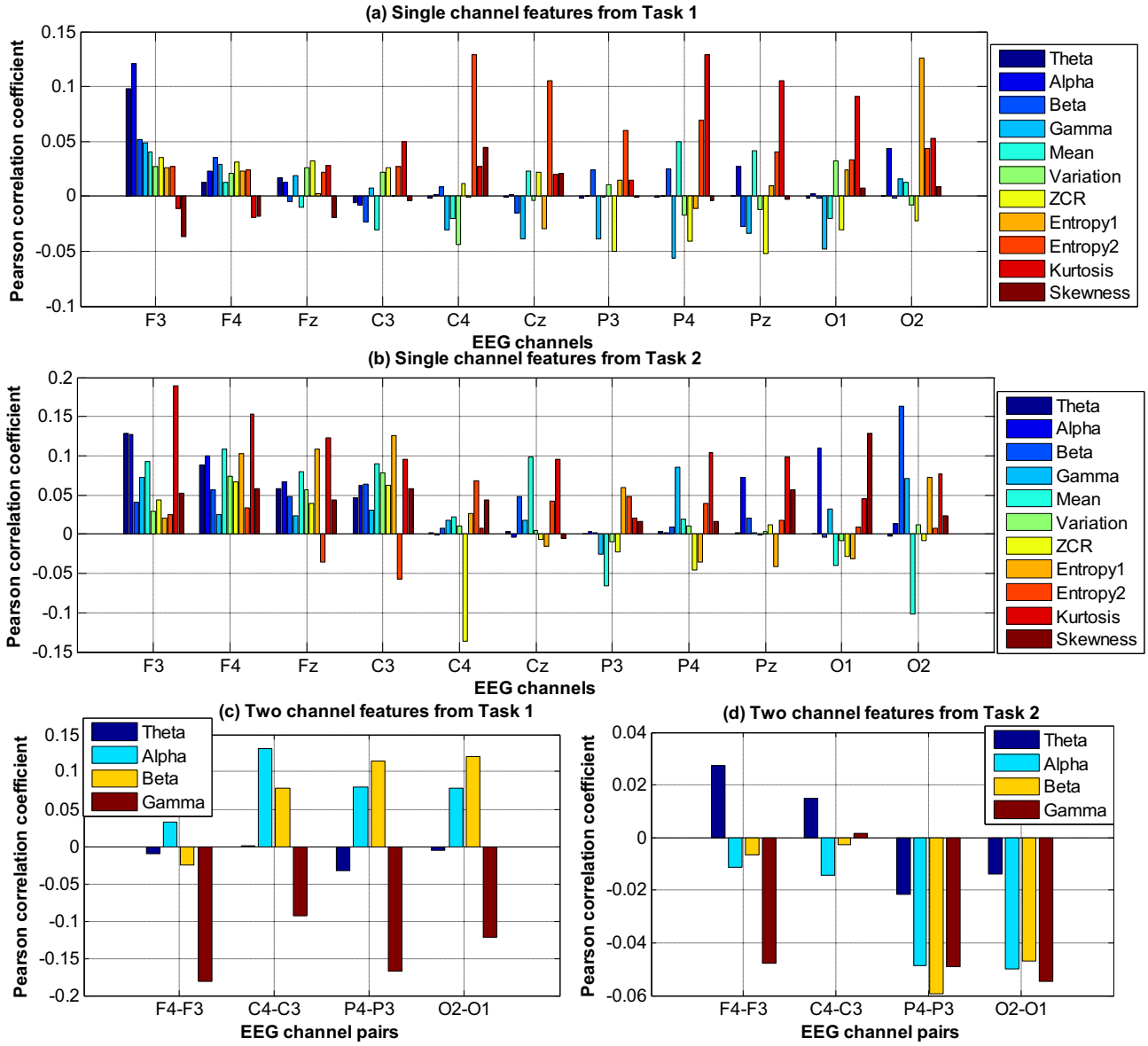

Fig. 2. Pearson correlation analysis on the time courses of the extracted EEG signals and the target mental fatigue labels for Tasks 1 and 2. In each column of data, the coefficient value is averaged across all participants for each task.

theta, alpha, beta (13–30 Hz), and gamma (30–40 Hz) bands, mean, variance, zero-crossing rate (ZCR), Shannon entropy (denoted as Entropy 1), log-energy entropy (denoted as Entropy 2), kurtosis, and skewness were computed. That is, $11 \times 11 = 121$ single channel features were calculated. Besides, the differences of the theta, alpha, beta and gamma power between F4-F3, P4-P3, C4-C3, and O2-O1 channel pairs were extracted. In total, $121 + 16 = 137$ EEG features were prepared for each segment. The participant-average Pearson correlation coefficients between these features and the target mental fatigue states are also shown in the figure. In Fig. 2(a) and (b), the positive correlation of the frontal theta and alpha power, log-energy entropy in parietal channels is shared in two tasks. The consistent negative correlation of the F4-F3, P4-P3, and O2-O1 power differences are found in Fig. 2(c) and (d). These facts indicate the common properties exist in the EEG data distribution across two tasks.

3. Method

The ELM based deep learning classifier is employed to discover salient EEG feature combinations for recognizing binary mental fatigue levels. In this section, the standard H-ELM method is first briefly reviewed. Then, we present the architecture and the details of learning algorithms of the DD-ELM specifically designed as a task-generic mental fatigue classifier.

3.1. Hierarchical ELM

The standard ELM network is initialized via a single-hidden-layer ANN f_S [33]. Given the input EEG feature vector $\mathbf{x} \in R^D$, the f_S can be formularized as follows.

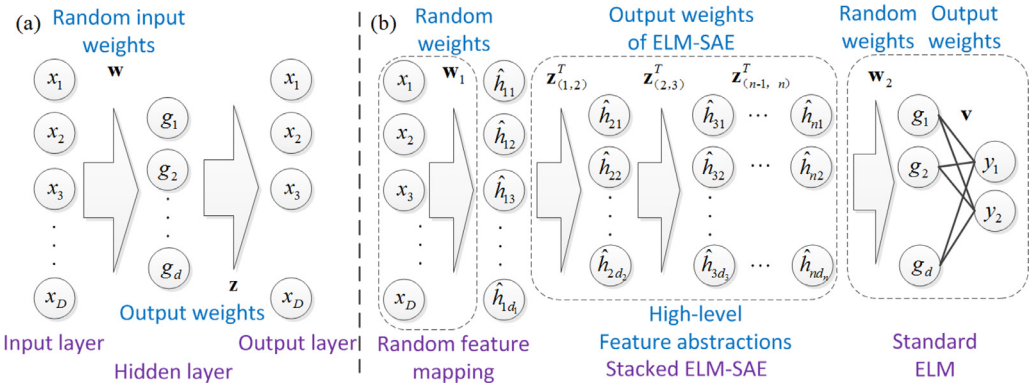

Fig. 3. Network architectures of (a) an ELM-SAE and (b) an H-ELM.

$$f_S(\mathbf{x}) = \sum_{i=1}^{d} v_i \cdot g_i(\mathbf{w}_i \cdot \mathbf{x} + b_i). \tag{1}$$

In the equation, $\mathbf{w}_i$ and b_i are the input weights and bias of the ith hidden node with random values. The terms g_i, d and v_i denote the activation function, number of the hidden neurons and output weight, respectively.

Training an ELM network is equivalent to minimizing the norm of the output weights $\mathbf{v}$ as well as the fitting error.

$$\min_{\mathbf{v}} \Phi(\mathbf{v}) = \|\mathbf{v}\|_2^2 + C\|\mathbf{H}\mathbf{v} - \mathbf{Y}\|_2^2. \tag{2}$$

Given N training instances with each denoted by $\mathbf{x}_j$, $\mathbf{H}$ is a $N \times d$ matrix, where the entry of jth row and ith column is $g_i(\mathbf{w}_i \cdot \mathbf{x}_j + b_i)$. The matrix $\mathbf{Y}$ is with the size of $N \times K$ and each row denotes a training label of $\mathbf{x}_j$ with K target classes. The regularization parameter defined by C is used to balance the minimum of the norm and the fitting error.

According to [41], Eq. (2) could be solved as follows,

$$\mathbf{v} = \mathbf{H}^*\mathbf{Y}, \tag{3}$$

where $\mathbf{H}^*$ is the Moore–Penrose generalized inverse of $\mathbf{H}$ regularized by C. The computation of $\mathbf{H}^*$ can be achieved by,

$$\mathbf{H}^* = \mathbf{H}^T\left(\frac{1}{C} + \mathbf{H}\mathbf{H}^T\right)^{-1}. \tag{4}$$

By substituting Eqs. (3) and (4) into Eq. (1), the learned model of the ELM is,

$$\hat{\mathbf{y}} = f_S(\mathbf{x}) = h(\mathbf{x})\mathbf{H}^T\left(\frac{1}{C} + \mathbf{H}\mathbf{H}^T\right)^{-1}\mathbf{Y}, \tag{5}$$

with $h(\mathbf{x})$=$G(\mathbf{w}_i \cdot \mathbf{x}_j + b_i)$ denoting the hidden neuron activation.

Based on the standard ELM, an ELM sparse autoencoder (ELM-SAE) was proposed by Huang et al. [34] for feature representation as shown in Fig. 3(a). In particular, the training target $\mathbf{Y}$ is exactly as same as the input features, $\mathbf{Y} = \mathbf{X}$. According to the l_1 optimization and the fast iterative shrinkage-thresholding algorithm [34], the output weights $\mathbf{z}$ of ELM-SAE is calculated by,

$$\min_{\mathbf{z}} \Phi(\mathbf{z}) = \|\mathbf{z}\|^2 + C\left\|\tilde{\mathbf{H}}\mathbf{z} - \mathbf{X}\right\|_{l_1}. \tag{6}$$

Similar to Eq. (2), $\tilde{\mathbf{H}}$ is the matrix of the hidden neuron activations derived by the random weights and bias. The $\mathbf{X}$ matrix is an array of all training instances.

By incorporating a layer of random mapping $G(\mathbf{w}_i \cdot \mathbf{x} + b_i)$, a multilayered, stacked ELM-SAE, and the standard ELM defined in Eq. (5), a deep learning framework is elicited and denoted as H-ELM [34] as shown in following equation.

$$\begin{cases} \hat{\mathbf{H}}_1 = G(\mathbf{w}_i \cdot \mathbf{x} + b_i). \\ \hat{\mathbf{H}}_2 = G[\hat{\mathbf{H}}_1\mathbf{z}_{(1,2)}^T]. \\ \dots \\ \hat{\mathbf{H}}_n = G[\hat{\mathbf{H}}_{n-1}\mathbf{z}_{(n-1,n)}^T]. \\ \hat{\mathbf{Y}} = \hat{f}_S(\hat{\mathbf{H}}_n). \end{cases} \tag{7}$$

In Eq. (7), $\hat{\mathbf{H}}_n$ denotes the high level feature representation learned by the random features $\hat{\mathbf{H}}_1$, $\mathbf{z}_{(n-1,n)}$ is the fully connected weight of a ELM-SAE between $(n-1)$th and nth hidden layers, and $\hat{\mathbf{Y}}$ is the estimated output matrix. The architecture of the H-ELM is illustrated in Fig. 3(b).

