Self-Paced Dynamic Infinite Mixture Model for Fatigue Evaluation of Pilots' Brains

Edmond Q. Wu, *Member, IEEE*, MengChu Zhou, *Fellow, IEEE*, Dewen Hu, *Senior Member, IEEE*, Longjun Zhu, Zhiri Tang, Xu-Yi Qiu, Ping-Yu Deng, Li-Min Zhu, *Member, IEEE*, and He Ren

***Abstract*—Current brain cognitive models are insufficient in handling outliers and dynamics of electroencephalogram (EEG) signals. This article presents a novel self-paced dynamic infinite mixture model to infer the dynamics of EEG fatigue signals. The instantaneous spectrum features provided by ensemble wavelet transform and Hilbert transform are extracted to form four fatigue indicators. The covariance of log likelihood of the complete data is proposed to accurately identify similar components and dynamics of the developed mixture model. Compared with its seven peers, the proposed model shows better performance in automatically identifying a pilot's brain workload.**

***Index Terms*—Brain fatigue, dynamic mixture model, electroencephalogram (EEG), machine learning, self-paced learning.**

Manuscript received June 21, 2020; revised September 20, 2020; accepted October 19, 2020. This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB1305101; in part by the National Natural Science Foundation of China under Grant 61671293 and Grant U1933125; in part by the Chinese Military Commission Equipment Development Department under Grant 61400030601 and Grant 6142219180403; and in part by the Jiangxi Province Key Research and Development Program under Grant 20192BBE50065. This article was recommended by Associate Editor C.-T. Lin. *(Corresponding authors: MengChu Zhou; Longjun Zhu; Edmond Q. Wu.)*

Edmond Q. Wu is with the Institute of Aerospace Science and Technology, Shanghai Jiao Tong University, Shanghai 200240, China, and also with the Science and Technology on Avionics Integration Laboratory, China National Aeronautical Radio Electronics Research Institute, Shanghai 200233, China (e-mail: edmondqwu@163.com).

MengChu Zhou is with the Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, NJ 07102 USA, and also with the Institute of Systems Engineering and Collaborative Laboratory for Intelligent Science and Systems, Macau University of Science and Technology, Macau 999078, China (e-mail: zhou@njit.edu).

Dewen Hu is with the College of Intelligence Science and Technology, National University of Defense Technology, Changsha 410073, China (e-mail: dwhu@nudt.edu.cn).

Longjun Zhu is with School of Artificial Intelligence, Shanghai Normal University Tianhua College, Shanghai, China (e-mail: zhulj09@126.com).

Zhiri Tang is with the School of Physics and Technology, Wuhan University, Wuhan 430072, China (e-mail: gerintang@163.com).

Xu-Yi Qiu and Ping-Yu Deng are with the Science and Technology on Avionics Integration Laboratory, China National Aeronautical Radio Electronics Research Institute, Shanghai 200233, China (e-mail: qiu_xuyi@careri.com; deng_pingyu@careri.com).

Li-Min Zhu is with the State Key Laboratory of Mechanical System and Vibration, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: zhulm@sjtu.edu.cn).

He Ren is with Shanghai Aircraft Design and Research Institute, COMAC, Shanghai 200120, China (e-mail: renhe@comac.cc).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCYB.2020.3033005.

Digital Object Identifier 10.1109/TCYB.2020.3033005

I. Introduction

A. Motivation

HIGH concentration of flight behavior can cause a pilot's brain fatigue. Fatigue is usually accompanied by a decline in decision-making ability, concentration, and reaction speed. During the flight, factors, such as lack of sleep, long flight time, and bad weather, may cause such fatigue. This may lead to operational errors or misjudgments, which may cause danger during flight. In 2007, the European Aviation Safety Agency (EASA) was the first to issue a regulation clause 25.1302 on human factors. In 2013, the Federal Aviation Administration (FAA) officially accepted the regulation 1302 of EASA and placed it in the amendment to the regulation 25–137. The regulations related to human factors verification of large aircraft are mainly CS25.1302 and CS25.1523. These two regulations focus on the smallest crew members and their safe operation.

The requirements related to human factors in the cockpit are explained in the China Civil Aviation Regulations (CCARs). CCAR 25.771 stipulates minimum flight crew fatigue, which requires pilots to control flight without excessive concentration and physical exertion. It can be seen that FAA, CCAR, and EASA have made a lot of explanations on pilot operation performance, but there are many problems in the actual implementation process, such as lack of detailed or accurate pilot cognitive measurement and analysis schemes. Therefore, it is of great significance to identify and alert a pilot's cognitive state objectively and accurately.

In order to record a pilot's physiological data on aircraft, it is necessary to develop onboard pilot physiological monitoring equipment. We should consider wearability and electromagnetic compatibility with airborne electronics.

B. Related Work

1) Brain Workload: At present, the evaluation methods of brain fatigue and workload are divided into two categories: 1) subjective and 2) objective ones [4], [10]. The former reflects pilots' state through the analysis of questionnaires, such as the NASA's Task Load indeX (TLX) or Scottish Walkability Assessment Tool (SWAT) questionnaires [5]. The latter is achieved by measuring human physiological parameters, such as eye tracking, hand movement, and electroencephalogram (EEG) signals. Compared with the former, the latter can quantify fatigue workload. At present, the working frequency of physiological signal recorders is far lower than

that of avionics, and the interval is at least three orders of magnitude. They have very weak electromagnetic interference on the latter. Therefore, the development of a portable wearable EEG recorder is feasible for measuring the brain load of a pilot.

EEG signals have been applied to assess human cognitive stages in different task learning [5]–[8], [12], [13], [18], [19], [30]. Zheng *et al.* [8] propose a tensor-based multitask learning (MTL) method to detect multiclass motor activities from them. Their results indicate that MTL is superior to the state-of-the-art techniques. Li *et al.* [12] propose a multisource transfer learning method for emotion recognition. They use limited labeled EEG data to train a transfer model. Their experimental results show that the three-category classification accuracy on benchmark SEED is improved with their method over a nontransfer method. Zhang *et al.* [13] propose a temporally constrained sparse group spatial pattern (TSGSP) algorithm to improve the classification accuracy of motor imagery EEG. A linear support vector machine classifier is used to learn EEG features to accurately identify motor imagery tasks. Three public EEG datasets are used to verify the effectiveness of TSGSP. Outstanding classification performance (averaged accuracies are 88.5%, 83.3%, and 84.3% for the three datasets, respectively) confirms that it is a promising candidate for improving the performance of motor imagery-based BCIs. Kwak and Lee [18] propose an error correction regression framework to improve the decoding accuracy of ear-EEG. Their results demonstrate that their BCI based on ear-EEG can achieve reliable performance. Zhang *et al.* [19] propose a learning framework of convolutional and recurrent neural networks to deal with spatial and temporal information of raw EEG streams. Extensive experiments on a movement intention EEG dataset (108 subjects) show that the proposed framework achieves high accuracy of 98.3%.

How to find the extent of a cognitive stage is a challenging task, which includes feature representation and learning and cognitive status detection. Existing studies have found that the change of their four rhythms can express human brain activity. The energy of feature waves is closely related to the level of fatigue [4], [5], [26], [27], [56]. Dong *et al.* [17] select discriminant features among EEG frequency bands and electrodes via a Fisher score. In particular, the time–frequency representation with Morlet wavelet shows a significant difference in the power of $\gamma, \beta,$ and α in the fronto-central region, and the power of θ in the centroparietal region. The best classification accuracy of 75.5% is obtained by a support vector machine classifier with features of γ in the frontocentral region. To assess and identify brain fatigue state using EEG signals, there are two methods. In the first method, the trend and changes of four rhythms, i.e., $\delta, \theta, \alpha,$ and β, are used to find evidence of fatigue. Many scholars have concluded that during fatigue, the activities of δ and θ increase while β decreases [23]–[25]. In the second category, the power spectrum of the four rhythms are developed as new indicators, such as $(\alpha + \theta)/\beta$, $(\theta + \delta)/(\alpha + \beta)$, α/β, and θ/β. Jap *et al.* [26] conclude that during fatigue, the change in α is less than that in β. Eoh *et al.* [27] find that the values of $(\alpha + \theta)/\beta$ and α/β increase significantly, while those of α and β decrease. Li *et al.* [28] demonstrate that the values of these four indicators obtain an increase after long-distance driving. These indicators show more fatigue information than four rhythms themselves. However, the above reports focus only on the FFT-based spectrum technique to form evaluation indicators [26], [27]. Since FFT can only provide the average spectrum information in a given period, these indicators cannot generate fatigue monitoring information in real time. Different from such prior work, this study aims to extract the instantaneous spectrum from EEG signals.

2) Mixture Model: Different learning methods have promoted the development of cognitive science in human brain. Xie *et al.* [7] propose a generalized hidden-mapping transudative learning method to achieve transfer learning of several classical intelligent models. Their experimental results prove the effectiveness of their proposed method in epilepsy EEG recognition. Zhang *et al.* [11] propose a spatial–temporal recurrent neural network (STRNN) to integrate feature learning into a unified spatial-temporal dependency model. Their results on the public emotion datasets show that STRNN is more competitive than the state-of-the-art methods. Zheng *et al.* [14] design a six-electrode placement scheme above the ears to collect EEG signals to increase the feasibility and wearability of EEG equipment in practical applications. Experimental results show that the mean accuracy of four emotions (happy, sad, fear, and neutral) is 85.11%. More work can refer to variational Bayesian Gaussian mixture model [15], Bayesian network [16], and deep neural networks [4], [5], [33].

To evaluate fatigue quantitatively, probability density estimation is often used to identify the cognitive stages of brain. Typical models include mixture models, such as a Gaussian mixture model (GMM). They can reveal the overall shape of the latent density. They have been successfully applied to the unsupervised learning of EEG signals [20], [21], [31], [32]. Solis *et al.* [20] use a GMM to separate sources from artifacts, and merge the clustering results into a probability hypothesis density filter to estimate the parameters for an unknown number of sources. Their simulation results demonstrate the effectiveness of this method in improving tracking accuracy for multiple neural sources. Zhang *et al.* [21] utilize a GMM to depict patients' imagery EEG distribution features in two different modules, and achieve a relatively higher discrimination accuracy. Zandi *et al.* [15] propose the measures of similarity and dissimilarity based on a variational Bayesian GMM. Their method is evaluated by using 561 h of scalp EEG, including a total of 86 seizures. A high sensitivity value of 88.34% is claimed.

Obviously, it is difficult to detect the latent state of raw EEG signals through a supervised method. We appeal to unsupervised methods to detect their latent attributes. However, GMMs are not sufficient to handle outliers and dynamics of EEG signals. Hybrid noise or singular points often emerge in practical problems. If GMMs are used to find the latent manifolds of these EEG signals, the estimates of their parameters can be severely affected by the outliers. This leads to an inaccurate model that uses a Gaussian distribution to

approximate unknown anomalous data. Gaussian distribution lacks the ability to approximate outliers. In addition, additional components are required to capture the tails of data. Thus we need a new distribution to replace Gaussian one. An elliptically symmetric student-t distribution is proposed as a base distribution of mixture model [36]. The resulting model is called a student mixture model (SMM) [37]–[42], and infinite student's t-mixture model (iSMM) [50]. The latter focuses on the modeling of an infinite SMM (iSMM). Reference [37]'s experimental results on simulated data and noisy speech data show that SMM is robust to additive noise. Nguyen and Wu [38] propose to extend the Student's-t distribution to form a finite bounded asymmetrical student's-t mixture model (BASMM), which includes GMM and SMM as its two special cases. This new distribution is flexible enough to accommodate observations of different shapes, such as non-Gaussian, nonsymmetric, and bounded support data. Nguyen *et al.* [42] present a new asymmetric mixture model for model detection and selection. Their asymmetric distribution is modeled with multiple student distributions. Clearly, SMM cannot automatically determine the proper number of components, which is important and can have significant impact on a learned model. Fortunately, iSMM can handle this issue [50].

With the above-mentioned methods, it is difficult to separate continuous clusters with high similarity. To solve this problem, we propose a dynamic detection method via a designed variance function of a mixture component. The variance function can improve the detection accuracy of data with high similarity. Our goal is to develop a dynamic infinite student t-mixture model with changeable variance of a mixture component for the attribute labels of a pilot's EEG signals. The model is expected to enhance the ability to identify brain fatigue states.

3) Self-Paced Learning: To cope with complexity of cognitive data, we propose a self-paced learning strategy to calculate thresholds by a loss function that characterizes the complexity of sample data, thereby helping the model learn from "simple data" to "complex data." In a self-paced learning strategy [46], the sample with smaller loss is considered as a "simple" sample, and the sample is selected during training; otherwise, it is deselected.

Another important issue is to detect data dynamic. It is difficult to distinguish the attributes of some points in the current and next clusters. So far, self-paced learning idea has been applied in the data processing of some supervised learning methods, such as support vector machine [49], [53], multilabel learning [48], and deep reinforcement learning [47], [54]. However, it has not yet been applied to unsupervised learning theory. This motivates us to build a self-paced unsupervised learning method. In this work, we propose a t-distribution dynamic mixture model with a self-paced strategy. It can detect the dynamic mechanism of overlapping data and achieve effective data separation. To our knowledge, this is the first work to incorporate the easy-to-hard strategy like humans into the learning process of a mixture model. Traditionally, the selection and determination of EEG data labels have been a complex issue. Expert experience plays an important role in labeling data. This may not lead to completely correct conclusions. Therefore, this study aims to provide a fully automatic model to complete the EEG signals labeling and brain state recognition.

C. Proposed Model

Considering the above points, we integrate a t-distribution-mixture-model with a self-paced learning strategy to propose a fatigue recognition framework for detecting and identifying brain load assessments.

We intend to develop a self-paced dynamic infinite mixture model (SDiMM) that can learn and detect different fatigue degrees from EEG signals, and observe the fatigue load through sub-band feature representations related to the main degree of dynamic fatigue changes. Therefore, it is very suitable for fatigue detection tasks. It can be used to obtain the components of EEG signals, thereby solving the problem of fatigue labeling.

To sum up, the main contributions of this study are: 1) it proposes instantaneous spectrum features-based fatigue measures; 2) it proposes an SDiMM to detect the attribute of EEG signals, and explore the overlapping detection between a current state and next one; and 3) it proposes a self-paced learning strategy to learn the model from simple data to "complicated data."

As shown in Fig. 1, this study computes four rhythms from pilots by using finite impulse response (FIR) filters. Then, an ensemble wavelet transform (EWT)-Hilbert transform is adopted to obtain their instantaneous spectrum. The extracted features are trained and labeled by the proposed SDiMM. The remainder of this article is organized as follows. Section II discusses fatigue feature extraction from EEG signals. Section III presents SDiMM. Experimental results are given in Section IV. The conclusion is drawn in Section V.

II. Feature Learning

From the construction theory of a wavelet function, we know how to build a tight frame set of empirical wavelets. EWT can be defined in the same way as that of classic wavelet transform [43]. The coefficients are given as the inner products with the empirical wavelets

$$W_f^{\varepsilon}(n,t) = \langle f, \psi_n \rangle = \int f(\tau)\overline{\psi_n(\tau - t)}d(\tau) = \left(\hat{f}(\omega)\overline{\widehat{\psi}_n(\omega)}\right)^{\vee} \tag{1}$$

$$W_f^{\varepsilon}(0,t) = \langle f, \emptyset_1 \rangle = \int f(\tau)\overline{\emptyset_1(\tau - t)}d(\tau) = \left(\hat{f}(\omega)\overline{\widehat{\emptyset}_1(\omega)}\right)^{\vee} \tag{2}$$

where $W_f^{\varepsilon}(n,t)$ is an empirical wavelet transform, f is an empirical mode, and ψ_n is a mother wavelet function. $W_f^{\varepsilon}(0,t)$ is an approximation coefficient

$$\widehat{\psi_n}(\omega) = \begin{cases} 1, & \text{if } (1+\gamma)\omega_n \le |\omega| \le (1-\gamma)\omega_n \\ \cos\left[\frac{\pi}{2}\beta\left(\frac{1}{2\gamma\omega_{n+1}}(|\omega| - (1-\gamma)\omega_{n+1})\right)\right] & \\ & \text{if}(1-\gamma)\omega_{n+1} \le |\omega| \le (1+\gamma)\omega_{n+1} \\ \sin\left[\frac{\pi}{2}\beta\left(\frac{1}{2\gamma\omega_n}(|\omega| - (1-\gamma)\omega_n)\right)\right] & \\ & \text{if } (1-\gamma)\omega_n \le |\omega| \le (1+\gamma)\omega_n \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

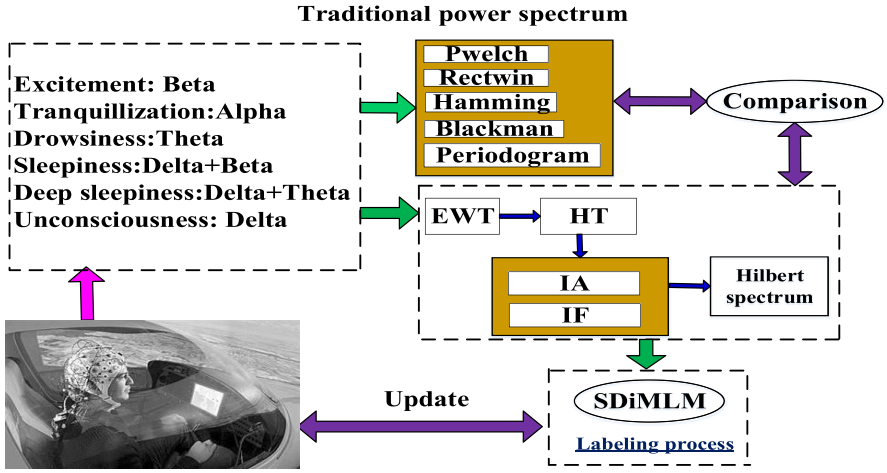

Fig. 1. Framework of SdiMM.

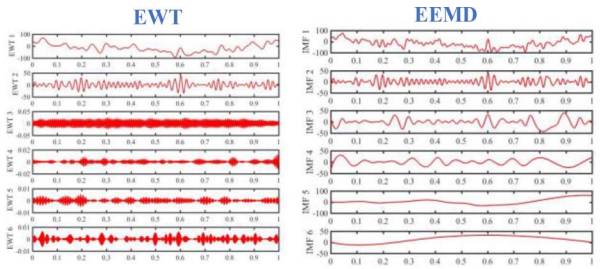

Fig. 2. Decomposition comparison between two empirical models.

and

$$\widehat{\phi_1}(\omega) = \begin{cases} 1, & \text{if } |\omega| \le (1-\gamma)\omega_n \\ \cos\left[\frac{\pi}{2}\beta\left(\frac{1}{2\gamma\omega_n}(|\omega| - (1-\gamma)\omega_n)\right)\right] & \\ \quad \text{if } (1-\gamma)\omega_n \le |\omega| \le (1+\gamma)\omega_n & \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

The reconstruction is obtained as

$$\begin{aligned} f(t) &= W_f^{\varepsilon}(0,t) * \emptyset_1(t) + \sum_{n=1}^{N} W_f^{\varepsilon}(n,t) * \psi_n(t) \\ &= \left(\widehat{W_f^{\varepsilon}}(0,\omega) * \widehat{\phi_1}(\omega) + \sum_{n=1}^{N} \widehat{W_f^{\varepsilon}}(n,\omega) * \widehat{\psi_n}(\omega)\right)^{\vee}. \end{aligned} \tag{5}$$

Following this formalism, we have the empirical mode:

$$\begin{cases} f_0(t) = W_f^{\varepsilon}(0,t) * \emptyset_1(t) \\ f_k(t) = \sum_{n=1}^{N} W_f^{\varepsilon}(k,t) * \psi_k(t). \end{cases} \tag{6}$$

The comparison of components is presented in Fig. 2 between EWT and ensemble empirical mode decomposition (EEMD). The former has stronger ability to detect the nature of a base function while the latter produces overlapping modes. EWT has more ability than EEMD to exhibit the details of EEG signals. Therefore, we select it to obtain instantaneous spectrum, which can provide more real-time information from EEG signals. Spectrum data from the Hilbert transform are the input of SDiMLM.

For each empirical mode $f_k(t)$, we have

$$y_k(t) = \frac{1}{\pi} p \int_{-\infty}^{+\infty} \frac{f_k(t)}{t-t'} dt \tag{7}$$

$$\begin{cases} \varphi_k(t) = \left[f_k^2(t) + y_k^2(t)\right]^{1/2} \\ \theta_k(t) = \arctan\left(\frac{y_k(t)}{f_k(t)}\right) \\ \omega_k = \frac{d\theta_k(t)}{dt} \end{cases} \tag{8}$$

where $y_k(t)$ is the Hilbert transform of $f_k(t)$. φ_k and $\theta_k(t)$ are, respectively, its instantaneous amplitude (IA) and phase from an empirical mode. ω_k is its instantaneous frequency (IF).

After performing HT at each empirical mode, we can express the data in the following form:

$$f(t) = \sum_{j=1}^{n} \varphi_k(t) \exp\left(i \int \omega_k(t) \mathrm{d}t\right). \tag{9}$$

This represents instantaneous amplitude and frequency as functions of time. This time–frequency distribution of the amplitude is defined as the Hilbert amplitude spectrum $H(\omega, t)$, or simply Hilbert spectrum. With the Hilbert spectrum defined, we can also define the marginal spectrum $h(\omega)$ as

$$h(\omega) = \int_0^T H(w,t) dt. \tag{10}$$

We can also define the instantaneous energy density as

$$I_E(t) = \int_0^T H^2(w,t) dw \tag{11}$$

and its marginal energy spectrum as

$$E_S(t) = \int_0^T H^2(w,t) dt. \tag{12}$$

Traditionally, we have the frequency indicators constructed from θ, α, and β as follows:

$$\begin{cases} (\alpha+\theta)/\beta \\ \alpha/\beta \\ (\alpha+\theta)/(\alpha+\beta) \\ \theta/\beta \end{cases}, \quad \text{where} \quad \begin{cases} \theta = \sum_{f=4}^{7} P_S(f) \\ \alpha = \sum_{f=8}^{13} P_S(f) \\ ß = \sum_{f=14}^{30} P_S(f). \end{cases} \tag{13}$$

Commonly, $P_s(f)$ denotes the magnitude of power spectral density at frequency f (Hz). FFT can only provide mean spectrum in a given time period; while HT can provide the instantaneous information of EEG signals. In this work, we use EWT-HT-based instantaneous spectrum to substitute the FFT-based spectrum. By monitoring instantaneous information, we can offer a real-time fatigue trend.

III. Self-Paced Dynamic Learning Model

This section presents the modeling process of SDiMM. In the self-paced learning algorithm [46], each iteration simultaneously selects easy samples and learns a new model' parameter. The number of samples selected is governed by a threshold that is evaluated by the loss function used in model training. SDiMM is proposed to find the latent mode of EEG fatigue. An expectation-maximization (EM) algorithm is proposed to learn its parameters. Moreover, the dynamic behavior between the current mode and next one in it is deeply discussed and modeled. Finally, an unsupervised learning model is proposed to deal with samples labeled by it. It consists of a self-paced iSMM and dynamic learning strategy.

A. Self-Paced Infinite Student Mixture Models

SMM is a latent variable model, which can find the attributes of each data point. In it, the latent stage of the observed data is difficult to determine. One way to overcome this difficulty is to establish an infinite mixture model. We propose a dynamic iSMM as follows:

$$\begin{cases} P(Y|\theta_s) = \sum_{j=1}^{N}\sum_{c=1}^{C}\lambda_{jc}\mathcal{S}\left(y_j|\mu_c, R_{jc}, v_{jc}\right) \\ \quad \sum_{c=1}^{C}\lambda_{jc} = 1 \\ \mathrm{Dir}\left(\lambda_{jc}|\eta\right) \sim \mathrm{Dirchlet}(\eta/c, \ldots, \eta/c) \\ \quad = \frac{\Gamma(\eta)\eta^C k}{\Gamma(\eta/c)}\prod_{c=1}^{c}\lambda_{jc}^{\eta/c-1} \\ \mathcal{S}\left(y_j|\mu_c, R_{jc}, v_{jc}\right) = \frac{\Gamma\left(\frac{d+v_{jc}}{2}\right)}{\Gamma\left(\frac{v_{jc}}{2}\right)(\pi v_{jc})^{\frac{d}{2}}}|R_c|^{\frac{1}{2}} \\ \quad \times\left[1+\frac{1}{v_{jc}}\left(y_j-\mu_c\right)^T R_c\mu_c\right]^{-\frac{d+v_{jc}}{2}} \end{cases} \tag{14}$$

where $\theta_s \equiv (\lambda_{j1}, \ldots, \lambda_{jC}, \mu_c, \ldots \mu_c, R_{j1}, \ldots R_{jC}, v_{j1}, \ldots v_{jC})$, parameter d is the feature dimension of EEG signals, μ_c is the mean of each group of EEG data, and R_c is precision in each group. $v_{jc} > 0$ is a tuning parameter to control the shape of t-distribution, which represents the degree of freedom. When v_{jc} decreases, the heavy tails are serious. As v_{jc} approaches infinity, and Student distribution becomes Gaussian distribution. $\Gamma(\cdot)$ represents a Gamma function.