3.2. DD-ELM as task-generic mental fatigue classifier

3.2.1. Framework of the classifier training and testing

The framework for task-generic mental fatigue classification is illustrated in Fig. 4. The labels of A-H and I-N denote the participant indices for Tasks 1 and 2, respectively. In Fig. 4(a), the EEG features from two tasks are independently used for training and testing with non-overlapping participant pools. The training and testing sets are then exchanged to examine the stability of the classification performance (Fig. 4(b)). Note that the order of the testing feature set labeled by the participant index is fixed from A to H or from I to N when fed to the classifier.

3.2.2. Network architecture of the DD-ELM

The architecture of the DD-ELM is shown in Fig. 5. The deep model consists of an input layer, a stacked ELM-SAE for feature abstracting, a hidden layer for high-level feature representation, another stacked ELM-SAE for feature reconstruction, and a standard ELM for mental fatigue perdition. Different from the H-ELM used for automatic feature engineering, the hierarchical architecture in the DD-ELM is designed for feature filtering. Therefore, we removed the first hidden layer for random feature mapping. The reconstructed features are expected to be task-generic with task-specific components properly filtered.

In each layer of the stacked ELM-SAE, the random feature mapping is used for EEG feature reconstruction,

$$\hat{h}_{j,i} = \frac{1}{1 + e^{-(\mathbf{w}_{j,i} \cdot \mathbf{x} + b_{j,i})}}. \tag{8}$$

where the logistic sigmoid function is used as the activation function and the values of $\mathbf{w}_{j,i}$ and $b_{j,i}$ are random according to the ELM theory. Each entry of the hidden activation $\hat{\mathbf{h}}_j$ is $\hat{h}_{j,i}$ with $i = 1, 2, \dots, d_j$.

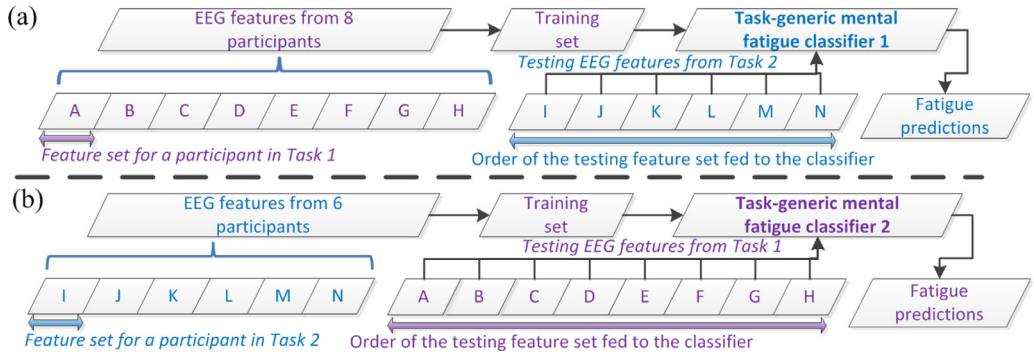

Fig. 4. Framework of the task-generic mental fatigue classifier for training and testing: (a) Case 1: EEG data from Task 1 and 2 are used for training and testing, respectively. (b) Case 2: EEG data from Tasks 2 and 1 are used for training and testing, respectively.

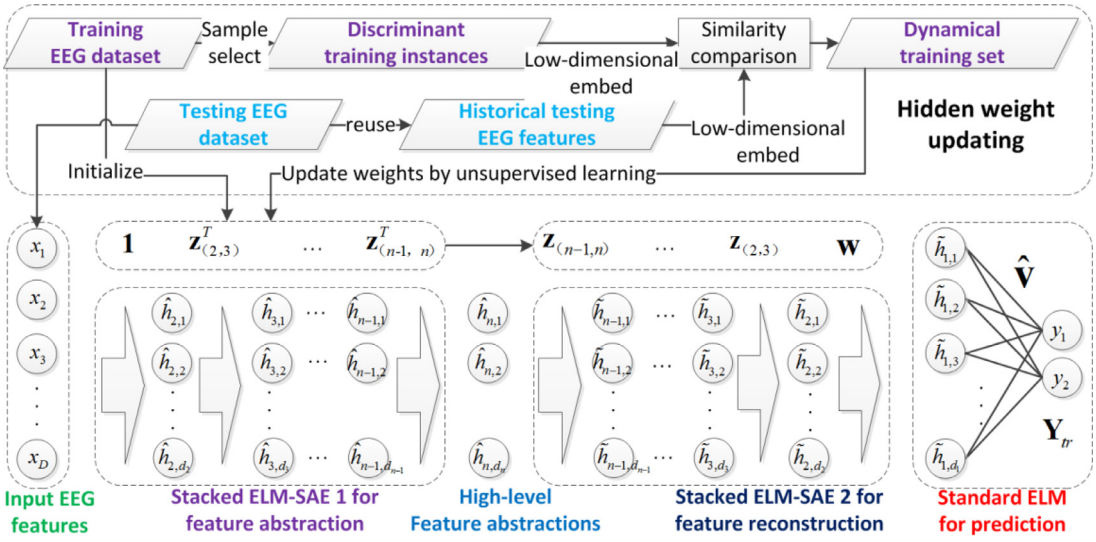

Fig. 5. Network architecture of the DD-ELM and the scheme for hidden-weight updating.

According to Eq. (7), the hidden feature representations in the DD-ELM network can be represented as follows.

$$\begin{cases} \hat{\mathbf{H}}_2 = \mathbf{X}. \\ \hat{\mathbf{H}}_3 = \tilde{G}[\hat{\mathbf{H}}_2 \mathbf{z}_{(1,2)}^T]. \\ \cdots \\ \hat{\mathbf{H}}_n = \tilde{G}[\hat{\mathbf{H}}_{n-1} \mathbf{z}_{(n-1,n)}^T] \\ \tilde{\mathbf{H}}_{n-1} = \tilde{G}[\hat{\mathbf{H}}_n \mathbf{z}_{(n-1,n)}]. \\ \cdots \\ \tilde{\mathbf{H}}_1 = \tilde{G}[\tilde{\mathbf{H}}_2 \mathbf{z}_{(1,2)}]. \end{cases} \tag{9}$$

In Eq. (9), the high-level feature representations, $\hat{\mathbf{H}}_n$, are used for feature reconstruction by the tied weighs and $\tilde{\mathbf{H}}_1$ is the random features in the last ELM layer.

According to Eq. (4), the output weight can be computed based on the filtered random feature matrix $\tilde{\mathbf{H}}_1$. The mental fatigue prediction $\hat{\mathbf{y}}$ is elicited as follows.

$$\hat{\mathbf{y}} = \tilde{h}_1(\mathbf{x})\hat{\mathbf{v}} = \tilde{h}_1(\mathbf{x})\tilde{\mathbf{H}}_1^T \left(\frac{1}{C} + \tilde{\mathbf{H}}_1 \tilde{\mathbf{H}}_1^T\right)^{-1} \mathbf{Y}_{tr}. \tag{10}$$

Based on Eqs. (5), (6), (8)–(10), the weight initialization algorithm of the DD-ELM is listed in Table 1 (denoted as Algorithm 1). The function of the algorithm is to generate the initial network parameters based on the training EEG dataset.

3.2.3. Dynamical training set of the DD-ELM

Different from the H-ELM with static network parameters, we employed a weight-updating scheme for modeling a DD-ELM.

Table 1
The pseudo codes for network weight initialization for the DD-ELM (denoted as Algorithm 1).

Start DD−ELM Weight Initialization
Set the training set S_{tr}
$S_{tr} = \{(\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2), ..., (\mathbf{x}_N, \mathbf{y}_N)\}$
Generate $\hat{\mathbf{H}}_2$ via S_{tr}
for $i = 2 : n - 1$
Build $\hat{\mathbf{H}}_{i+1}$ via ramdom mapping
$\min_{\mathbf{z}_{(i,i+1)}} \Phi[\mathbf{z}_{(i,i+1)}] = \|\mathbf{z}_{(i,i+1)}\|^2 + C\|\hat{\mathbf{H}}_{i+1}\mathbf{z}_{(i,i+1)} - \hat{\mathbf{H}}_i\|_{l_1}$
$\hat{\mathbf{H}}_{i+1} = \tilde{G}[\hat{\mathbf{H}}_i \mathbf{z}_{(i,i+1)}^T]$
End for
$\tilde{\mathbf{H}}_{n-1} = \tilde{G}[\hat{\mathbf{H}}_n \mathbf{z}_{(n-1,n)}]$
for $i = 2 : n - 1$
$\tilde{\mathbf{H}}_{n-i} = \tilde{G}[\tilde{\mathbf{H}}_{n-i+1} \mathbf{z}_{(n-i,n-i+1)}]$
End for
Build $\tilde{\mathbf{H}}_1$ via ramdom mapping
$\hat{\mathbf{v}} = \tilde{\mathbf{H}}_1^T (1/C + \tilde{\mathbf{H}}_1 \tilde{\mathbf{H}}_1^T)^{-1} \mathbf{Y}_{tr}$
Return $\mathbf{z}, \hat{\mathbf{v}}$
End DD−ELM Weight Initialization

Since the EEG data were recorded under different task environments, the dynamical weight of the deep network is expected to adapt the cross-task distribution variation of the reconstructed features.

The prerequisite for the weight updating is to build a dynamical training set. As shown in Fig. 5, the discriminant training instances

are first selected as follows.

$$\min d_{tr,\mathbf{y}^{(i)}}(\mathbf{x}_{j,\mathbf{y}}^{(i)}) = \left\| \mathbf{x}_{j,\mathbf{y}}^{(i)} - \frac{1}{N_{\mathbf{y}}} \sum_{m=1}^{N_{\mathbf{y}}} \mathbf{x}_{m,\mathbf{y}}^{(i)} \right\|, \mathbf{x}_{j,\mathbf{y}}^{(i)} \in S_{tr}^{(i)}. \tag{11}$$

In Eq. (11), $\mathbf{x}_{j,\mathbf{y}}^{(i)}$ denotes the jth training instances from the ith participant's training set $S_{tr}^{(i)}$ with the mental fatigue label $\mathbf{y}$. The number of the training instances of each class are defined by $N_{\mathbf{y}}$. By defining the solution of Eq. (11) as $\hat{\mathbf{x}}_{tr,j^*}$, a discriminant training set can be built as $\hat{X}_{tr}$.

Since the historical testing EEG data can be simultaneously stored, we reuse it to analyze the temporal similarity between the small portion of testing instances and the discriminant training instances. The goal of the similarity comparison is to select the EEG features with the similar statistical properties across two tasks.

Given a predetermined step size T_s of the stored testing instances according to the time sequence, the testing set with the size of N_{te} updates the dynamical training set for N_{te}/T_s times. The testing instances in kth step is denoted by $\hat{X}_{Ts}(k)$. Due to the high dimensionality of EEG features in $\hat{X}_{tr}$ and $\hat{X}_{Ts}(k)$, we compute the low-dimensional embeddings according to [41] before performing similarity comparison.