In order to operationalize self-paced learning, we need a strategy for simultaneously selecting the easy samples and learning parameter θ at each iteration. The parameter update involves optimizing an objective function that depends on θ as

$$\theta_{t+1} = \underset{w\in R^d}{\mathrm{argmin}}\left(\sum_{i=1}^{n} f(y_i, \theta)\right) \tag{15}$$

where $f(\cdot,\cdot)$ is the negative log likelihood of EM. To model the self-paced learning strategy, a modified optimization problem with binary variable v_i that indicates whether the ith sample is easy or not, is proposed in this study. First, only easier samples contribute to the objective function. Then, we have the following mixed-integer program:

$$(\theta_{t+1}, v_{t+1}) = \underset{w\in R^d, v\in\{0,1\}^n}{\mathrm{argmin}}\left(\sum_{i=1}^{n} v_i f(y_i, \theta) - \frac{1}{K}\sum_{i=1}^{n} v_i\right) \tag{16}$$

where K is a weight that determines the number of samples to be considered: if K is large, the problem prefers to consider only "easy" samples with a small value of f (high likelihood, or far from the margin).

A subset is easy if current θ can be fit to the mixture model with a small f. We iteratively decrease the value of K in order to estimate parameters θ via self-paced learning. As K approaches 0, more samples are included until problem (16) reduces to problem (15). We thus begin with only a few easy examples, gradually introduce more until the entire training dataset is used.

To optimize problem (16), we note that it can be relaxed such that each variable v_i is allowed to take any value in the interval [0, 1]. This relaxation is tight; that is, for any value of θ, an optimum value of v_i is either 0 or 1 for all samples. If $f(x_i, y_i, \theta) < 1/K$, then $v_i = 1$ yields the optimal objective function value. Similarly, if $f(x_i, y_i, \theta) > 1/K$, then the objective is optimal when $v_i = 0$.

Given the set of observations $\{y_j\}$, the overall log likelihood for θ in SMM is

$$\log L(\theta) = \sum_{j=1}^{N}\log\left(\sum_{c=1}^{C}\lambda_c S\left(y_j\middle|\mu_c, \sum_c, v_c\right)\right). \tag{17}$$

To represent the attributes of EEG signals, we define label indicator vectors as $\{z_n\}_{n=1}^{N}$. Now, if we treat both y_j and z_{cj} as being known, we have the log likelihood of the whole sample data as follows:

$$\log L_C(\theta) = \sum_{j=1}^{N}\sum_{c=1}^{C} z_{cj}\left(\log\lambda_C + \log S\left(y_j\middle|\mu_c, \sum_c, v_c\right)\right). \tag{18}$$

Unfortunately, the above expression is very tough to deal with. We know that Gamma and Gaussian distributions form a conjugated pair. Moreover, student distribution has the same mean but precision different from that of Gaussian one, which is an infinite mixture of the latter. However, it does have a convenient factorization, i.e., we can present a Gamma-distributed random variable U (shape-rate parametrization)

$$U \sim \Gamma\left(\frac{v}{2}, \frac{v}{2}\right) \tag{19}$$

and a random variable Y whose distribution conditional on $U = u$ is Gaussian

$$Y|U \sim \mathcal{N}(\mu, \Sigma/u) \tag{20}$$

where U obeys Gamma distribution $\Gamma(\cdot)$, which only depends on a single parameter v. Y obeys normal Gaussian distribution $\mathcal{N}(\cdot)$.

The marginal distribution of Y is student t-distribution with location μ, scale $\mathbf{\Lambda}$, and degrees-of-freedom v. This gives us a joint distribution of Y and U as follows:

$$\begin{cases} S(Y|\mu\Sigma, v) = \int_0^{+8} \mathcal{N}(Y|\boldsymbol{\mu}, \Sigma/u)\mathcal{G}\left(u|\frac{v}{2}, \frac{v}{2}\right)du \\ \mathcal{N}(Y|\boldsymbol{\mu}, \Sigma/u) = (2\pi)^{-\frac{d}{2}}|\Sigma/u|^{\frac{1}{2}} \\ \quad \times \exp\left\{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^T\Sigma/u(\mathbf{y}-\boldsymbol{\mu})\right\} \\ \mathcal{G}\left(u|\frac{v}{2}, \frac{v}{2}\right) = \frac{(\frac{v}{2})^{\frac{v}{2}}}{\Gamma(\alpha)}u^{\frac{v}{2}-1}\exp\left(-\frac{v}{2}u\right) \end{cases} \tag{21}$$

where $\xi > 0$. $\mathcal{N}(Y|\boldsymbol{\mu}, \Sigma/u)$ is Gaussian distribution, and $\mathcal{G}(u|\alpha, \beta)$ is Gamma one.

We can obtain the log likelihood of the probability distribution function (PDF) of a multivariate Student-t distribution as follows:

$$\begin{aligned} \log S\left(y_j, u_j \middle| \mu_c, \sum_c, v\right) = &-\frac{1}{2}p\log(2\pi) - \frac{1}{2}\log\left|\frac{\sum_c}{u_j}\right| \\ &- \frac{1}{2}\left(y_j-\mu_c\right)'\left(\sum_c/u_j\right)^{-1} \\ &\times \left(y_j-\mu_c\right) - \log\Gamma\left(\frac{1}{2}v\right) \\ &+ \frac{1}{2}v\log\left(\frac{1}{2}v\right) + \frac{1}{2}v\left(\log u_j - u_j\right) \\ &- \log u_j. \end{aligned} \tag{22}$$

In (22), each data point y_j and each component c are known, and the scale variable u_j given z_{cj} is latent. Then, the latent variable model of the iSMM with an infinite number of components can be represented as

$$\begin{cases} p\left(z_j \mid \theta_{\mathbb{S}}\right) = \prod_{c=1}^{C}\lambda_c^{z_{jc}} \\ p\left(u_j \mid z_j, \theta_{\mathbb{S}}\right) = \prod_{c=1}^{C}\mathcal{G}\left(u_{jc}|\frac{v_c}{2}, \frac{v_c}{2}\right)^{z_{jc}} \\ p\left(y_j \mid u_j, z_j, \theta_{\mathbb{S}}\right) = \prod_{m=1}^{M}\mathcal{N}\left(y_i|\mu_i, u_{jc}\sum_c\right)^{z_{jc}} \end{cases} \tag{23}$$

where y_j and u_j are from the same component c, and each observation y_j is conditionally dependent on the latent variable and model parameters. Latent variable u_j is also conditionally dependent on latent variable z_j.

$U = \{u_j\}_{j=1}^N$ is a set of scale vectors. We can marginalize it over the latent variables and obtain (14). From this viewpoint, Bayesian inference formulation of iSMM is complete when a particular prior is proposed for the parameters. Obviously, we can select a suitable distribution as the conjugate prior of likelihood function (23).

In iSMM, the prior on the mixture proportions is from Dirichlet distribution $\mathrm{Dir}(z|\eta_0)$, and the joint prior of the mean and precision is from Gaussian–Wishart distribution $\mathcal{W}_G(\mu_m, \Lambda_m^{\pm}|\theta_{\mathcal{N}\mathcal{W}_0})$. $\{v_c\}_{c=1}^C$ has no conjugate prior. We can relax the restrictions of hyperparameters to obtain broad priors of model parameters. The resulting joint prior on model parameters can be defined again as

$$\begin{cases} p(\theta_{\mathbb{S}}|\mathrm{iSMM}) = \mathrm{Dir}(z|\eta_0)\prod_{m=1}^{M}\mathcal{W}_G\left(\boldsymbol{\mu}_m, \Lambda_m^{\pm}|\boldsymbol{\theta}_{\mathcal{N}\mathcal{W}_0}\right) \\ \mathrm{Dir}(z|\eta) = \frac{\Gamma(\eta)\eta^C}{\Gamma(\eta+N)}\prod_{c=1}^{c}\Gamma(N_c) \\ \mathcal{W}_G\left(\mu_m, \Lambda_m^{\pm}|\theta_{\mathcal{N}\mathcal{W}_0}\right) = \mathcal{N}\left(\mu_m|m_0, \eta_0\Lambda_m^{\pm}\right)\mathcal{W}\left(\Lambda_m^{\pm}|S_0^{\mp}, \upsilon_0\right) \\ \mathcal{W}\left(\Lambda^{\pm}|S^{\mp}, \upsilon\right) = G|\Lambda^{\pm}|^{\frac{\upsilon-d-1}{2}}\exp\left(-\frac{1}{2}\mathrm{tr}\left\{S^{\pm}\Lambda^{\pm}\right\}\right) \end{cases} \tag{24}$$

where G is a normalizing constant. $\mathcal{W}_G(\mu_m, \Lambda_m|\theta_{\mathcal{N}\mathcal{W}_0})$ is the joint distribution from Gaussian and Wishart distributions. $\mathrm{Dir}(z|\eta)$ is Dirichlet distribution, and $\mathcal{W}(\Lambda^{\pm}|S^{\mp}, \upsilon)$ is Wishart distribution.

By adding U as an additional latent variable, and substituting (22) into (18), we update the log likelihood of the whole sample data as

$$\begin{aligned} \log L_C(\theta) = \sum_{j=1}^{N}\sum_{c=1}^{C} z_{cj}\Bigg[&\log\alpha_C - \frac{1}{2}p\log(2\pi) \\ &- \frac{1}{2}\log\left|\frac{\Sigma_c}{u_j}\right| - \frac{1}{2}\left(y_j-\mu_C\right)'\left(\frac{\Sigma_c}{u_j}\right)^{-1} \\ &\times \left(y_j-\mu_C\right) - \log\Gamma\left(\frac{1}{2}v\right) \\ &+ \frac{1}{2}v\log\left(\frac{1}{2}v\right) \\ &+ \frac{1}{2}v\left(\log u_j - u_j\right) - \log u_j\Bigg]. \end{aligned} \tag{25}$$

In order to enforce the consistency of component locations across time, we introduce a prior on the location parameter that penalizes large changes over consecutive time steps. This prior has the following joint multivariate Gaussian distribution:

$$\begin{cases} p_{pri}\left(\boldsymbol{\mu}_c, \boldsymbol{Q}_c^{\pm}\right) = \prod_{t=2}^{T}N\left(\left(\mu_{ct}-\mu_{c(t-1)}\right)|u, \left(r\boldsymbol{Q}_c^{\pm}\right)^{-1}\right) \\ \quad \times W\left(\mathbf{Q}_c^{\pm}|\mathbf{S}^{\mp}, \nu\right) \\ \mathcal{W}\left(\boldsymbol{Q}_c^{\pm}|\mathbf{S}^{\mp}, \nu\right) = \frac{1}{G}\left|\boldsymbol{Q}_c^{\pm}\right|^{\frac{\upsilon-d-1}{2}}\exp\left(-\frac{1}{2}\mathrm{tr}\left(\mathbf{S}^{\pm}\boldsymbol{Q}_c^{\pm}\right)\right) \\ \mu_{ct} \sim \mu_{c(t-1)} + \mathcal{N}(u, Q) \end{cases} \tag{26}$$

where Gaussian parameters $(\boldsymbol{\mu}_c, \boldsymbol{Q}_c^{\pm})$ obey Gaussian–Wishart priors. $t \in \{1, \ldots, T\}$ denotes the time frame for an EEG signal. u is the mean of μ_{ct}, and r is the relative precision of μ_{ct}. ν is the number of degrees of freedom for $\boldsymbol{Q}_c^{\pm}$. $W(\boldsymbol{Q}_c^{\pm}|\mathbf{S}^{\mp}, \nu)$ is the Wishart distribution. $\boldsymbol{Q}_c^{\pm}$ is a $d \times d$ symmetric matrix of random positive variables. $\mathbf{S}^{\pm}$ is a positive-definite matrix of size $d \times d$, and its inverse matrix $\mathbf{S}^{\mp}$ represents the scale matrix for $\boldsymbol{Q}_c^{\pm}$. Then, if $\upsilon \geq d$, $\boldsymbol{Q}_c^{\pm}$ has a Wishart distribution with υ degrees of freedom. The dynamic behavior can be controlled by covariance matrix Q.

Next, we obtain the maximum *a posteriori* (MAP) estimate of the fitted parameters θ by maximizing the log posterior, which is equivalent (up to an additive constant) to

$$\begin{aligned} f(y_i, \theta) &= L(\theta) \\ &= \sum_{j=1}^{N}\log p\left(y_j \middle| \mu_c, \sum_c, v_c\right) \\ &\quad + \sum_{c=1}^{C}\log p_{\mathrm{pri}}\left(\mu_{ct}\middle|\mu_{c(t-1)}, Q\right). \end{aligned} \tag{27}$$

B. Expectation–Maximization (EM) for SDiMM

Compared with Gaussian distribution, we have a fixed solution for estimating the parameters of Student-t distribution based on the maximum-likelihood principle. EM can train an

approximate mixture model via u as a latent variable from a Gamma prior.

For convenience, assume that there is exactly one observation per step, to be relaxed later. The dynamic covariance Q is assumed known. Adding dynamic behavior, we can obtain the complete-data log likelihood as follows:

$$\begin{aligned} f(y_i,\theta) &= \log L_C(\theta) \\ &= \sum_{j=1}^{N}\sum_{c=1}^{C} z_{jc}\big(\log\lambda_c + \log\log p\big(y_j|\mu_c,\,\Sigma_c,\,v_c\big)\big) \\ &\quad + \sum_{j=1}^{N}\zeta_j\sum_{t=2}^{T}\sum_{c=1}^{C}\log p_{pri}\big(\mu_{ct}\big|\mu_{c(t-1)},\;Q\big). \end{aligned} \tag{28}$$

Therefore, we add additional latent variable U in (27) to have

$$\begin{aligned} \log L_C(\theta) &= \sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{c=1}^{C} z_{cj} \\ &\quad\times\Bigg[\log\lambda_c - \frac{1}{2}p\log(2\pi) \\ &\qquad - \frac{1}{2}\log\left|\sum_c\right| - \frac{1}{2}u_{jc_t}\big(y_{jc_t}-\mu_{jc_t}\big)' \\ &\qquad\times\sum_{c}^{-1}\big(y_{jc_t}-\mu_{jc_t}\big) \\ &\qquad - \log\Gamma\left(\frac{1}{2}\nu\right) + \frac{1}{2}\nu\log\left(\frac{1}{2}\nu\right) \\ &\qquad + \frac{1}{2}\nu\big(\log u_{jc_t} - u_{jc_t}\big) - \log u_{jc_t}\Bigg] \\ &\quad + \sum_{j=1}^{N}\zeta_j\sum_{t=2}^{T}\sum_{c=1}^{C} \\ &\quad\times\left[-\frac{1}{2}p\log(2\pi) - \frac{1}{2}\log|Q| \right. \\ &\qquad \left. - \frac{1}{2}\Big(\mu_{jc_t}-\mu_{jc_{(t-1)}}\Big)'Q^{-1}\Big(\mu_{jc_t}-\mu_{nc_{(t-1)}}\Big)\right]. \end{aligned} \tag{29}$$

Relying on the math in [45], we define the expression for the posterior membership as

$$\begin{cases} \tau_{cj} = P\big(Z_{cj}=1|y_j,\widehat{\theta}\big) = \dfrac{\widehat{\alpha}_c f_{mvt}\Big(y_j;\widehat{\mu}_c,\widehat{\sum_c},\nu\Big)}{f_{mot}\big(y_j;\widehat{\theta}\big)} \\ u_{cj} = E_{U_j}\big(U_j|Z_{cj}=1|y_j,\widehat{\theta}\big) = \dfrac{\nu+p}{\nu+\frac{1}{2}u_j(y_j-\widehat{\mu}_c)'\widehat{\sum_c}^{-1}(y_j-\widehat{\mu}_c)} \end{cases} \tag{30}$$

where f_{mot} is the PDF of mixture of t-distribution, and f_{mvt} is one of the multivariate t-distribution. u_{cj} turns out to be a correction term for the non-Gaussian-ness of observations. It accounts for the longer tails by weighting the far away points less. In the Gaussian limit ($\nu\to\infty$), we have $u_{cj}\to 1$.

In fact, there is no Z or U in the additional term. In other words, it does not affect the E-step. Considering the posterior membership, the new objective function is then

$$\begin{aligned} Q\big(\theta|\overline{\theta}\big) &= \sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{c=1}^{C}\tau_{jc_t} \\ &\quad\times\left[\log\lambda_c - \frac{1}{2}\log\left|\sum_c\right| - \frac{1}{2}u_{jc_t}\big(y_{j_t}-\mu_{jc_t}\big)' \right. \\ &\qquad \left.\times\sum_{\boldsymbol{C}}^{-\mathbf{1}}\big(y_{j_t}-\mu_{jc_t}\big)+\cdots\right] \\ &\quad + \sum_{j=1}^{N}\zeta_j\sum_{t=2}^{T}\sum_{c=1}^{C}\left[-\frac{1}{2}\Big(\mu_{jc_t}-\mu_{jc_{(t-1)}}\Big)'Q^{-1}\right. \\ &\qquad \left.\times\Big(\mu_{jc_t}-\mu_{jc_{(t-1)}}\Big)+\cdots\right]. \end{aligned} \tag{31}$$

Furthermore, we can obtain the compact expression of the above objective function by defining μ_{jc_0} and letting the second sum start at $t=1$, i.e.,

$$\begin{aligned} Q\big(\theta|\overline{\theta}\big) &= \sum_{j=1}^{N}\sum_{t=1}^{N}\sum_{c=1}^{C} \\ &\quad\times\left(\tau_{jc_t}\left[\log\alpha_c - \frac{1}{2}\log\left|\sum_c\right| \right.\right. \\ &\qquad \left. -\frac{1}{2}u_{ct}\big(y_{jc_t}-\mu_{jc_t}\big)'\sum_{\boldsymbol{C}}^{-\mathbf{1}}\big(y_{jc_t}-\mu_{jc_t}\big)+\cdots\right] \\ &\qquad -\frac{1}{2}\Big(\mu_{jc_t}-\mu_{jc_{(t-1)}}\Big)'Q^{-1}\Big(\mu_{jc_t}-\mu_{jc_{(t-1)}}\Big) \\ &\qquad \left.+\cdots\right). \end{aligned} \tag{32}$$

Unlike a standard SMM model, the M-step of the proposed model updates μ that depends on Σ. Thus, we first estimate μ by holding Σ constant, and then (as in the stationary case) use the updated μ to estimate Σ. Thus, our update equations for $\widehat{\lambda}_c$ and $\widehat{\sum}_c$ remain unchanged

$$\begin{cases} \widehat{\lambda}_c = \arg\max_{\lambda_c} Q\big(\theta|\overline{\theta}\big) = \dfrac{\sum_j\zeta_j\sum_{t=1}^{T}\tau_{ct}}{T} \\ \widehat{\sum}_c = \arg\max_{\sum_c} Q\big(\theta|\overline{\theta}\big) = \dfrac{\sum_{t=1}^{T}\tau_{ct}u_{ct}(y_t-\widehat{\mu}_c t)(y_t-\widehat{\mu}_{ct})'}{\sum_{t=1}^{T}\tau_{ct}}. \end{cases} \tag{33}$$

Each component c is independent of the rest. Updating $\hat{u}_t$ needs to solve the following optimization problem:

$$\begin{aligned} \min_{\{\mu_1,\ldots,\mu_T\}}\;\sum_{t=1}^{T}\Big[&\tau_t u_t(y_t-\mu_t)'\Sigma^{-1}(y_t-\mu_t) \\ &+ (\mu_t-\mu_{t-1})'Q^{-1}(\mu_t-\mu_{t-1})\Big]. \end{aligned} \tag{34}$$

In fact, (34) is an unconstrained quadratic optimization problem, which can be solved by inverting a single $T_p\times T_p$ matrix. Furthermore, a recursive algorithm with a block-tridiagonal structure is proposed to solve the problem with T inversions of $p\times p$ matrices by a Kalman filter [55].

Since we have multiple observations, the forward pass can be performed more efficiently [52]. Here, we derive μ as a dynamic program for solving an optimization problem related to (34). Its forward and backward pass algorithms [51], [52] are presented as follows.

Forward Pass: Let us introduce "cost-to-go" functions $J_{1|1}, \ldots, J_{T|T}$

$$J_{t|t}(\mu_t) = \min_{\{\mu_1,\ldots,\mu_{t-1}\}} \sum_{s=1}^{t} \Big[\tau_s u_s (y_s - \mu_s)' \Sigma^{-1} (y_s - \mu_s) + (\mu_s - \mu_{s-1})' Q^{-1} (\mu_s - \mu_{s-1}) \Big]. \tag{35}$$

$J_{t|t}$ indicates how our cumulative cost is affected by μ_t under the conditions of the optimal choice of $\mu_1, \ldots,$ and μ_{t-1}. We assume that these cost functions are quadratic, to be justified later. In terms of $\mu_{t|t}$ and a positive definite $P_{t|t}$, we have

$$J_{t|t}(\mu_t) = \big(\mu_t - \mu_{t|t}\big)' P_{t|t}^{-1} \big(\mu_t - \mu_{t|t}\big) + \text{constant}. \tag{36}$$

The values of $\mu_{t|t}$ and $P_{t|t}$ are given in the recursive update

$$\begin{cases} P_{t|t} = \big(\tau_t u_t \Sigma^{-1} + (P_{t-1|t-1} + Q)^{-1}\big)^{-1} \\ \mu_{t|t} = P_{t|t} \Big| \Big(\tau_t u_t \Sigma^{-1} y_t + \big(P_{t-1|t-1} + Q\big)^{-1} \mu_{t-1|t-1}\Big). \end{cases} \tag{37}$$

If $J_{t|t}$ is known (i.e., $\mu_{t|t}$ and $P_{t|t}$ are known), a new cost function $J_{t+1|t}$ that contains the dynamic penalty for the next time step can be defined as

$$\begin{aligned} J_{t+1|t}(\mu_{t+1}) &= \min_{\mu_t} \Big[J_{t|t}(\mu_t) + (\mu_{t+1} - \mu_t)' Q^{-1} (\mu_{t+1} - \mu_t) \Big] \\ &= \min_{\mu_t} \Big[\big(\mu_t - \mu_{t|t}\big)' P_{t|t}^{-1} \big(\mu_t - \mu_{t|t}\big) \\ &\qquad + (\mu_{t+1} - \mu_t)' Q^{-1} (\mu_{t+1} - \mu_t) \Big]. \end{aligned} \tag{38}$$

Differentiating (38) regarding μ_t and setting it to zero, we obtain the optimal choice of μ_t, we have

$$\mu_t = \operatorname*{argmin}_{\mu_t} [\cdots] = \Big(P_{t|t}^{-1} + Q^{-1}\Big)^{-1} \Big(P_{t|t}^{-1} \mu_{t|t} + Q^{-1} \mu_{t+1}\Big). \tag{39}$$

Backward Pass: Suppose that we know the values of $\mu_{t+1}, \ldots,$ and μ_T that minimize the objection function as

$$\begin{aligned} \{\widehat{\mu}_{t+1}, \ldots, \widehat{\mu}_T\} = \operatorname*{argmin}_{\{\widehat{\mu}_{t+1},\ldots,\widehat{\mu}_T\}} & \sum_{t=1}^{T} \\ & \times \Big[\tau_s u_s (y_s - \mu_s)' \Sigma^{-1} (y_s - \mu_s) \\ & \quad + (\mu_s - \mu_{s-1})' Q^{-1} (\mu_s - \mu_{s-1}) \Big]. \end{aligned} \tag{40}$$

Since the optimal $\mu_{t+1}, \ldots, \mu_T$ are known, they can be treated as constants

$$\begin{aligned} \widehat{\mu}_t &= \arg \min_{\mu_t} \sum_{s=1}^{T} \Big[\tau_s u_s (y_s - \mu_s)' \Sigma^{-1} (y_s - \mu_s) \\ &\qquad + (\mu_s - \mu_{s-1})' Q^{-1} (\mu_s - \mu_{s-1}) \Big] \\ &= \arg \min_{\mu_t} \sum_{s=1}^{t} \Big[\tau_s u_s (y_s - \mu_s)' \Sigma^{-1} (y_s - \mu_s) \\ &\qquad + (\mu_s - \mu_{s-1})' Q^{-1} (\mu_s - \mu_{s-1}) \Big] \\ &\quad + (\widehat{\mu}_{t+1} - \mu_t)' Q^{-1} (\widehat{\mu}_{t+1} - \mu_t) + \text{constants}. \end{aligned} \tag{41}$$

Fig. 3. Flight simulator and monitoring system.

Furthermore, we can determine the optimal value of μ_t in (39) as

$$\begin{aligned} \widehat{\mu}_t &= \operatorname*{argmin}_{\mu_t} \Big[J_{t|t}(\mu_t) + (\widehat{\mu}_{t+1} - \mu_t)' Q^{-1} (\widehat{\mu}_{t+1} - \mu_t) \Big] \\ &= \Big(P_{t|t}^{-1} + Q^{-1}\Big)^{-1} \Big(P_{t|t}^{-1} \mu_{t|t} + Q^{-1} \widehat{\mu}_{t+1}\Big) \\ &= \Big(I - P_{t|t}\big(P_{t|t} + Q\big)^{-1}\Big) \mu_{t|t} + \Big(I - Q\big(P_{t|t} + Q\big)^{-1}\Big) \widehat{\mu}_{t+1} \\ &= \mu_{t|t} + P_{t|t}\big(P_{t|t} + Q\big)^{-1} \big(\widehat{\mu}_{t+1} - \mu_{t|t}\big). \end{aligned} \tag{42}$$

To initialize the recursion, we minimize the overall objective function $J_{T|T}$. The initial $\widehat{\mu}_T$ is set as

$$\widehat{\mu}_T = \operatorname*{argmin}_{\mu_T} J_{T|T}(\mu_T) = \mu_{T|T}. \tag{43}$$

Then, we use (42) to obtain $\{\widehat{\mu}_1, \ldots, \widehat{\mu}_T\}$. Repeating for each component c is used to update $\widehat{\mu}_{ct}$ in the M-step. About the inference of these updates, we can refer to [51] and [52].