Given the $\hat{X}_{tr}$ with $\hat{\mathbf{x}}_{tr,j} \in \hat{X}_{tr}$, an adjacent graph is built by connecting the edges between each instance $\hat{\mathbf{x}}_{tr,j}$ and its M nearest neighbors. The weight of each edge e_{ij} is determined by the Gaussian kernel,

$$e_{ij} = \exp\left(-\frac{\left\| \hat{\mathbf{x}}_{tr,i} - \hat{\mathbf{x}}_{tr,j} \right\|^2}{2{\sigma_1}^2} \right). \tag{12}$$

The value of the σ_1 indicates the width parameter of the Gaussian kernel. That is, when σ_1 increases, the weight of the edges e_{ij} between each two pair of $\hat{\mathbf{x}}_{tr,i}$ and $\hat{\mathbf{x}}_{tr,j}$ is close to one. When such value decreases, e_{ij} is close to zero. To this end, the variation and the average of the e_{ij} can be controlled by selecting the proper σ_1. Therefore, the σ_1 can be considered as a hyper-parameter for the dynamical ELM model. Since the functionality of Eq. (12) is to generate the continuously distributed weight values for the edge in the adjacent graph, the σ_1 does not significantly influences the discriminant capability of the classifier. The parameter setting for σ_1 in this study can be found in Section 4.1.

The embeddings $\tilde{\mathbf{x}}_{tr,i}$ of $\hat{\mathbf{x}}_{tr,j}$ can be derived by minimizing the summation of the weighted distances between the neighbors $\tilde{\mathbf{x}}_{tr,i}$ and $\tilde{\mathbf{x}}_{tr,j}$ in the low dimensional feature space, i.e.,

$$\min \Psi(\tilde{\mathbf{x}}_{tr}) = \sum_{j=1}^{\hat{N}_{tr}} \sum_{i=1}^{\hat{N}_{tr}} \left\| \tilde{\mathbf{x}}_{tr,i} - \tilde{\mathbf{x}}_{tr,j} \right\|^2 e_{ij}, \tilde{\mathbf{x}}_{tr} \in R^{d_M}. \tag{13}$$

In the equation, $\hat{N}_{tr}$ and d_M denote the number of the instances in $\hat{X}_{tr}$ and the reduced dimensionality, respectively.

Given $\hat{X}_{Ts}(k)$ with $\hat{\mathbf{x}}_{Ts,i}(k) \in \hat{X}_{Ts}(k)$, the low-dimensional embedding $\tilde{\mathbf{x}}_{Ts,i}(k)$ can be computed through the nonlinear mapping between $\tilde{\mathbf{x}}_{tr}$ and $\hat{\mathbf{x}}_{tr}$ [42,43]. The Gaussian kernel is maximized to find the highest similarity R between each $\tilde{\mathbf{x}}_{tr,j}$ and the center vector of $\tilde{\mathbf{x}}_{Ts,i}(k)$.

$$\max R(\tilde{\mathbf{x}}_{tr,j}) = \exp\left(-\left\| \tilde{\mathbf{x}}_{tr,j} - \sum_{i}^{T_s} \tilde{\mathbf{x}}_{Ts,i}(k) \right\|^2 / 2{\sigma_2}^2 \right). \tag{14}$$

Given the solution of Eq. (14) as $\hat{\mathbf{x}}_{tr,j^+}$, the dynamical training set can be built by incorporating the stored testing instances $\hat{X}_{Ts}(k)$ and a group of $\hat{\mathbf{x}}_{tr,j^+}$

According to Eqs. (12)–(14), the algorithm for updating the dynamical training set of the DD-ELM is summarized in Table 2 (denoted as Algorithm 2), where N_p and $size(\cdot)$ represent the number of the participants in the training set and a function for eliciting the sample size, respectively. The dynamical training set at the $(k+1)$th testing step is defined by $\hat{X}_{dtr}(k+1)$. We also introduce a constant p_e to control the sample size of $\hat{X}_{dtr}(k+1)$.

Table 2
The pseudo codes for building the dynamical training set of the DD-ELM (denoted as Algorithm 2).

Start building the dynamical training set
Set $\hat{X}_{tr}^{(i)} = \emptyset, \hat{X}_{tr} = \emptyset, \hat{X}_{dtr} = \emptyset$
Set $S_{tr}^{(i)}$ as the training set participant i
for $i = 1 : N_p$
  while size $(\hat{X}_{tr}^{(i)}) < \hat{N}_{tr}/N_p$
    given $\mathbf{x}_{j,\mathbf{y}}^{(i)} \in S_{tr}^{(i)}$
    min $d_{tr,\mathbf{y}^{(i)}}(\mathbf{x}_{j,\mathbf{y}}^{(i)})$
    $\hat{X}_{tr}^{(i)} \leftarrow \hat{\mathbf{x}}_{tr,j^*}$
    remove $\hat{\mathbf{x}}_{tr,j^*}$ from $S_{tr}^{(i)}$
  End while
  $\hat{X}_{tr} = \hat{X}_{tr} \cup \hat{X}_{tr}^{(i)}$
End for
for $i = 1 : \hat{N}_{tr}$
  for $j = 1 : \hat{N}_{tr}$
  $e_{ij} = \exp(-\|\hat{\mathbf{x}}_{tr,i} - \hat{\mathbf{x}}_{tr,j}\|^2 / 2{\sigma_1}^2)$
  End for
End for
min $\Psi(\tilde{\mathbf{x}}_{tr}) = \sum_{j=1}^{\hat{N}_{tr}} \sum_{i=1}^{\hat{N}_{tr}} \|\tilde{\mathbf{x}}_{tr,i} - \tilde{\mathbf{x}}_{tr,j}\|^2 e_{ij}$
generate mapping $\phi: \tilde{\mathbf{x}}_{tr} \rightarrow \hat{\mathbf{x}}_{tr}$
given $\hat{\mathbf{x}}_{Ts}(k) \in \hat{\mathbf{X}}_{Ts}(k)$
$\tilde{\mathbf{x}}_{Ts}(k) = \phi[\hat{\mathbf{x}}_{Ts}(k)]$
while size$(\hat{X}_{dtr}) < \hat{N}_{tr} \cdot p_e$
  max $R(\tilde{\mathbf{x}}_{tr,j})$
  $\hat{X}_{dtr} \leftarrow \hat{\mathbf{x}}_{tr,j^+}$
  remove $\hat{\mathbf{x}}_{tr,j^+}$ from $\hat{X}_{tr}$
End while
$\hat{X}_{dtr}(k+1) = \hat{X}_{Ts}(k) \cup \hat{X}_{dtr}$
Return $\hat{X}_{dtr}(k+1)$ at each testing step
End updating the dynamical training set

Table 3
The pseudo codes for the testing procedure of the DD-ELM.

Start mental fatigue prediction by DD−ELM
Set testing set $X_{Te} = \bigcup_{k=1}^{N_{te}/T_s} \hat{X}_{Ts}(k)$
$\hat{X}_{Ts}(i) \cap \hat{X}_{Ts}(j) = \emptyset$
Define F_k as the DD − ELM model at k^{th} step
Call Algorithm 1
Return $\{\mathbf{z}(1), \hat{\mathbf{v}}\}$
$F_1 \leftarrow \{\mathbf{z}(1), \hat{\mathbf{v}}\}$
for $k = 1 : N_{te}/T_s$
  Call Algorithm 2
  Return $\hat{X}_{dtr}(k+1)$
  for $i = 2 : n_s, n_s \le n - 1$
    Build $\hat{\mathbf{H}}_{i+1}$ via ramdom mapping
    min $\Phi[\mathbf{z}_{(i,i+1)}(k+1)] = \|\mathbf{z}_{(i,i+1)}(k+1)\|^2 +$
    $C\|\hat{\mathbf{H}}_{i+1}(k+1)\mathbf{z}_{(i,i+1)}(k+1) - \hat{\mathbf{H}}_i(k+1)\|_{l_1}$
    $\hat{\mathbf{H}}_{i+1}(k+1) = \tilde{G}[\hat{\mathbf{H}}_i(k+1)\mathbf{z}_{(i,i+1)}^T(k+1)]$
  End for
  Return $\mathbf{z}(k+1) = [\mathbf{z}_{(1,2)}(k+1), \mathbf{z}_{(2,3)}(k+1),$
    $..., \mathbf{z}_{(n_s-1,n_s)}(k+1)]$
    $F_{k+1} \leftarrow \{\mathbf{z}(k+1), \hat{\mathbf{v}}\}$
    Return F_{k+1}
End for
End mental fatigue prediction of DD − ELM

3.2.4. Dynamical weight updating of the DD-ELM

The testing procedure of the DD-ELM is presented by the pseudo codes in Table 3. The testing set X_{Te} is split into N_{te}/T_s non-overlapping subsets of equal size according to the time sequence.

The initial DD-ELM model F_1 is generated by Algorithm 1. It facilities the mental fatigue prediction at the first time step, where

Table 4
The selected hyper-parameters for the DD-ELM.

Hyper-parameter of DD-ELM	Case 1		Case 2	
	Session 1	Session 2	Session 1	Session 2
Number of hidden layers	10	10	10	10
Number of hidden neurons	10	10	10	10
Regularization parameter	2^{-2}	1	2	1
Step size	58	145	90	540
Weight in dynamical training set	0.1	0.2	0.2	0.2
Number of hidden layers for weight updating	5	3	2	3
The number of the discriminant instances	10	10	10	10
The number of the nearest neighbors	12	12	12	12
Widths in two Gaussian kernels	1	1	1	1
The output dimensionality of the embeddings	10	10	10	10

Note: The session-specific parameter settings are marked in bold.

the first T_s EEG instances are fed to F_1. Then, the dynamical training set $\hat{X}_{dtr}$ is initialized and updated by Algorithm 2. In particular, $\hat{X}_{dtr}(k+1)$ is constructed by discriminant training instances and the historical testing data $\hat{X}_{Ts}(k)$ of previous time step k. According to Eq. (6), the 2nd to n_sth layers of hidden weights in each stacked ELM-SAE are iteratively updated and elicit the DD-ELM model F_{k+1} at $k+1$ time step. Note that the input and output weights $\mathbf{w}$ and $\hat{\mathbf{v}}$ of the ELM in the decision layer are fixed through the testing procedure.

4. Results

4.1. Task-generic mental fatigue classification via DD-ELM

According to the classification paradigm presented in Fig. 4, the task-generic mental fatigue recognition results were separately computed for all session 1 or 2. The classification accuracy (*ACC*) was calculated twice for each classification case to investigate the repeatability of the classifier performance.

All the weights connecting each two consecutive hidden layers and the bias of all hidden neurons are set to the pseudorandom values drawn from the uniform distribution on the open interval $(-1, 1)$ and $(0, 1)$, respectively. Such scheme is consistent with the original ELM approach proposed in [33]. All the hyper-parameters of the DD-ELM are selected by the cross-validation and the grid search on the training set as follows. We first determine the structural hyper-parameter, i.e., number of hidden layers and hidden neurons from candidate sets {2, 3, ..., 11} and {10, 20, ..., 100}, respectively. Then, we select the regularization parameter within $\{2^{-4}, 2^{-3}, ..., 2^4, 2^5\}$. Then, the step size is optimized by exploring the sets of {29, 58, ..., 290} and {90, 180, ..., 900} for two experimental cases, respectively. The parameter for weighting the training and testing instances in the dynamical training set is selected from {0.1, 0.2, ..., 1}. The number of hidden layers for weight updating is selected from {1, 2, ..., 10} while the number of the discriminant instances are set to 10 to reduce the computational time. In the end, the number of the nearest neighbors, the widths in two Gaussian kernels and the output dimensionality of the embeddings are fixed as 12, 1, 1, and 10, respectively, since these parameters does not significantly influence the classifier's performance. The results of the hyper-parameter selection are summarized in Table 4. Note that the session-specific parameter settings are highlightened.