IV. Experiment

A. Experimental Setup

To collect a pilot's physiological signals during flight simulation, the experiments are implemented in a 6-DOF full flight simulator, as shown in Fig. 3. It provides a virtual environment for real-time flight simulation with motion cockpit, high-frame-rate vision, animated audio, and haptic controls with high fidelity and good immersion. The adopted flight dynamic model is a C919 commercial airplane.

Participants: We invite 40 Chinese male pilots. Before the test, all participants flew for 15–25 h per week. During the test, they are paired with 20 pairs of flight crew, one in each crew as pilot flying and the other as pilot monitoring. All pilots pass the routine medical examination, fully rest before the mission. EEG signals are recorded by a 64-channel EEG system during flight simulation. The signals are sampled at 160 Hz. Every pilot is required to operate the flight simulator. They go through four flight load phases, as shown in Fig. 4. In each phase, the effective acquisition time is half an hour. They are divided into ten equal parts, and the middle eight equal parts are selected to form eight datasets. In each dataset, each flight phase contains a 3-min EEG signal.

How to choose an effective channel group is important for assessing the stage of fatigue workload. Chakraborty *et al.* [22] use the standard common spatial pattern filter to generate the optimal electrode set. Their method yields excellent results in electrode selection with respect to classification accuracy of cognitive tasks. Qi *et al.* [25] propose a spatiotemporal-filtering-based channel selection (STECS) algorithm, which

TABLE I
EEG FREQUENCY AND CORRESPONDING DESCRIPTIONS

WAVE	Frequency	Status	Amplitude of wave	Electrode	Electrode position
δ	1~4Hz	unconsciousness	20~200	T	41,42,43,44
				O	61,62,63
θ	4~8Hz	drowsiness	100~150	T	41,42,43,44
				P	47,48,49,50,51,52,53,54,55
α	8~9 Hz (slow α)	mental confusion	20~100	O	61,62,63
	9~12 Hz (middle α)	work with inspiration			
	12~14 Hz (fast α)	wide awake			
β	14~30Hz	Excitement	5~20	F	30,31,32,33,34,35,36,37,38
				P	47,48,49,50,51,52,53,54,55
δ+θ	1~8Hz	deep sleepiness	20~200	PO	56,57,58,59,60
δ+β	1~4Hz and 14~30Hz	sleepiness	20~200	TP	45,46
				TF	39,40

Note: T: Temporal Lobe; O: Occipital lobe; P: Parietal lobe; F: Frontal lobe

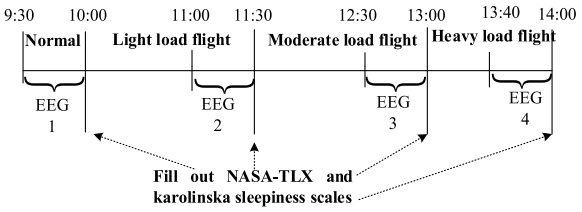

Fig. 4. EEG collection process.

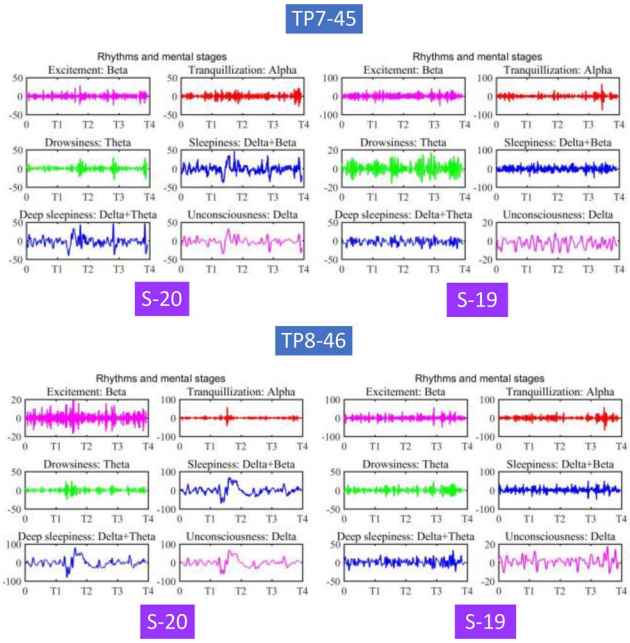

Fig. 5. Rhythms and its combination.

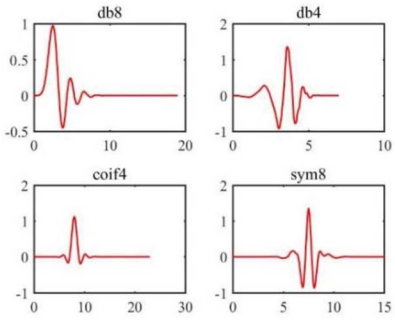

Fig. 6. Typical wavelet function.

can automatically identify a designated number of discriminative channels. The performance of STECS is assessed on three MI EEG datasets. Their results indicate that STECS can yield comparable classification performance in a half of the channels. We refer to the relevant literature on the cognitive evaluation of brain fatigue and find that there are many options for the topological location of the area of interest [32]. The corresponding experimental results also verify the function of the relevant brain area. In these achievements, the electrodes in the five regions of T, P, F, TP, and TF are the most commonly used. Therefore, this work selects them to form our cognitive indicators. Here, we present the rhythms and their corresponding mental stages on four electrodes of regions-of-interest in Table I and Fig. 5. "S" represents a pilot in the following figures.

B. Feature Learning

Because EEG signals are weak electrical signals, they are easily polluted by various noise in a device during their collection process. Therefore, we must remove the noise to reduce its impact on subsequent feature extraction. Some reports point out that wavelet transform can only reduce low frequency noise, while wavelet packet transform can reduce both high frequency and low frequency noise [4], [5]. Wavelet noise reduction usually uses a threshold to control the wavelet coefficients to achieve effective noise removal. It has two key issues. One is the choice of a suitable wavelet base, and the other is the choice of a suitable threshold. Related work can refer to [4] and [5]. In the process of collecting EEG signals, the effects of electromyogram (EMG) and electrooculogram (EOG) are also mixed in EEG signals. Some of them are distributed in high frequency parts, such as EMG signals, and some parts are in low frequencies, such as EOG signals. Here, we introduce some typical wavelet function visualization, such as db8, db4, sym8, and coiflet4 in Fig. 6. The waveform of the Db8 wavelet base is closest to that of the EEG signal. Obviously, db8 is a good choice to decompose EEG signal, because it can reduce the noise from the high and low frequency parts of EEG signal at the same time. The reconstructed wavelet coefficients are used to recover the denoised EEG signals. Four FIR filters are used to extract four rhythms from the denoised EEG signals. Their parameter settings are shown in Table II. Through bandpass filters, we obtain the EEG signal band of 0.1–30 Hz, which does not

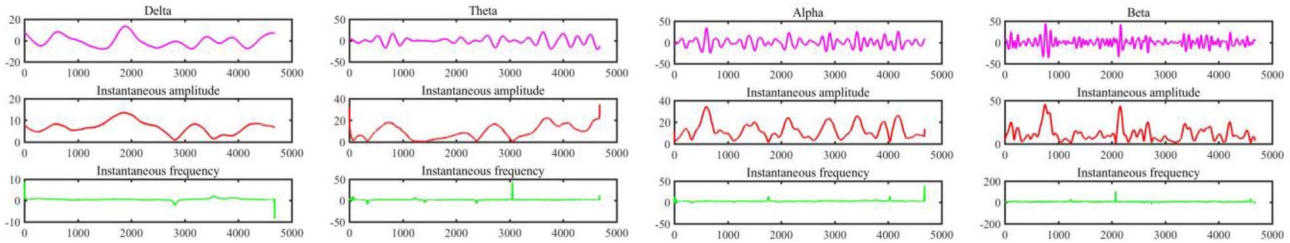

Fig. 7. Rhythm waves and their instantaneous amplitude and frequency.

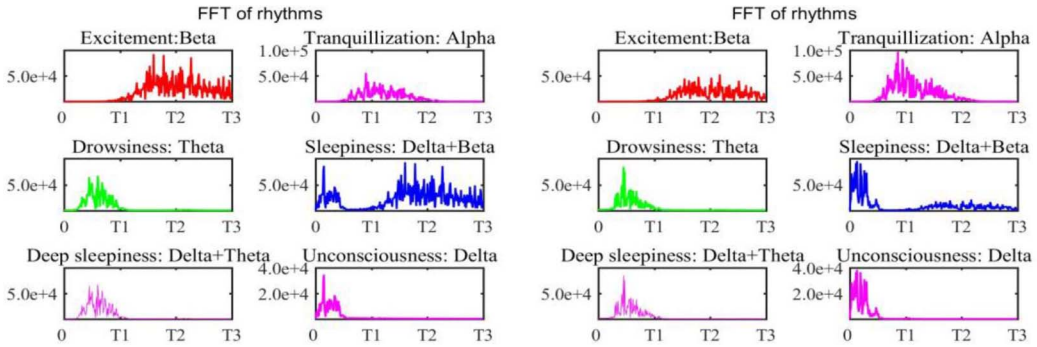

Fig. 8. FFT of rhythms.

TABLE II
FIR FILTER PARAMETERS SETUP

	FSF	FPF	SPF	SSF	FSA	PR	SSA	DF
δ	0.1	0.5	3	3.5	0.001	0.0575	0.0001	20
θ	3.5	4	7	7.5	0.001	0.0575	0.0001	20
α	7.5	8	12	12.5	0.001	0.0575	0.0001	20
β	12.5	12.5	30	30.5	0.001	0.0575	0.0001	20

Note: First Stopband Frequency (FSF), First Passband Frequency (FPF), Second Passband Frequency (SPF), Second Stopband Frequency (SSF), First Stopband Attenuation (FSA), Passband Ripple (PR), Second Stopband Attenuation (SSA), Density Factor (DF)

include the Gamma rhythm. In this way, interference, such as EMG can be removed.

EWT is used to decompose four rhythms. Then, we obtain their instantaneous amplitudes and frequencies as shown in Fig. 7. In addition, we can compute Hilbert amplitude spectrum and marginal energy spectrum of the featured wavebands according to (10). Traditionally, the power spectrum curve area of these four rhythms is obtained via FFT technology, while we use different time–frequency analysis techniques, such as HT. In Fig. 8, an FFT of rhythms from two channels is presented. They are similar in the frequency domain and difficult to distinguish. In Fig. 9, Hilbert marginal spectrum of four rhythms is shown. They are clearly different. Compared with power spectrum from FFT, they have richer information of brain workload. Obviously, the EWT-based spectrum has more frequency information than the FFT-based one. In this work, we use the instantaneous spectrum as the input of the models.

C. Detecting Cognitive Status

SDiMM is implemented in a MATLAB 2017a environment. To prove its effectiveness, tenfold cross-validation is used for the model. The purpose of this study is to identify the latent status of a pilot's mental status as determined by SDiMM. The four indicators of fatigue detection, such as $(\alpha+\theta)/\beta$, α/β, $(\alpha+\theta)/(\alpha+\beta)$, and θ/β, are used to judge the psychological state of the pilot. These mental stages contain wide awake (nonfatigue), mental confusion (microfatigue), and drowsiness (extreme-fatigue).

In order to test the validity of our model, we have intentionally added some noise and outlier data points in some benchmark data to test SDiMM as shown in Table III. The clustering quality is evaluated through WSS and BSS, where WSS stands for within sum of squares, and BSS stands for between sum of squares. For the results from clustering, smaller WSS and larger BSS imply better classification performance. The smaller WSS/BSS, the better clustering ability of the model.

As shown in Table III, our model has the best WSS/BSS among the five mixture models. MSE of our model is also the best. In particular, in datasets Iris and Gauss, SMM and iSMM provide weaker detection performance compared to GMM. Perhaps, the components of these two datasets contain more Gaussian components. Note that GMMs are more suitable for the detection and identification of components with Gaussian distribution. Clearly, the proposed SDiMM model can deal with some significant abnormal patterns.