By implementing the DD-ELM algorithms, the task-generic mental fatigue classification results for two representative participants under Case 1 are shown in Fig. 6. In Fig. 6(a) and (b), we illustrate the details of the classification performance at each time instant of participant K. The testing classification accuracy is achieved at 0.7551 and 0.6551 for sessions 1 and 2, respectively. It is shown most of the EEG feature vectors have been correctly recognized as low or high mental fatigue levels. The EEG features located at the switch point of the binary classes are more difficult for discrimination. In Fig. 6(c) and (d), the accuracy for each session of participant L is 0.7586 and 0.9758. It is shown that the classification performance is improved along with the increase of the time index in Fig. 6(c). Such observation partially reflects the merit of the scheme for dynamical weight updating in the DD-ELM. It is also interesting that almost perfect classification is achieved in Fig. 6(d).

Table 5
Mental fatigue confusion matrices summated over all participants by DD-ELM for the task-generic classification paradigm of Case 1.

Session index	Predicted mental fatigue	Target mental fatigue		Classification performance
		Low	High	
Session 1		$P_{pre}=0.6615$	$P_{npv}=0.6514$	$P_{acc}=0.6563$
	Low	**557**	285	$P_{sen}=0.6402$
	High	313	**585**	$P_{spe}=0.6724$
Session 2		$P_{pre}=0.6751$	$P_{npv}=0.7053$	$P_{acc}=0.6890$
	Low	**634**	305	$P_{sen}=0.7287$
	High	236	**565**	$P_{spe}=0.6494$

Note: The largest number of instances in each column for each session is marked in bold.

Table 6
Mental fatigue confusion matrices summated over all participants by DD-ELM for the task-generic classification paradigm of Case 2.

Session index	Predicted mental fatigue	Target mental fatigue		Classification performance
		Low	High	
Session 1		$P_{pre}=0.6729$	$P_{npv}=0.6498$	$P_{acc}=0.6605$
	Low	**2249**	1093	$P_{sen}=0.6247$
	High	1351	**2507**	$P_{spe}=0.6963$
Session 2		$P_{pre}=0.7031$	$P_{npv}=0.6486$	$P_{acc}=0.6716$
	Low	**2139**	903	$P_{sen}=0.5941$
	High	1461	**2697**	$P_{spe}=0.7491$

Note: The largest number of instances in each column for each session is marked in bold.

The detail classification results for two participants under Case 2 are illustrated in Fig. 7. In Fig. 7(a) and (b), the testing accuracy of the participant A is achieved at 0.9844 and 0.5544 for the EEG data of sessions 1 and 2, respectively. Although the classification performance is very satisfactory for session 1, quite a few EEG samples belonging to low fatigue class are misclassified as high class in Fig. 7(b). For testing data from participant C, the accuracy is 0.9222 and 0.8555 for two sessions. Compared to Fig. 6(c), similar observation is found in Fig. 7(d), where the EEG features extracted with consecutive time indices are incorrectly predicted at the switch point of two classes.

The participant-specific testing accuracy (denoted by P_{acc}) for all cases and sessions are shown in Fig. 8. Under Case 1, all participants, except for participant I with session 2, the P_{acc} values are higher than the random classification rate 0.5. Under Case 2, only the P_{acc} values of participant E with session 2 and participant G with session 1 are lower than 0.5. By employing the two-tailed t-test, we found the DD-ELM can lead to significant ($p<0.05$) higher average accuracy across participants A to N with (t=2.9321, p=0.012) and (t= 3.684, p=0.003) than the random case for sessions 1 and 2, respectively.

In Tables 5 and 6, the overall confusion matrices under two cases are listed. Taken the upper matrix in Table 5 as an example,

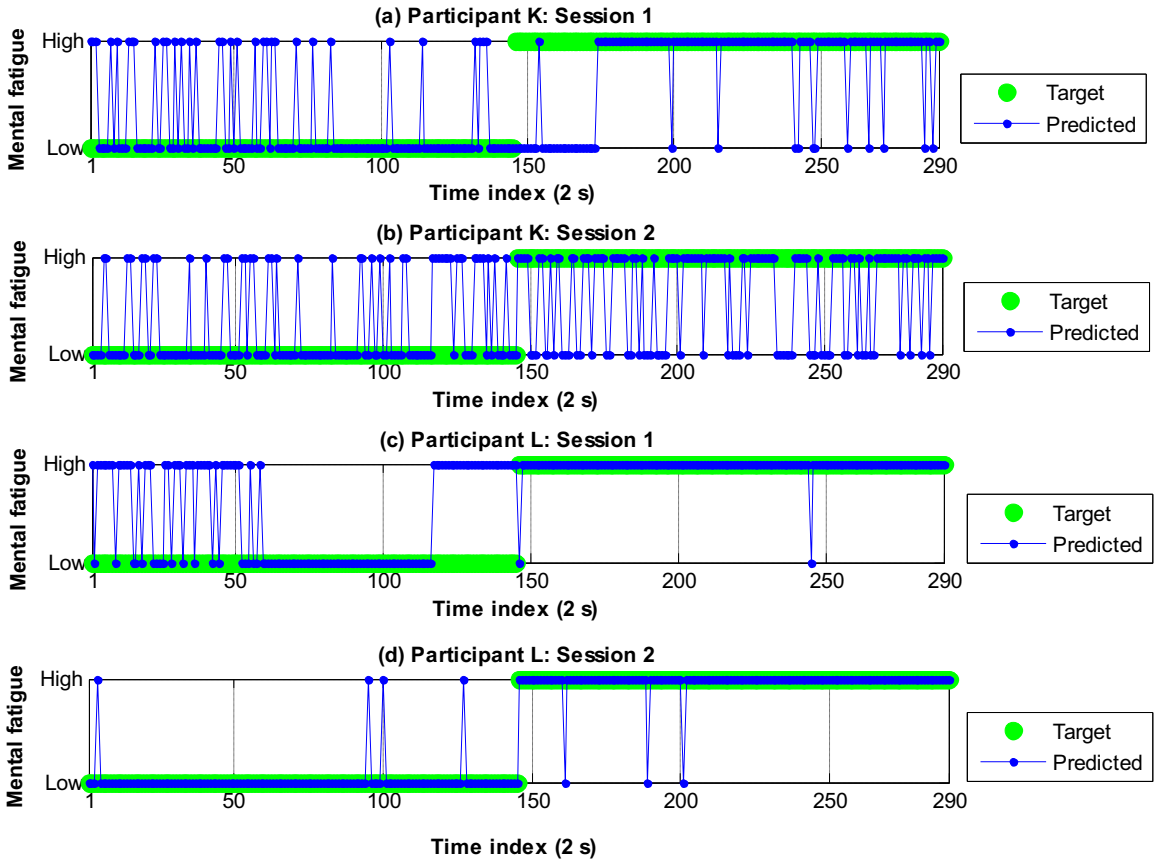

Fig. 6. Task-generic mental fatigue estimation results for Participants K and L under the classification paradigm of Case 1. The time courses of the target and predicted mental fatigue levels are shown and compared.

557 and 585 testing instances from low and high classes are correctly estimated under Case 1. It leads to an average P_{acc} of 0.6563. On the other hand, 313 instances belonging to low fatigue are misclassified as high fatigue. Given low fatigue level as the positive class $(+1)$, the sensitivity and the precision denoted by P_{sen} and P_{pre} can be computed as 0.6402 and 0.6615, respectively. Similarly, 285 high-fatigue instances are incorrectly predicted as low fatigue. By defining the high fatigue as the negative class (-1), the specificity and the negative prediction value (denoted by P_{spe} and P_{npv}) are elicited as 0.6724 and 0.6514, respectively. In general, the participant average P_{acc} for all cases and sessions are higher than 0.65 when the DD-ELM fatigue classifier is implemented. However, it is noted the DD-ELM tend to misclassify EEG samples as high fatigue state under Case 2.

4.2. Classification performance comparison

To validate the effectiveness of the DD-ELM, we first introduced the H-ELM model for performance comparison. The regularization parameter C of the H-ELM was selected under the same condition as shown in Section 4.1. Then, we computed the classification accuracy across two cases and sessions in Fig. 9. For each column data in the bar plot, the value of the participant-average P_{acc} from the testing set is illustrated. The performance of the structural hyper-parameters, i.e., the number of the hidden layers and the neurons in each hidden layer, has been exhausted. Although the optimal H-ELM structure seems to be case-specific and session-dependent, we still find stable accuracy values across all conditions across Fig. 9(a)–(d) when 11 hidden layers with each of 10 neurons are employed. However, the P_{acc} is lower than 0.6 for all parameter combinations, which indicates the superiority of the DD-ELM with that value higher than 0.65.

We also investigated three shallow learning machines, i.e., the classical ELM, the least square support vector machine (LSSVM) with the linear kernel, and the ANN with the single hidden layer, across different hyper-parameters. The participant-average values of the training and the testing P_{acc} for two cases are plotted in Fig. 10. For the ELM classifier, the number of the hidden neurons was selected within a candidate set {10, 20, ..., 500}. The training P_{acc} stably improves along with the increase of the hidden neurons. However, the testing P_{acc} for both of the two cases does not drastically vary and the optimal value is achieved with small size of the hidden nodes. For the LSSVM, we examined the P_{acc} across different regularization parameters $\{2^{-14}, 2^{-13}, ..., 2^{-15}\}$. It is shown the training and the testing performance is inconsistent with the optimal testing P_{acc} achieved with small parameter values. The similar observation is found for the ANN classifier in Fig. 10(e) and (f). In particular, for three classifiers under all parameter conditions, the testing accuracy is much lower than 0.65 derived by the DD-ELM.

The participant-average classification accuracy for 15 classifiers under two Cases are shown in Fig. 11, respectively. The ANN, LSSVM and ELM classifiers employ optimal hyper-parameters regarding the testing performance in Fig. 10. In addition, three classical shallow learning machines, i.e., logistic regression (LR), naive Bayesian (NB), and K-nearest neighbor (KNN) classifiers, were examined. We also combined the principal component analysis (PCA) preprocessing with 50 principal components for shallow models to form five hybrid classifiers, i.e., PCA-NB, PCA-KNN, PCA-LSSVM, PCA-NB, and PCA-ELM. Finally, two deep learning primitives, H-ELM and stacked autoencoder (SAE), were investigated under the

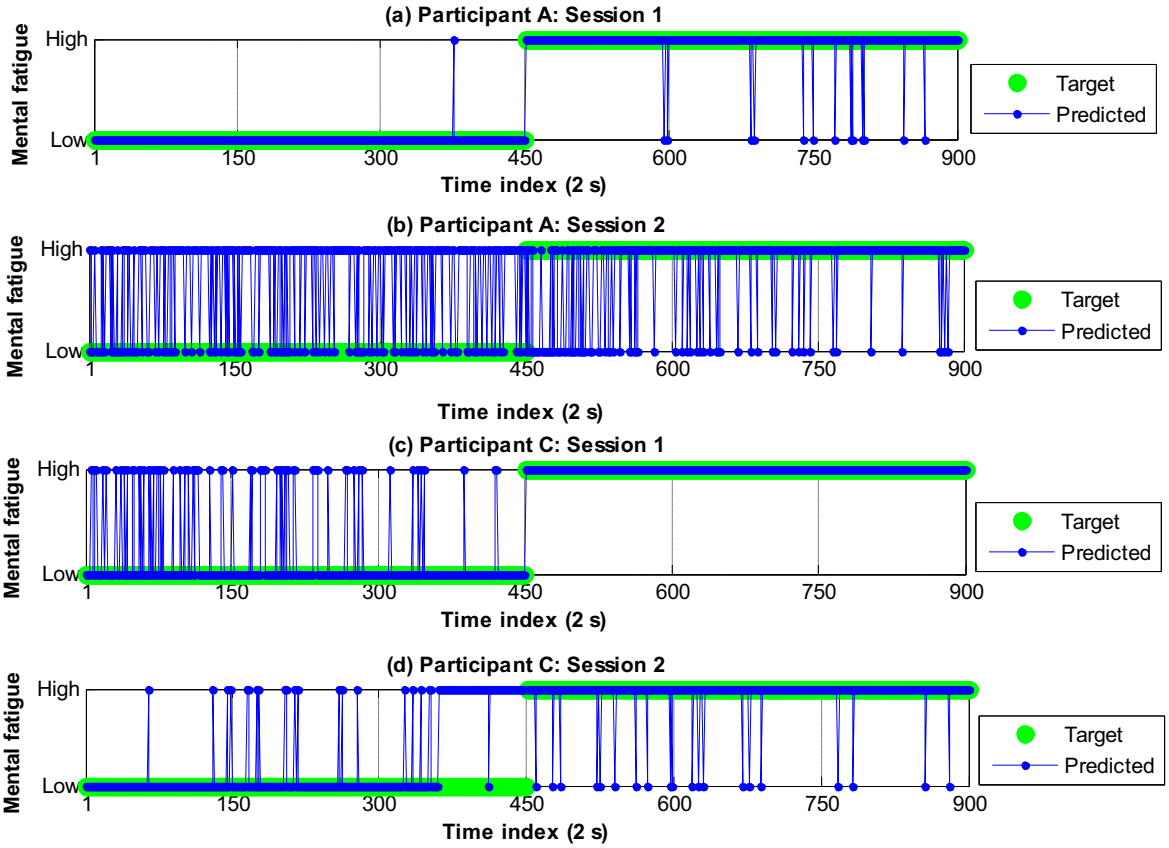

Fig. 7. Task-generic mental fatigue estimation results for Participants A and C under the classification paradigm of Case 2.

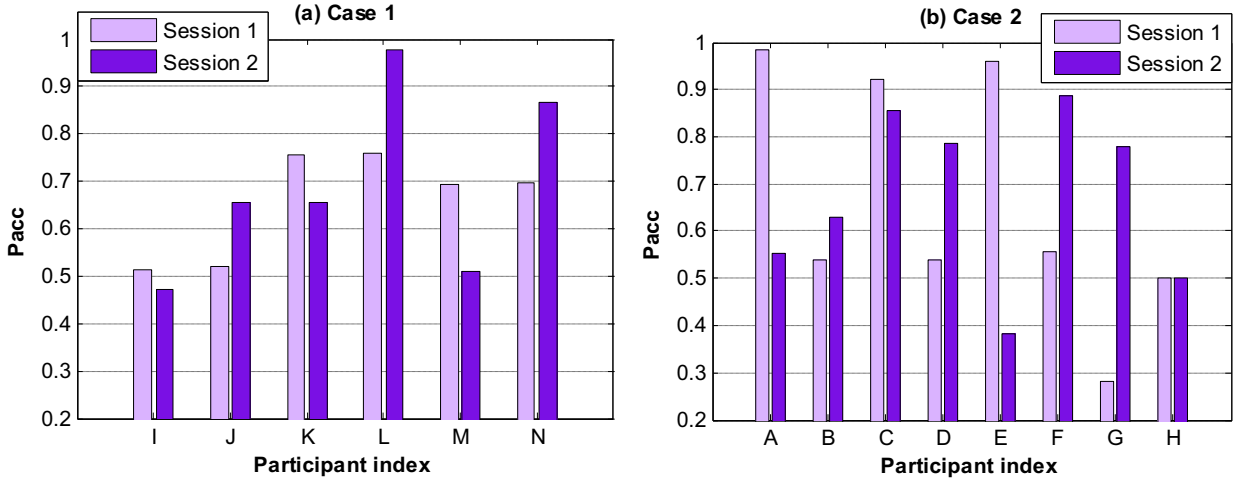

Fig. 8. Participant-specific mental fatigue classification accuracy for each classification case by DD-ELM.

optimal network structure. The performance of the DD-ELM is consistent with those shown in Fig. 8.

From Fig. 11, it is shown the DD-ELM achieves the highest value of the mean testing accuracy. Based on the two-tailed *t*-test, we find such improvement is significant against all 14 classifiers ($p < 0.05$). It can be observed that the generalization capability of the conventional shallow learning machines except for LSSVM and ELM are undermined with experimental paradigm of Case 2. The reason behind is that the size of the training instances is limited when the EEG data of Task 2 are used to build the classifier while only the ELM and LSSVM introduce a regularization term to control the fitting capacity. The H-ELM also shows better performance than that of the SAE because of such merit inherited from the ELM. The superiority of the DD-ELM is closely related to the dynamical weight updating. It facilitates the flexibility of the deep network and improves transferability for learning high-level features across two similar but different data distributions.

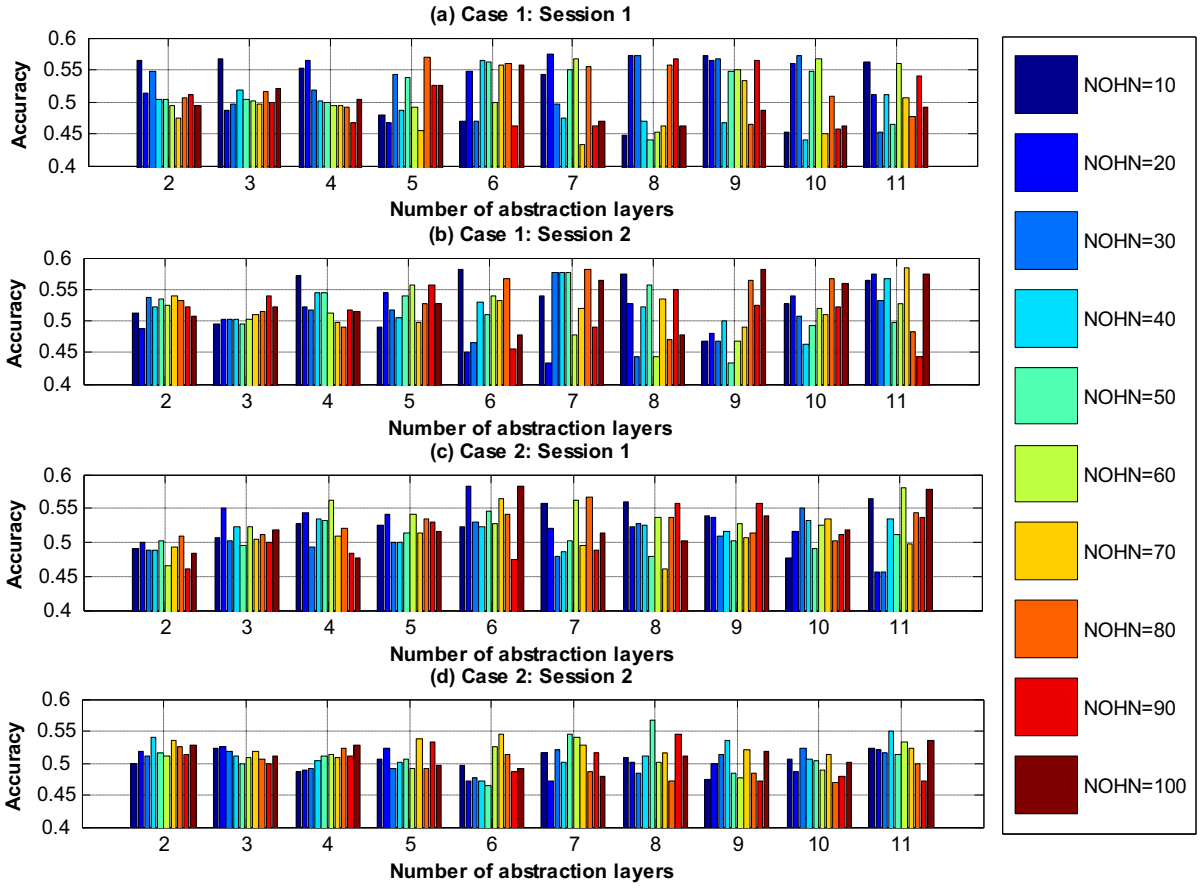

Fig. 9. Task-generic mental fatigue classification results computed via the H-ELM: (a), (b) Case 1: testing accuracy of Sessions 1 and 2. (c), (d) Case 2: testing accuracy of Sessions 1 and 2. The term, NOHN, denotes the number of hidden neurons in a hidden layer.

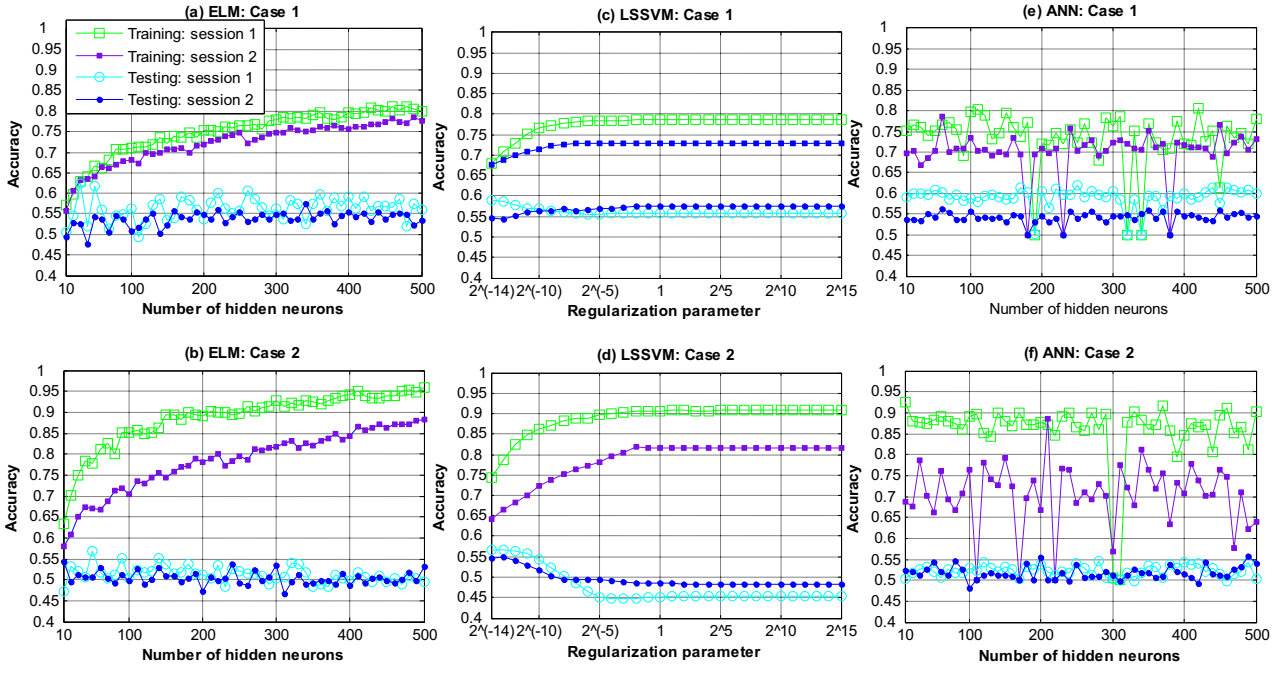

Fig. 10. Task-generic mental fatigue classification results under different hyper-parameters across (a), (b) the ELM classifier, (c), (d) the linear LSSVM classifier, (e), (f) and the ANN with single hidden layer. The training and testing accuracies on all sessions and cases are shown. In subfigures (b)–(f), all the legends are as same as those in subfigure (a).