Next, we use a Treelet transform [29] to reduce the correlation of these spectrum data. Treelet is a novel construction of multiscale-based data that extends wavelets to nonsmooth signals. It is well suited as a dimensionality reduction and feature selection tool. The results from Treelet transform are shown in Fig. 10. These processed features are more prominent. Obviously, the difference in these spectrum data can be clearly found. They can be utilized as the input of these mixture models. Furthermore, Hilbert marginal spectrum data can be obtained as shown in Fig. 11. These features have more ability to indicate the latent state of pilot brain workload. In particular, the overlap phenomenon of these features is obvious. In

TABLE III
PERFORMANCE COMPARISON AMONG DIFFERENT MIXTURE MODELS

Benchmark	index	GMM	iGMM	SMM	iSMM	SDiMM
Thyroid	MSE	4.0578	3.8657	3.7525	**3.5180**	**3.3902**
	WSS/BSS	4.0767	3.6132	3.0863	**2.8430**	**2.7654**
Seed	MSE	5.5924	4.6510	4.4052	**3.6713**	**3.1729**
	WSS/BSS	1.4538	1.2091	0.9143	**0.8716**	**0.6586**
Iris	MSE	3.4512	3.8560	4.8157	**4.4077**	**1.3858**
	WSS/BSS	1.3682	1.2639	3.7963	**0.7846**	**0.5154**
Gauss	MSE	2.0995	2.5613	3.1042	**2.8178**	**0.1479**
	WSS/BSS	1.2756	1.0567	4.0023	**0.5182**	**0.1298**
Bezdekiris	MSE	3.6458	3.5398	2.7734	**2.4307**	**0.9277**
	WSS/BSS	1.4322	1.3927	1.0250	**0.6726**	**0.3553**

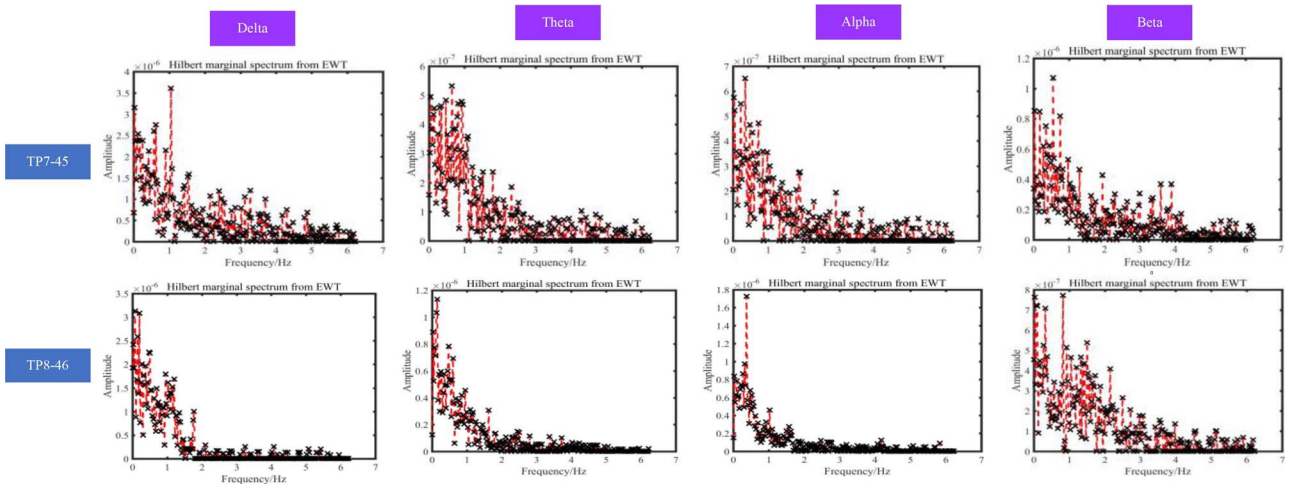

Fig. 9. Hilbert marginal spectrum.

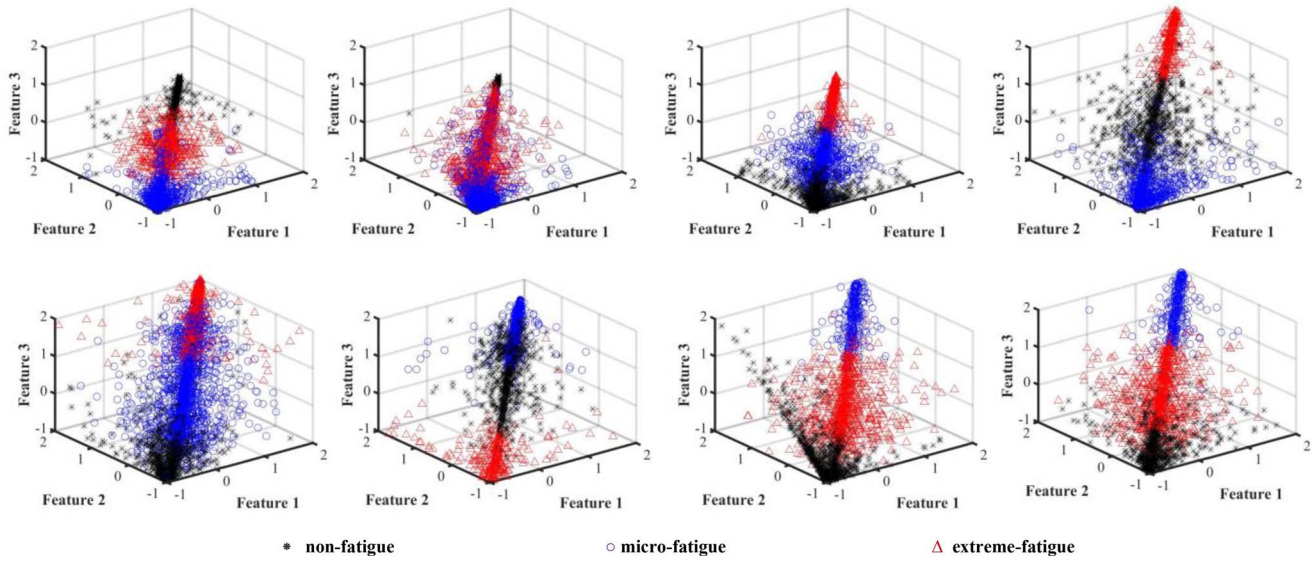

Fig. 10. Feature decorrelation results from eight EEG datasets.

addition, outliers of these features are also observed. Eight models are used to train them. The corresponding results are in Table IV and Fig. 12. Eight datasets present an average recognition accuracy of 91.78%, which is the average of the detection accuracy from the 8 datasets.

In Table IV, the *k*-means method delivers good performance on all datasets. However, it lacks the ability to detect similar overlapping patterns. For the type of hidden state of a dataset, *a priori* setting is required. Therefore, its overall learning ability is inferior to the infinite mixture models. The detection ability of iGMM is better than the finite GMM. Its recognition ability is usually very close to SMM. This also shows that the ability of SMM to learn abnormal pattern points is stronger than that of GMM. On DATA 3, 5, and 6–8, the

TABLE IV
PERFORMANCE COMPARISON AMONG DIFFERENT CLUSTER METHODS

	index	k-means	DBSCAN	HMM	GMM	iGMM	SMM	iSMM	SDiMM
DATA1	MSE	33.56	40.08	82.36	99.32	33.43	33.57	**19.66**	**19.32**
	WSS/BSS	0.44	0.72	2.99	9.36	0.41	0.42	**0.23**	**0.22**
	Time (s)	10.15	11.04	14.64	23.73	27.90	40.85	**53.03**	**58.8**
DATA2	MSE	49.01	35.63	94.44	103.40	35.70	35.87	**25.96**	**24.09**
	WSS/BSS	0.85	0.50	7.98	33.74	0.49	0.50	**0.35**	**0.31**
	Time (s)	10.15	11.20	16.47	21.02	48.27	30.66	**35.57**	**56.88**
DATA3	MSE	35.01	47.96	89.38	92.51	35.25	35.06	**25.47**	**20.88**
	WSS/BSS	0.47	0.77	4.32	5.34	0.44	0.44	**0.34**	**0.25**
	Time (s)	10.15	10.34	24.65	38.99	42.04	48.04	**57.81**	**50.29**
DATA4	MSE	36.18	49.22	94.89	100.11	34.71	36.24	**35.56**	**35.12**
	WSS/BSS	0.51	0.81	8.25	15.69	0.43	0.49	**0.48**	**0.44**
	Time (s)	10.17	11.18	16.97	31.53	47.36	45.82	**48.51**	**54.11**
DATA5	MSE	36.23	36.31	96.26	91.64	36.34	36.26	**26.21**	**24.09**
	WSS/BSS	0.52	0.52	9.47	6.15	0.49	0.49	**0.34**	**0.31**
	Time (s)	10.17	10.25	25.50	30.42	47.70	43.03	**51.86**	**56.00**
DATA6	MSE	110.02	49.22	70.68	95.36	34.71	34.62	**29.31**	**20.67**
	WSS/BSS	497.70	0.81	1.80	6.52	0.43	0.43	**0.39**	**0.24**
	Time (s)	10.14	11.27	23.02	31.12	47.36	49.38	**49.92**	**52.38**
DATA7	MSE	36.24	34.73	88.34	90.29	34.80	34.69	**26.21**	**20.95**
	WSS/BSS	0.46	0.46	4.08	4.60	0.43	0.43	**0.34**	**0.25**
	Time (s)	10.15	11.17	17.02	32.63	37.36	33.06	**48.24**	**43.14**
DATA8	MSE	48.30	49.52	101.02	40.26	38.35	38.17	**29.30**	**27.13**
	WSS/BSS	0.80	0.85	13.72	0.60	0.54	0.51	**0.38**	**0.35**
	Time (s)	10.46	11.15	15.55	20.76	35.48	37.03	**48.21**	**56.34**

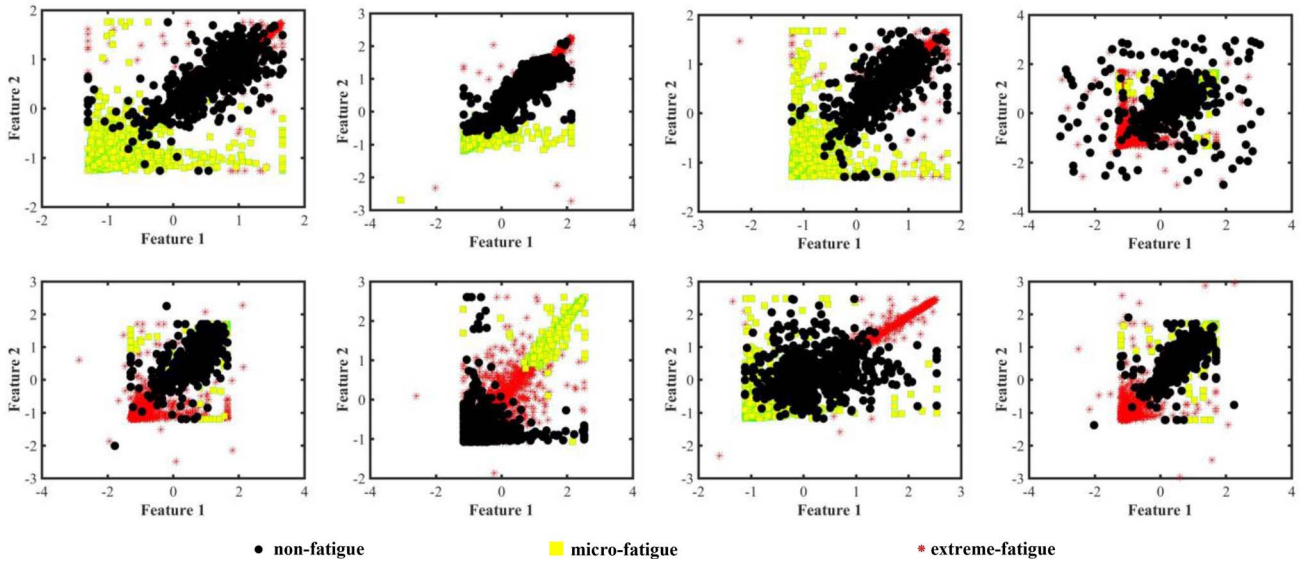

Fig. 11. Feature representation from eight EEG datasets.

detection capability of SDiMM is more than that of iSMM, which indicates that these five datasets have more similar cognitive behaviors. It is difficult for iSMM to distinguish them. The covariance of the joint distribution of SDiMM can effectively learn continuous similar cognitive behaviors. HMM does not perform well on these datasets, perhaps due to its arbitrarily given initial parameters of the model. In addition, HMM is suitable for single-sequence hidden state learning. For the reasoning of multiple-input hidden state reasoning, its ability seems insufficient. The running time of k-means and DBSCAN is very close. The running time of the infinite mixture models is more than that of the finite mixture models, because *a priori* reasoning of the model parameters takes more time. The running time of HMM is greater than that of k-means and DBSCAN, and less than that of the infinite mixture models.

As a dynamic infinite mixture model, SDiMM can accurately detect the nature of data, which is attractive and very useful. The latent manifolds of brain workload can be captured by some mixture models. SDiMM has the highest test accuracy in assessing the workload of pilot brains. It enjoys a competitive advantage with other hybrid models (such as GMM, iGMM, SMM, and iSMM) and is able to detect the latent fatigue manifolds of EEG signals.