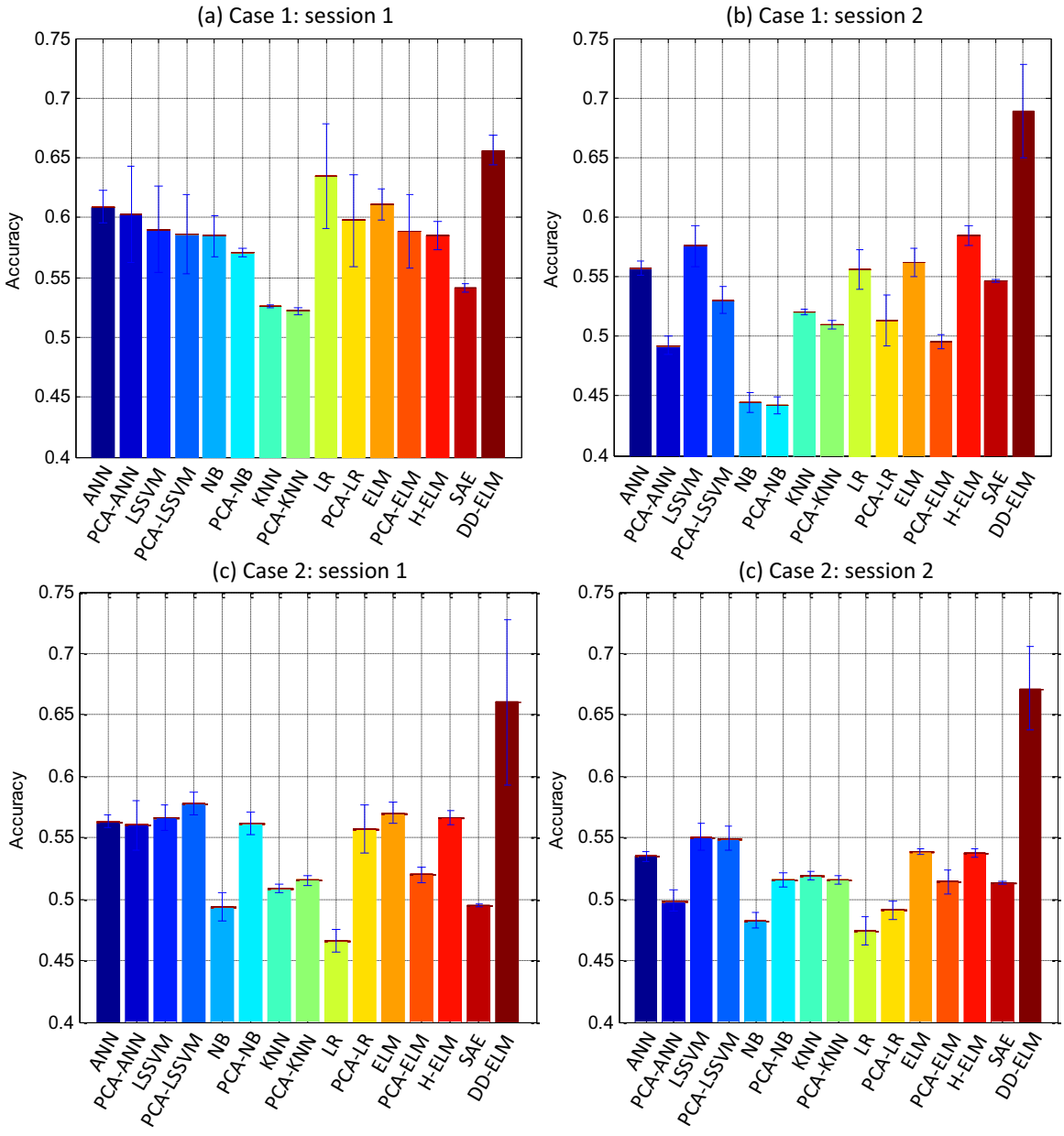

Fig. 11. Task-generic mental fatigue classification accuracy comparison across 15 classifiers.

4.3. DD-ELM performance with different parameter settings

In DD-ELM framework, we found the task-generic classifier's performance is closely related to the combination of two hyper-parameters, i.e., the regularization weight C and the number of the dynamical hidden layers n_s. Different from the classical autoencoder trained by the back propagation algorithm, the C value in each stacked ELM-SAE controls the accuracy for feature reconstruction. The EEG feature abstractions from the input layer can be recovered in an accurate (or approximate) manner with large (or small) C employed. Therefore, it indicates the strength of the noise filtering through the hierarchical structure in DD-ELM. On the other hand, given each stacked ELM-SAE with $n-1$ hidden layers, only the weights in the 2nd to n_sth layers ($n_s \leq n-1$) are dynamically updated. Such scheme exploits the merit that neuron activations in deep layers are more stable and noise-free than that from shallow ones. In addition, it reduces the computational cost for dynamical classification by avoiding recomputing whole network's weights at each time step.

The mean testing accuracy of the DD-ELM classifier derived by using different parameter combinations are shown in Fig. 12. In Fig. 12(a), the optimum performance is achieved by sufficiently large C ($C \geq 1$) when only two hidden layers are involved for weight updating. The generalizability is impaired with the increasing n_s value. In Fig. 12(b), the best performance is achieved at $C = 0.125$ and 5 dynamical layers. Such difference shows the session-to-session variability of the EEG data distribution. The variability can be originated from the EEG signals recorded on separate days while the day-to-day transition of the physiological states leads to the feature nonstationarity. For Fig. 12(c) and (d), we found the highest accuracy is reached at small C values combined with a single dynamical layer. The reason behind is that the fewer EEG instances are available for training under Case 2 and the model performance can be improved via large filtering strength of stacked

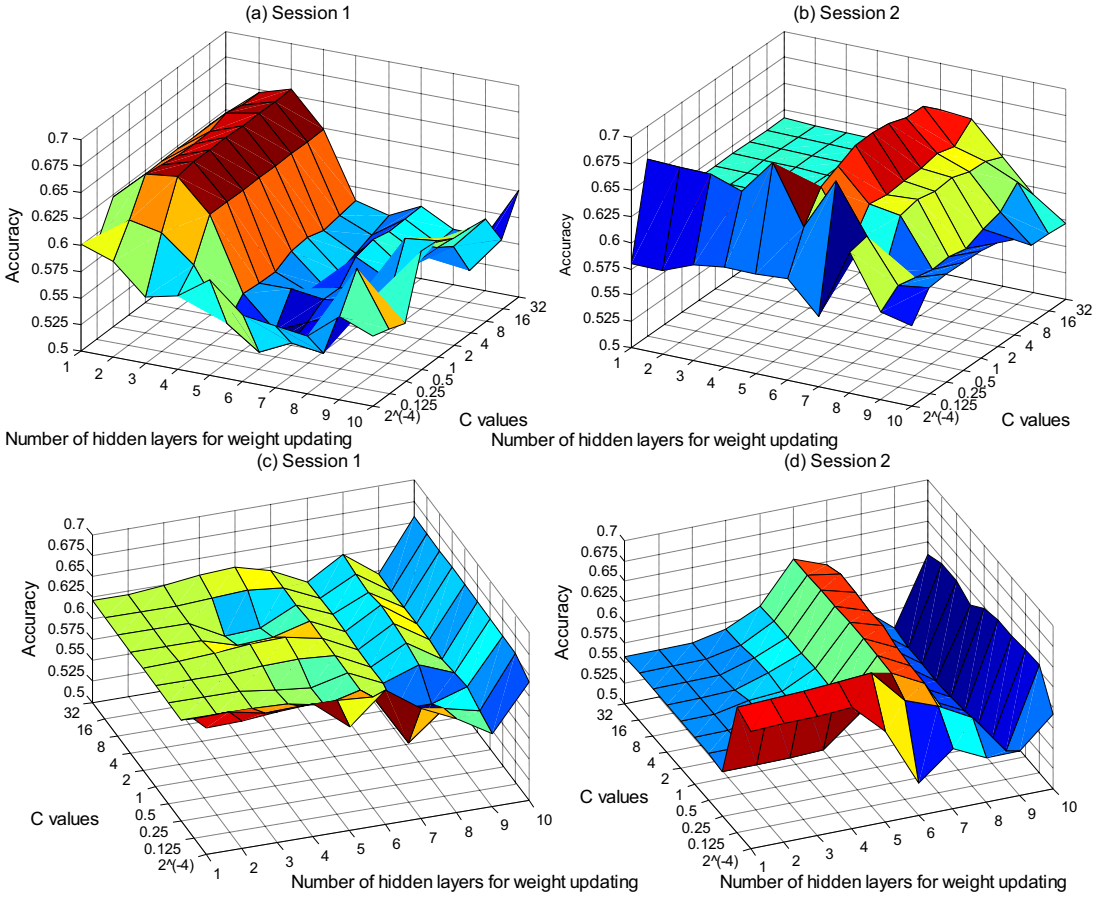

Fig. 12. Participant-average testing accuracy vs. different parameter combinations for DD-ELM.

ELM-SAE. Base on that, we recommend selecting small n_s for DD-ELM while appropriate C value should be consistent with the size of the training set.