D. Discussion

However, many researchers have not selected the central electrode as the area of interest for fatigue detection. We want to explore the contribution of the brain center position to fatigue detection. On the basis of the original channels, $C1$ and $C2$ are added, respectively, to form three new datasets,

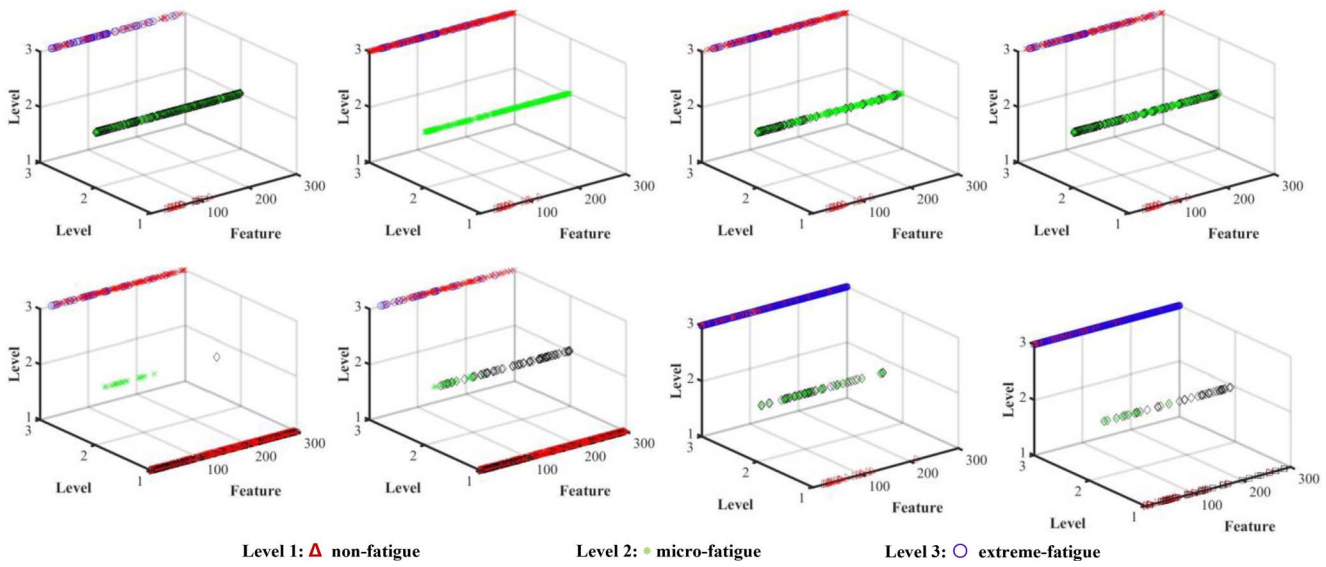

Fig. 12. Cognition detection results from eight datasets.

TABLE V
COMPARISON OF COGNITIVE ABILITIES IN BRAIN REGIONS OF INTEREST USING WSS/BSS

DATA1	DATA2	DATA3	DATA4	DATA5	DATA6	DATA7	DATA8
0.22	0.31	**0.25**	0.44	0.31	**0.24**	**0.25**	0.35
DATA1+C1	DATA2+C1	DATA3+C1	DATA4+C1	DATA5+C1	DATA5+C1	DATA7+C1	DATA8+C1
0.24	**0.23**	0.33	0.33	**0.33**	0.32	0.25	**0.31**
DATA1+C2	DATA2+C2	DATA3+C2	DATA4+C2	DATA5+C2	DATA6+C2	DATA7+C2	DATA8+C2
0.40	**0.23**	0.46	**0.32**	0.35	0.32	0.26	0.43
DATA1+C1+C2	DATA2+C1+C2	DATA3+C1+C2	DATA4+C1+C2	DATA5+C1+C2	DATA5+C1+C2	DATA7+C1+C2	DATA8+C1+C2
0.41	0.24	0.35	0.35	0.36	0.36	0.27	0.45

namely, DATA + $C1$, DATA + $C2$, and DATA + $C1$ + $C2$. The test results in Table V show that the four datasets without central brain electrodes can obtain higher detection accuracy than the dataset with central electrode positions. The detection accuracy of the dataset containing $C1$ is slightly higher than that of the dataset containing $C2$. However, the overall performance of the dataset containing $C1 + C2$ has decreased. Obviously, the electrodes in the center of the brain have not improved the recognition ability of the SDiMM model.

Attribute labeling of a brain cognitive state is one of the hot topics in brain science in recent years. For unlabeled EEG data, unsupervised methods, such as K-means, GMM, and DBSCAN, are often used to obtain their attribute categories, form labels of EEG data, and obtain the level of pilot fatigue cognitive state. These methods ignore the dynamic mechanism of brain cognition and lack precise division of overlapping behaviors with similar modalities in brain cognition. The proposed joint distribution of SMM can be used to model the dynamic behavior of brain cognition. Through covariance of joint distribution, similar patterns of brain cognition can be accurately detected. This is the advantage of dynamic mixture models over deep learning models.

This study investigates the instantaneous spectrum features provided by the EWT-HT technique, and proposes novel indicators to characterize the workload of the brain. The attributes of EEG signals can be detected automatically and labeled by the proposed SDiMM. Our experimental results indicate that the self-paced dynamic iSMM and SMM can well identify pilots' fatigue state, and have high practical value. Compared with GMMs and SMMs model, SDiMM has excellent ability to identify the level of pilots' fatigue. Therefore, pilot fatigue identification based on the EWT-HT-SDiMM model is effective.

V. CONCLUSION

We have addressed the problem of brain workload evaluation and proposed a solution within the framework of self-paced learning strategy and student t-distribution-mixture-model. It is found that SMM and SDiMM possess higher robustness and yield better accuracy than other methods, such as GMM and k-means. Experimental results on pilots' fatigue status recognition demonstrate higher robustness and accuracy than other techniques in the literature. Noise-performance analysis shows that SDiMM is more robust when facing outlier of data points. The testing results from five benchmark datasets and practical EEG data indicate that SDiMM has stronger ability to detect the latent manifolds of sample data. The advantages of the proposed method can be further summarized as follows.

1) It can help the model learn from "easy data" to complicated data. Our method simultaneously selects easy samples and updates the parameters of SDiMM.
2) It is able to model the dynamic behavior of continuous sample segments, and provide the attribute identification of overlapping segments via a changeable covariance matrix.

3) It builds an unsupervised multilayer learning network to extract the abstract feature representation of workload and provide the higher accuracy model to detect real-time brain status with new EEG signals.

In summary, the proposed model realizes one complete evaluation procedure of pilots' brain workload from the pretreatment of raw EEG signals to latent status classification of workload. The proposed workload evaluation system is a promising approach for pilots monitoring and a good addition to investigation on quantified brain activities for rhythm activity disorders or cognition behavioral changes, such as drowsiness disorder analysis.

A possible extension of this work is to study the use of other instantaneous features and dynamic behavior modeling techniques, such as SMM with a hidden Markov model. In addition, the current version of an EEG device is not convenient for pilots to have on their head during aviation. A helmet-mounted EEG recorder should be developed for pilots to wear and attention should be paid to its electromagnetic compatibility with avionics in aircraft. Recently proposed methods [57], [58] should be explored to accelerate the training of the proposed models.

REFERENCES

[1] J. A. Caldwell, M. M. Mallis, J. L. Caldwell, M. A. Paul, J. C. Miller, and D. F. Neri, "Fatigue countermeasures in aviation," *Aviation Space Environ. Med.*, vol. 80, no. 1, pp. 29–59, 2009.

[2] C. Reis, C. Mestre, H. Canhão, D. Gradwell, and T. Paiva, "Sleep complaints and fatigue of airline pilots," *Sleep Sci.*, vol. 9, no. 2, pp. 73–77, 2016

[3] *Implementation Guide of Fatigue Risk Management System (FRMS) for the Civil Aviation Operators*, Int. Air Transp. Assoc., Montreal, QC, Canada, 2011.

[4] E. Q. Wu, X. Y. Peng, C. Z. Zhang, J. X. Lin, and S. F. Sheng, "Pilots' fatigue status recognition using deep contractive autoencoder network," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 10, pp. 3907–3919, Oct. 2019.

[5] E. Q. Wu *et al.*, "Detecting fatigue status of pilots based on deep learning network using EEG signals," *IEEE Trans. Cogn. Develop. Syst.*, early access, Jan. 1, 2020, doi: 10.1109/TCDS.2019.2963476.

[6] D. Iacoviello, A. Petracca, M. Spezialetti, and G. Placidi, "A classification algorithm for electroencephalography signals by self-induced emotional stimuli," *IEEE Trans. Cybern.*, vol. 46, no. 12, pp. 3171–3180, Dec. 2016.

[7] L. Xie, Z. Deng, P. Xu, K. Choi, and S. Wang, "Generalized hidden-mapping transductive transfer learning for recognition of epileptic electroencephalogram signals," *IEEE Trans. Cybern.*, vol. 49, no. 6, pp. 2200–2214, Jun. 2019.

[8] Q. Zheng, Y. Wang, and P. A. Heng, "Multitask feature learning meets robust tensor decomposition for EEG classification," *IEEE Trans. Cybern.*, early access, Oct. 31, 2019, doi: 10.1109/TCYB.2019.2946914.

[9] A. H. Mooij, B. Frauscher, M. Amiri, W. M. Otte, and J. Gotman, "Differentiating epileptic from non-epileptic high frequency intracerebral EEG signals with measures of wavelet entropy," *Clin. Neurophysiol.*, vol. 127, no. 12, pp. 3529–3536, 2016.

[10] E. Debie *et al.*, "Multimodal fusion for objective assessment of cognitive workload: A review," *IEEE Trans. Cybern.*, early access, Sep. 23, 2019, doi: 10.1109/TCYB.2019.2939399.

[11] T. Zhang, W. Zheng, Z. Cui, Y. Zong, and Y. Li, "Spatial–temporal recurrent neural network for emotion recognition," *IEEE Trans. Cybern.*, vol. 49, no. 3, pp. 839–847, Mar. 2019.

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

[13] Y. Zhang, C. S. Nam, G. Zhou, J. Jin, X. Wang, and A. Cichocki, "Temporally constrained sparse group spatial patterns for motor imagery BCI," *IEEE Trans. Cybern.*, vol. 49, no. 9, pp. 3322–3332, Sep. 2019.

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

[15] A. S. Zandi, R. Tafreshi, M. Javidan, and G. A. Dumont, "Predicting epileptic seizures in scalp EEG based on a variational Bayesian gaussian mixture model of zero-crossing intervals," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 5, pp. 1401–1413, May 2013.

[16] L. He, D. Hu, M. Wan, Y. Wen, K. M. Deneen, and M. C. Zhou, "Common Bayesian network for classification of EEG-based multiclass motor imagery BCI," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 46, no. 6, pp. 843–854, Jun. 2016.

[17] S.-Y. Dong, B.-K. Kim, and S.-Y. Lee, "EEG-Based classification of implicit intention during self-relevant sentence reading," *IEEE Trans. Cybern.*, vol. 46, no. 11, pp. 2535–2542, Nov. 2016.

[18] N. S. Kwak and S. W. Lee, "Error correction regression framework for enhancing the decoding accuracies of ear-EEG brain–computer interfaces," *IEEE Trans. Cybern.*, vol. 50, no. 8, pp. 3654–3667, Aug. 2020.

[19] D. Zhang, L. Yao, K. Chen, S. Wang, X. Chang, and Y. Liu, "Making sense of spatio-temporal preserving representations for EEG-Based human intention recognition," *IEEE Trans. Cybern.*, vol. 50, no. 7, pp. 3033–3044, Jul. 2020.

[20] F. J. Solis, A. Maurer, J. Jiang, and A. Papandreou-Suppappola, "Adaptive EEG artifact suppression using Gaussian mixture modeling," in *Proc. 49th Asilomar Conf. Signals Syst. Comput.*, Pacific Grove, CA, USA, 2015, pp. 607–611.

[21] H. Zhang, Y. Liu, J. Liang, J. Cao, and L. Zhang, "Gaussian mixture modeling in stroke patients' rehabilitation EEG data analysis," in *Proc. IEEE 35th Annu. Int. Conf. Eng. Med. Biol. Soc. (EMBC)*, Osaka, Japan, 2013, pp. 2208–2211.

[22] B. Chakraborty, L. Ghosh, and A. Konar, "Optimal selection of EEG electrodes using interval type-2 fuzzy-logic-based semiseparating signaling game," *IEEE Trans. Cybern.*, early access, Feb. 20, 2020, doi: 10.1109/TCYB.2020.2968625.

[23] A. Belyavin and N. A. Wright, "Changes in electrical activity of the brain with vigilance," *Electroencephalogr. Clin. Neurophysiol.*, vol. 66, no. 2, pp. 137–144, 1987.