4.4. Computational cost analysis on task-generic classifiers

Table 7 lists the computational costs of all 15 classifiers used for task-generic mental fatigue recognition. For each classifier, we ran the algorithm defined in Section 4.2 on a participant's EEG dataset of a single session. For each experimental trial, the training and testing of a specific classifier were carried out and the summation of the CPU time was recorded. Two cases are separately investigated while 50 trials of each case were performed to elicit the mean and s.d. values. All algorithms were programmed by using Matlab® 2011b software and tested via a laptop computer with Windows 8® operating system, AMD® CPU 2.0 GHz and 8 GB RAM configurations. In particular, the ANN classifier was built via Matlab® Neural Network Toolbox (Ver. 7.0.2). The LSSVM classifier was achieved by using the LSSVM toolbox developed by Suykens and Vandewalle [44]. The ELM and H-ELM codes are developed by Huang et al. and are publicly available through [33,34,41]. The details of the convolutional neural network (CNN) and the deep belief network (DBN) can be found in Section 4.5. The 50 trials of the experiments are applied in order to eliminate the hardware or software factors that can affect the computational time of the whole program. For instance, when a only a single trial is carried out, the valued of the derived CPU time may be distorted by some transient electromagnetic noise. Therefore, the average results of multiple trials can be quite suitable to characterize the computational burden of each classification algorithms. The selection on number of experimental trials is required to be sufficiently large. Here, 50 trials are acceptable since we found the average CPU time has been already converged after more than 30 trials are performed.

Table 7
Participant-average CPU time (in sec) for training and testing a task-generic classifier using single-session EEG data. The mean and s.d. values are computed across 50 repeated trials.

Classifier	Case 1		Case 2	
	Mean	S.D.	Mean	S.D.
ANN	**456.45**	**231.79**	**81.34**	**24.55**
PCA-ANN	275.09	215.69	42.37	14.82
LSSVM	142.01	4.13	5.89	0.33
PCA-LSSVM	136.89	5.21	4.74	0.36
NB	*0.77*	0.14	*0.74*	*0.07*
PCA-NB	1.14	*0.13*	0.94	0.10
KNN	5.65	0.33	5.57	0.64
PCA-KNN	3.17	0.23	2.89	0.25
LR	6.32	0.45	12.5	0.33
PCA-LR	2.10	0.17	1.46	0.20
ELM	1.33	0.17	1.30	0.19
PCA-ELM	1.51	0.18	1.44	0.16
H-ELM	13.03	0.56	4.00	0.29
SAE	17.13	0.75	6.06	0.40
CNN	3.98	0.24	1.71	0.08
DBN	11.02	0.15	4.17	0.21
DD-ELM	27.40	1.28	47.24	2.58

Note: The smallest and largest values in each column are marked in bold and italic styles, respectively.

In general, the NB and the ANN classifiers require the lowest and the highest CPU time for each trial, respectively. Under Case 1, the performance of the DD-ELM is superior to PCA-ANN, LSSVM,

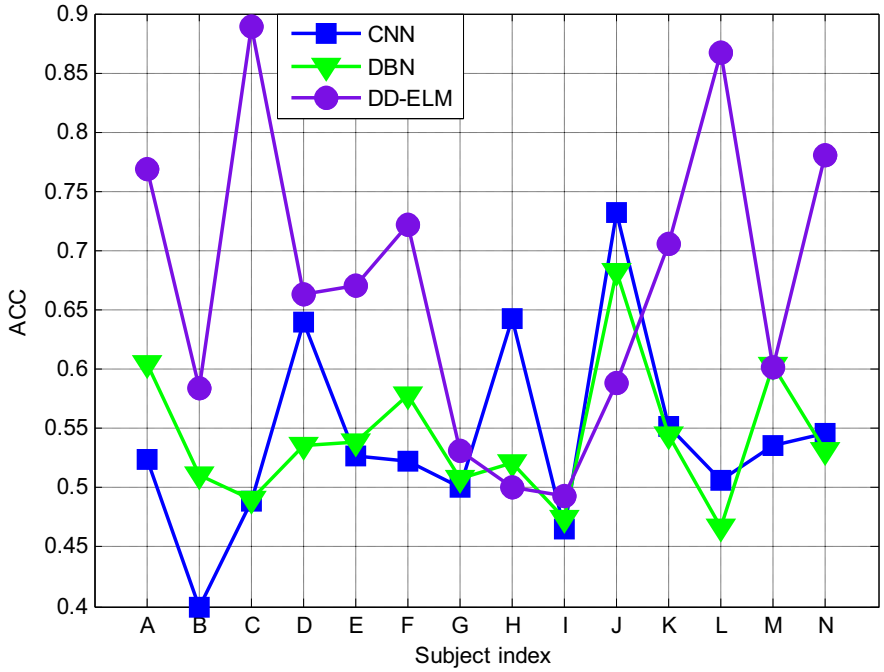

Fig. 13. Participant-specific testing accuracy for different deep learning models.

and PCA-LSSVM. Although the accuracy of the ANN, LSSVM, H-ELM and ELM are comparable, the ELM and H-ELM require much less time for training and testing. The DD-ELM inherits such merit and achieves relatively lower computational cost compared to the conventional ANN trained by the BP algorithm. Opposite to the static classifiers, the mean CPU time of the DD-ELM increases under Case 2. The potential reason is due to a larger testing set while more loops for iteration are required on weight updating. The solution to reduce the computational cost for large testing sample is to consistently enlarge the step size for dynamical classification. By incorporating the results in Fig. 11 and Table 6, the DD-ELM possesses a higher generalizability and an acceptable computational burden, which leads to the competitive capability for identifying the abnormal mental fatigue level in a task-generic manner.

4.5. Performance comparison against other deep learning models

To further validate the DD-ELM's performance, we newly introduced two deep learning models to recognize the mental fatigue levels via task-generic EEG data. The first deep classifier is the convolutional neural network with hybrid structure (denoted as CNN for simplicity). The deep CNN model is built via several consecutive convolutional and subsampling layers. Since the 2-D input data is required for such CNN classifier. We reshape the EEG feature vector of 137 dimensionality as the data matrix with the size of 12 by 12. For each instance, the value the last 7 pixels ($12^2 - 137 = 7$) is set to 1. The logistic sigmoid function is applied to preprocess the features within the open interval (0, 1). After the higher-level feature abstractions are computed, an additional layer for feature standardization is implemented. The classification is achieved by introducing a Bayesian classifier at the end of the CNN. Moreover, another deep learning method, deep belief network (DBN), is applied for the performance comparison. The DBN is constructed by stacking several Restricted Boltzmann Machines (RBMs). In each hidden layer, the parameters (neuron weights and bias) of the RBM model can be pre-trained by using back-propagation algorithm. The fine-tuning is further carried out to optimize the parameters of the whole DBN by adding a decision layer at the end of the network.

After all hyper-parameters of the CNN and the DBN models are optimized via grid search and the cross-validation technique, the mental fatigue classification accuracies of the two deep models are compared against the DD-ELM in Fig. 13. The results of all 14 subjects from both of two experimental cases are shown. For most subjects' EEG datasets, the DD-ELM outperforms the CNN and DBN. The exceptions are only found for subjects H and J. In general, the DD-ELM achieves the highest performance (0.6689) while the CNN and the DBN are comparable (0.5418 and 0.5421, respectively). According to the *t*-test, we found the superiority of the DD-ELM is significant against CNN ($t = -2.9195$, $p = 0.011$) and DBN ($t = -3.2367$, $p = 0.006$).

5. Discussions

Different from our previous work [36] where the EEG data used for OFS identification were collected via exactly the same experimental paradigm, we employ two similar but different mental tasks as fatigue stimuli. Designing such task-generic mental fatigue classifier can significantly reduce the time cost for task-specific neurophysiological data collection. By investigating the correlation coefficient between the time courses of the target labels and the EEG features, we found the consistency exists for theta and alpha power located in prefrontal cortex and the inter-hemisphere difference of high frequency power. This fact indicates there is an EEG feature subset possessing stable statistical properties along with the accumulation of the mental fatigue even for different modes of human–machine interaction. It thus facilitates the feature engineering by learning the transferable, combined EEG indicators.

One of the challenges for applying deep learning approaches on classifying EEG can be the inconsistent data distributions across training and testing set caused by the difference of the stimuli modes and the participant pools across two tasks. As shown in Fig. 10, the merit of the deep neuron network for feature abstraction is not fully exploited since the performance of the SAE is lower than the LSSVM for most cases. Although we exhausted different number of hidden layers and nodes for the H-ELM, the binary classification ac-

curacy is still below 60% in Fig. 9. Such facts imply the classical deep learning methods with static network parameters may not adapt the nonstationary, heterogeneous EEG features.

The main contribution of study lies in the feasible schemes of dynamical weight updating for the deep ELM network. Compared to our recent study [29] where the ensemble DBN model with switched structure was applied to adapt the EEG individual difference, we fuse the fixed network structure and the dynamical weights together to balance the computational cost and the generalizability. Since the testing accuracy of the deep ELM model with the varied network structure is still limited, we adopt additional hyper-parameters in DD-ELM to specify the updating frequency and the topology of the dynamical hidden layers. The effectiveness of such scheme can be partially validated in Fig. 12, where the proper combination of the dynamical layer number and regularization parameter significantly improves the testing performance.

Another advantage of the DD-ELM is that the weight updating is achieved by the unsupervised feature learning. In [39], the BSVM model was used to predict the cognitive state of the operator while the updating for support vectors requires the testing labels in past time steps. However, the ELM autoencoder generated the updated weights without the labels of the historical EEG instances. Although the DD-ELM framework possesses additional procedures for selecting training instances and the low-dimensional embeddings, the computational cost is still acceptable. These facts indicate the potential of the DD-ELM for real-time mental fatigue prediction.

The idea of the DD-ELM is originated by combining the transfer learning and the ELM theory such that its framework is different from the deep network constructed by the CNN. The CNN consists of several convolution and subsampling layers that could effectively handle the 2-D input data. The convolution layer used as the pattern filter forms a partially connected network since only a patch of neurons shares the weights with those in the consecutive layer. It leads to advantages in extracting the local spatial features for high-resolution images. On the other hand, the DD-ELM adopts the fully connected auto-encoders to achieve the feedforward pass. Such design is suitable for the current study as input EEG patterns are computed in the temporal or frequency domains independent from spatial locations. Moreover, training of the DD-ELM with the random feature mapping and the minimization of the weight norm is characterized by low-time cost. The deep CNN tuned by the hierarchical back-propagation algorithm possesses lower training speed since both of the pre-training and the fine turning are required. Most importantly, the CNN uses the static weights through the testing implementation while the DD-ELM adapts the variation of the testing pattern distribution with dynamical weights. It leads to the superiority of the DD-ELM in handling the nonstationary, cross-task EEG data.

The limitation of the present study can be summarized into two aspects.

(1) Although task-generic manner has been applied, the size of the instances is still not sufficiently large for training the deep model. The feature work is to further rich the participant pool for both of two tasks with multiple types of the fatigue factors. It can help to better justify the DD-ELM model.
(2) The testing accuracy is not perfect for binary mental fatigue recognition because of the complexity of the high dimensional EEG features. The cross-task EEG feature selection is expected to further improve the deep classifier's performance. The cross-task, multiclass mental fatigue estimation can be also taken into consideration.

6. Conclusion

In this study, we developed a task-generic mental fatigue estimator, DD-ELM, to recognize the EEG features into different fatigue levels. The DD-ELM was modified from the hierarchical ELM to improve its robustness against the varied statistical property of neurophysiological data originated from different modes of human–machine interactions. The proposed method employed two ELM-based, stacked autoencoders as the adaptive filter to eliminate the inter-task variation of 137 EEG indicators. The weights in shallow layers of the DD-ELM were iteratively updated according to the low-dimensional embeddings fused from historical testing instances and discriminant training samples. The effectiveness of the DD-ELM has been investigated on two mental tasks designed via AutoCAMS software. The average classification rate on binary fatigue degree is 67%, which is significantly superior to 14 classical classifiers with optimal hyper-parameters selected. Both of the improvement of the classification performance as well as the acceptable computational cost shows the potential of the DD-ELM in real-time fatigue estimation via EEG signals.

Conflict of interests

None declared.

Acknowledgments

This work is sponsored by the National Natural Science Foundation of China under Grant no. 61703277 and the Shanghai Sailing Program (17YF1427000).

References

[1] L. Reinerman-Jones, G. Matthews, J.E. Mercado, Detection tasks in nuclear power plant operation: vigilance decrement and physiological workload monitoring, Saf. Sci. 88 (2016) 97–107.
[2] G.R.J. Hockey, A.W.K. Gaillard, O. Burov, Operator Functional State: The Assessment and Prediction of Human Performance Degradation in Complex Tasks, Ios Publ., the Netherlands, Amsterdam, 2003.
[3] J. Zhang, J. Xia, J.M. Garibaldi, P.P. Groumpos, R. Wang, Modeling and control of operator functional state in a unified framework of fuzzy inference petri nets, Comput. Methods Programs Biomed. 144 (2017) 147–163.
[4] J. Christensen, J. Estepp, G. Wilson, C. Russell, The effects of day-to-day variability of physiological data on operator functional state classification, Neuroimage 59 (1) (2012) 57–63.
[5] M. Grozdanovic, G.L. Janackovic, The framework for research of operators' functional suitability and efficiency in the control room, Int. J. Ind. Ergon. (2018). (in press) http://dx.doi.org/10.1016/j.ergon.2016.10.009.
[6] T. Heine, G. Lenis, P. Reichensperger, T. Beran, O. Doessel, B. Deml, Electrocardiographic features for the measurement of drivers' mental workload, Appl. Ergon. 61 (2017) 31–43.
[7] J.C. Lo, E. Sehic, K.A. Brookhuis, S.A. Meijer, Explicit or implicit situation awareness? Measuring the situation awareness of train traffic controllers, Transp. Res. Part F: Traffic Psychol. Behav. 43 (2016) 325–338.
[8] S. Charbonnier, R.N. Roy, S. Bonnet, A. Campagne, EEG index for control operators' mental fatigue monitoring using interactions between brain regions, Expert Syst. Appl. 52 (2016) 91–98.
[9] S.K.L. Lal, A. Craig, A critical review of the psychophysiology of driver fatigue, Biol. Psychol. 55 (2001) 173–194.
[10] I. Sasahara, N. Fujimura, Y. Nozawa, Y. Furuhata, H. Sato, The effect of histidine on mental fatigue and cognitive performance in subjects with high fatigue and sleep disruption scores, Physiol. Behav. 147 (2015) 238–244.
[11] K.C.H.J. Smolders, Y.A.W. de Kort, Bright light and mental fatigue: effects on alertness, vitality, performance and physiological arousal, J. Environ. Psychol. 39 (2014) 77–91.
[12] Y. Shigihara, M. Tanaka, A. Ishii, S. Tajima, E. Kanai, M. Funakura, Y. Watanabe, Two different types of mental fatigue produce different styles of task performance, Neurol. Psychiatry Brain Res. 19 (2013) 5–11.
[13] R. Parasuraman, M. Mouloua, R. Molloy, Effects of adaptive task allocation on monitoring of automated systems, Human Factor 38 (4) (1996) 665–679.
[14] J. Liu, C. Zhang, C. Zheng, EEG-based estimation of mental fatigue by using KPCA–HMM and complexity parameters, Biomed. Signal Process. Control 5 (2010) 124–130.
[15] E. Grandjean, Fitting the Task to the Man, Taylor and Francis, London, 1988.
[16] O. Okogbaa, R. Shell, D. Filipusic, On the investigation of the neurophysiological correlates of knowledge worker mental fatigue using the EEG signal, Appl. Ergon. 25 (1994) 355–365.

[17] O.N. Markand, Alpha rhythms, J. Clin. Neurophysiol. 7 (1990) 163–189.
[18] S.K.L. Lal, A. Craig, Driver fatigue: psychophysiological effects, Proceedings of the Fourth International Conference on Fatigue and Transportation, Australia, 2000.
[19] S. Makeig, T. Jung, Changes in alertness are a principal component of variance in the EEG spectrum, Neuroreport 7 (1995) 213–216.
[20] M. Simon, E.A. Schmidt, W.E. Kincses, M. Fritzsche, A. Bruns, C. Aufmuth, C. Bogdan, W. Rosenstiel, M. Schrauf, EEG alpha spindle measures as indicators of driver fatigue under real traffic conditions, Clin. Neurophysiol. 122 (2011) 1168–1178.
[21] C. Papadelis, Z. Chen, C. Kourtidou-Papadeli, P.D. Bamidis, I. Chouvarda, E. Bekiaris, N. Maglaveras, Monitoring sleepiness with on-board electrophysiological recordings for preventing sleep-deprived traffic accidents, Clin. Neurophysiol. 118 (2007) 1906–1922.
[22] V. Aleksandra, R. Vlada, C. Andrew, P. Dejan, Automatic recognition of alertness and drowsiness from EEG by an artificial neural network, Med. Eng. Phys. 24 (2002) 349–360.
[23] M. Kiymik, M. Akin, A. Subasi, Automatic recognition of alertness level by using wavelet transform and artificial neural network, J. Neurosci. Methods 139 (2004) 231–240.
[24] A. Yildiz, M. Akin, M. Poyraz, G. Kirbas, Application of adaptive neuro-fuzzy inference system for vigilance level estimation by using wavelet-entropy feature extraction, Expert Syst. Appl. 36 (2009) 7390–7399.
[25] V. Mervyn, X. Li, K. Shen, E. Wilder-Smith, Can SVM be used for automatic EEG detection of drowsiness during car driving, Saf. Sci. 47 (2009) 115–124.
[26] S. Vieira, W.H.L. Pinaya, A. Mechelli, Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: methods and applications, Neurosci. Biobehav. Rev. 74 (2017) 58–75.
[27] M. Hajinoroozi, Z. Mao, T.-P. Jung, C.-T. Lin, Y. Huang, EEG-based prediction of driver's cognitive performance by deep convolutional neural network, Signal Process. Image Commun. 47 (2016) 549–555.
[28] G.E. Hinton, S. Osindero, Y.W. Teh, A fast learning algorithm for deep belief nets, Neural Comput. 18 (7) (2006) 1527–1554.
[29] Z. Yin, J. Zhang, Cross-subject recognition of operator functional states via EEG and switching deep belief networks with adaptive weights, Neurocomputing 260 (2017) 349–366.
[30] C.L. Baldwin, B.N. Penaranda, Adaptive training using an artificial neural network and EEG metrics for within- and cross-task workload classification, Neuroimage 59 (2012) 48–56.
[31] Y. Ke, H. Qi, L. Zhang, S. Chen, X. Jiao, P. Zhou, X. Zhao, B. Wan, D. Ming, Towards an effective cross-task mental workload recognition model using electroencephalography based on feature selection and support vector machine regression, Int. J. Psychophysiol. 98 (2015) 157–166.
[32] J. Tang, C. Deng, G.-B. Huang, Extreme learning machine for multilayer perceptron, IEEE Trans. Neural Netw. Learn. Syst. 27 (2016) 809–821.
[33] G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, Extreme learning machine: theory and applications, Neurocomputing 70 (2006) 489–501.
[34] G.-B. Huang, X. Ding, H. Zhou, Optimization method based extreme learning machine for classification, Neurocomputing 74 (2010) 155–163.
[35] G.-B. Huang, L. Chen, C.K. Siew, Universal approximation using incremental constructive feedforward networks with random hidden nodes, IEEE Trans. Neural Netw. 17 (2006) 879–892.
[36] S. Lu, X. Qiu, J. Shi, N. Li, Z.H. Lu, P. Chen, M. Yang, F. Liu, W. Jia, Y. Zhang, A pathological brain detection system based on extreme learning machine optimized by bat algorithm, CNS & Neurol. Disord. Drug Targets 16 (2017) 23–29.
[37] Y. Zhang, G. Zhao, J. Sun, X. Wu, Z. Wang, H. Liu, V. Govindaraj, T. Zhan, J. Li, Smart pathological brain detection by synthetic minority oversampling technique, extreme learning machine, and Jaya algorithm, Multimed. Tools Appl. 6 (2017) 1–20, doi:10.1007/s11042-017-5023-0.
[38] J. Zhang, Z. Yin, R. Wang, Nonlinear dynamic classification of momentary mental workload using physiological features and NARX-model-based least-squares support vector machines, IEEE Trans. Hum.-Mach. Syst. 47 (2017) 536–549.
[39] J. Zhang, Z. Yin, R. Wang, Recognition of mental workload levels under complex human-machine collaboration by using physiological features and adaptive support vector machines, IEEE Trans. Hum.-Mach. Syst. 45 (2015) 200–214.
[40] C. Ting, M. Mahfouf, A. Nassef, D. Linkens, G. Panoutsos, P. Nickel, A. Roberts, G.R.J. Hockey, Real-time adaptive automation system based on identification of operator functional state in simulated process control operations, IEEE Trans. Syst. Man Cybern. – Part A: Syst. Hum. 40 (2010) 251–262.
[41] G.-B. Huang, H. Zhou, X. Ding, R. Zhang, Extreme learning machine for regression and multiclass classification, IEEE Trans. Syst. Man Cybern. – Part B: Cybern. 12 (2012) 513–529.
[42] M. Belkin, P. Niyogi, Laplacian Eigenmaps and spectral techniques for embedding and clustering, Advances in Neural Information Processing Systems, The MIT Press, Cambridge, MA, USA, 2002, pp. 585–591.
[43] Z. Yin, J. Zhang, Cross-session classification of mental workload levels using EEG and an adaptive deep learning model, Biomed. Signal Process. Control 33 (2017) 30–47.
[44] J.A.K. Suykens, J. Vandewalle, Least squares support vector machine classifiers, Neural Process. Lett. 9 (1999) 293–300.

Zhong Yin received his Ph.D. in control science and engineering from the East China University of Science and Technology. He has been a Lecturer at School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, China, since 2015. His research interests include intelligent human–machine systems, biomedical signal processing and pattern recognition.

Jianhua Zhang received his Ph.D. in Electrical Engineering from Ruhr-Universität Bochum, Germany in 2005. Between 2005 and 2006 he was a Research Associate at Intelligent Systems Research Laboratory, The University of Sheffield, UK. He has been a Full Professor of Control Systems Engineering at Department of Automation, East China University of Science and Technology since 2007. He was a Visiting Professor at Control Systems Group, Technische Universität Berlin, Germany in 2011 (for 6 months), 2012 (for a month), 2014 (for a month), and 2015 (for 2 months). His research interests are in computational intelligence, machine learning, intelligent systems and control, intelligent data modeling and analysis, pattern recognition, biomedical signal processing, modeling and control of complex engineering and biomedical systems, human–machine systems, brain-machine interaction, computational neuroergonomics, and cognitive neuroengineering. So far he has published one research monograph, three translated books, and more than 110 referred journal and conference proceedings papers.