[24] A. Subasi, "Automatic recognition of alertness level from EEG by using neural network and wavelet coefficients," *Expert Syst. Appl.*, vol. 28, no. 4, pp. 701–711, 2005.

[25] F. Qi *et al.*, "Spatiotemporal-filtering-based channel selection for single-trial EEG classification," *IEEE Trans. Cybern.*, early access, Jan. 21, 2020, doi: 10.1109/TCYB.2019.2963709.

[26] B. T. Jap, S. Lal, and P. Fischer, "Using EEG spectral components to assess algorithms for detecting fatigue," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 2352–2359, 2009.

[27] H. J. Eoh, M. K. Chung, and S. H. Kim, "Electroencephalographic study of drowsiness in simulated driving with sleep deprivation," *Int. J. Ind. Ergon.*, vol. 35, no. 4, pp. 307–320, 2005.

[28] W. Li, Q.-C. He, X.-M. Fan, and Z.-M. Fei, "Evaluation of driver fatigue on two channels of EEG data," *Neurosci. Lett.*, vol. 506, no. 2, pp. 235–239, 2012.

[29] A. B. Lee, B. Nadler, and L. Wasserman, "Treelets—An adaptive multi-scale basis for sparse unordered data," *Ann. Appl. Stat.*, vol. 2, no. 2, pp. 435–471, 2008.

[30] E. Q. Wu *et al.*, "Novel nonlinear approach for real-time fatigue EEG data: An infinitely warped model of weighted permutation entropy," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 6, pp. 2437–2448, Jun. 2020.

[31] J. Zhang, Z. Yin, and R. Wang, "Pattern classification of instantaneous cognitive task-load through GMM clustering, laplacian eigenmap, and ensemble SVMs," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 14, no. 4, pp. 947–965, Jul./Aug. 2017.

[32] R. P. Balandong, R. F. Ahmad, M. N. M. Saad, and A. S. Malik, "A review on EEG-based automatic sleepiness detection systems for driver," *IEEE Access*, vol. 6, pp. 22908–22919, 2018.

[33] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," *IEEE Trans. Auton. Mental Develop.*, vol. 7, no. 3, pp. 162–175, Sep. 2015.

[34] X. Wei and C. Li, "The student's *t*-hidden Markov model with truncated stick-breaking priors," *IEEE Signal Process. Lett.*, vol. 18, no. 6, pp. 355–358, Jun. 2011.

[35] R. M. Mehmood, R. Du, and H. J. Lee, "Optimal feature selection and deep learning ensembles method for emotion recognition from human–brain EEG sensors," *IEEE Access*, vol. 5, pp. 14797–14806, 2017.

[36] T. M. Nguyen and Q. M. Wu, "A nonsymmetric mixture model for unsupervised image segmentation," *IEEE Trans. Cybern.*, vol. 43, no. 2, pp. 751–765, Apr. 2013.

[37] H. Sundar, C. S. Seelamantula, and T. V. Sreenivas, "A mixture model approach for formant tracking and the robustness of student's-*t* distribution," *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, no. 10, pp. 2626–2636, Dec. 2012.

[38] T. M. Nguyen and Q. M. Wu, "Bounded asymmetrical student's-*t* mixture model," *IEEE Trans. Cybern.*, vol. 44, no. 6, pp. 857–869, Jun. 2014.
[39] H. Zhang, Q. M. Wu, T. M. Nguyen, and X. Sun, "Synthetic aperture radar image segmentation by modified student's *t*-mixture model," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 4391–4403, Jul. 2014.
[40] C. Archambeau and M. Verleysen, "Robust Bayesian clustering," *Neural Netw.*, vol. 20, no. 1, pp. 129–138, 2007.
[41] M. Nguyen and Q. M. Wu, "Robust student's-*t* mixture model with spatial constraints and its application in medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 31, no. 1, pp. 103–116, Jan. 2012.
[42] T. M. Nguyen, Q. M. Jonathan Wu, and H. Zhang, "Asymmetric mixture model with simultaneous feature selection and model detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 2, pp. 400–408, Feb. 2015.
[43] J. Gilles, "Empirical wavelet transform," *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 3999–4010, Aug. 2013.
[44] H. E. Rauch, F. Tung, and C. T. Striebel, "Maximum likelihood estimates of linear dynamic systems," *J. AIAA*, vol. 3, no. 8, pp. 1445–1450, 1965.
[45] G. J. McLachlan and D. Peel, *Finite Mixture Models*. New York, NY, USA: Wiley, 2000.
[46] M. Kumar, B. Packer, and D. Koller, "Self-paced learning for latent variable models," in *Advances in Neural Information Processing Systems*. La Jolla, CA, USA: Neural Inf. Process. Syst., 2010.
[47] Z. Ren, D. Dong, H. Li, and C. Chen, "Self-paced prioritized curriculum learning with coverage penalty in deep reinforcement learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 6, pp. 2216–2226, Jun. 2018.
[48] C. Li, F. Wei, J. Yan, X. Zhang, Q. Liu, and H. Zha, "A self-paced regularization framework for multilabel learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 6, pp. 2660–2666, Jun. 2018.
[49] L. Lin, K. Wang, D. Meng, W. Zuo, and L. Zhang, "Active self-paced learning for cost-effective and progressive face identification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 1, pp. 7–19, Jan. 2018.
[50] X. Wei and C. Li, "The infinite student's *t*-mixture for robust modeling," *Signal Process.*, vol. 92, no. 1, pp. 224–234, 2012.
[51] A. Calabrese and L. Paninski, "Kalman filter mixture model for spike sorting of non-stationary data," *J. Neurosci. Methods*, vol. 196, pp. 159–169, Mar. 2011.
[52] B. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1979.
[53] P. Y. Zhang, S. Shu, and M. C. Zhou, "An online fault detection model and strategies based on SVM-grid in clouds," *IEEE/CAA J. Autom. Sinica*, vol. 5, no. 2, pp. 445–456, Mar. 2018.
[54] D. P. Bertsekas, "Feature-based aggregation and deep reinforcement learning: A survey and some new implementations," *IEEE/CAA J. Autom. Sinica*, vol. 6, no. 1, pp. 1–31, Jan. 2019.
[55] H. Fang, N. Tian, Y. Wang, M. Zhou, and M. A. Haile, "Nonlinear Bayesian estimation: From Kalman filtering to a broader horizon," *IEEE/CAA J. Autom. Sinica*, vol. 5, no. 2, pp. 401–417, Mar. 2018.
[56] Q. Wu and X. Chu, "Recognition of fatigue status of pilots based on deep contractive sparse auto-encoding network," in *Proc. 37th Chin. Control Conf.*, Wuhan, China, 2018, pp. 9220–9225.
[57] S. Gao, M. Zhou, Y. Wang, J. Cheng, H. Yachi, and J. Wang, "Dendritic neuron model with effective learning algorithms for classification, approximation and prediction," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 2, pp. 601–614, Feb. 2019.
[58] Q. Kang, S. Yao, M. Zhou, K. Zhang, and A. Abusorrah, "Enhanced subspace distribution matching for fast visual domain adaptation," *IEEE Trans. Comput. Social Syst.*, vol. 7, no. 4, pp. 1047–1057, Aug. 2020.

Edmond Q. Wu (Member, IEEE) received the Ph.D. degree in control theory and application from Southeast University, Nanjing, China, in 2009.

He is an Associate Professor with the Key Laboratory of System Control and Information Processing, Ministry of Education, Shanghai Jiao Tong University, Shanghai, China. He is also with the Science and Technology on Avionics Integration Laboratory, China National Aeronautical Radio Electronics Research Institute, Shanghai. His research interests include deep learning, cognitive modeling, and industrial information processing.

Mengchu Zhou (Fellow, IEEE) received the B.S. degree in control engineering from the Nanjing University of Science and Technology, Nanjing, China, in 1983, the M.S. degree in automatic control from the Beijing Institute of Technology, Beijing, China, in 1986, and the Ph.D. degree in computer and systems engineering from Rensselaer Polytechnic Institute, Troy, NY, USA, in 1990.

He joined the New Jersey Institute of Technology, Newark, NJ, USA, in 1990, where he is currently a Distinguished Professor of Electrical and Computer Engineering. He has led or participated in over 50 research and education projects with total budget over $12M, funded by National Science Foundation, Department of Defense, NIST, New Jersey Science and Technology Commission, and industry. He has over 900 publications including 12 books, over 600 journal papers (over 450 in IEEE transactions), and 29 book-chapters. He holds 27 patents and several pending ones. His research interests are in intelligent automation, Petri nets, Internet of Things, big data, cloud/edge computing, transportation, and energy systems.

Dr. Zhou was a recipient of the Excellence in Research Prize and Medal from NJIT, Humboldt Research Award for U.S. Senior Scientists from Alexander von Humboldt Foundation, and the Franklin V. Taylor Memorial Award and the Norbert Wiener Award from IEEE SMC Society, and the Edison Patent Award from the Research and Development Council of New Jersey, and Computer-Integrated Manufacturing UNIVERSITY-LEAD Award from Society of Manufacturing Engineers. He is the Founding Editor of IEEE Press Book Series on Systems Science and Engineering and an Editor-in-Chief of IEEE/CAA JOURNAL OF AUTOMATICA SINICA. He served as an Associate Editor of IEEE TRANSACTIONS ON ROBOTICS AND AUTOMATION, IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING, and IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, and an Editor of IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING. He served as a Guest-Editor for many journals, including IEEE INTERNET OF THINGS JOURNAL, IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, and IEEE TRANSACTIONS ON SEMICONDUCTOR MANUFACTURING. He is currently an Associate Editor of IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, IEEE INTERNET OF THINGS JOURNAL, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS, and *Frontiers of Information Technology and Electronic Engineering*. He was the General Chair of IEEE Conference on Automation Science and Engineering in August 2008, the General Co-Chair of IEEE International Conference on System, Man and Cybernetics (SMC) in October 2003, and IEEE International Conference on SMC in October 2019, a Founding General Co-Chair of IEEE International Conference on Networking, Sensing and Control in March 2004, and the General Chair of IEEE International Conference on Networking, Sensing and Control in April 2006. He was a Program Chair of IEEE International Conference on Mechatronics and Automation in August 2010, and 1998 and 2001 IEEE International Conference on SMC and 1997 IEEE International Conference on Emerging Technologies and Factory Automation. He has been among most highly cited scholars since 2012 and ranked top one in the field of engineering worldwide in 2012 by Web of Science. He is a Life Member of Chinese Association for Science and Technology–USA and served as its President in 1999. He is a Fellow of the International Federation of Automatic Control, American Association for the Advancement of Science, and the Chinese Association of Automation.

Dewen Hu (Senior Member, IEEE) was born in Hunan, China, in 1963. He received the B.Sc. and M.Sc. degrees from Xi'an Jiaotong University, Xi'an, China, in 1983 and 1986, respectively, and the Ph.D. degree from the National University of Defense Technology (NUDT), Changsha, China, in 1999.

Since 1986, he has been with NUDT. From 1995 to 1996, he was a visiting scholar with the University of Sheffield, Sheffield, U.K. He was promoted as a Professor in 1996. His research interests include image processing, system identification and control, neural networks, and cognitive science.

Dr. Hu is an Associate Editor of IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS.

Longjun Zhu received the master's degree in control theory and control engineering from the East China University of Science and Technology, Shanghai, China, in 2009.

She is currently a Lecturer with the School of Artificial Intelligence, Shanghai Normal University Tianhua College, Shanghai. Her research interest includes intelligent computing.

Zhiri Tang received the bachelor's and master's degree in microelectronics from Wuhan University, Wuhan, China, in 2017 and 2019, respectively.

His research interests include machine learning and cognitive computing.

Xu-Yi Qiu is currently pursuing the Ph.D. degree with Shanghai Jiao Tong University, Shanghai, China.

He is currently with the Science and Technology on Avionics Integration Laboratory, China National Aeronautical Radio Electronics Research Institute, Shanghai. His research interests include human–machine interaction and cognitive modeling.

Ping-Yu Deng is currently pursuing the Ph.D. degree with the Institute of Aerospace Science and Technology, Shanghai Jiao Tong University, Shanghai, China. He is a Visiting Professor with the Science and Technology on Avionics Integration Laboratory, China National Aeronautical Radio Electronics Research Institute, Shanghai, China. His research interests include avionics system design, analysis, and architecture.

Li-Min Zhu (Member, IEEE) received the B.E. (Hons.) and Ph.D. degrees in mechanical engineering from Southeast University, Nanjing, China, in 1994 and 1999, respectively.

He is currently the "Cheung Kong" Chair Professor, the Head of the Department of Mechanical Engineering, and the Vice Director of the State Key Laboratory of Mechanical System and Vibration, School of Mechanical Engineering, Shanghai Jiao Tong University. His research interests include mechanical signature analysis for machine and manufacturing process monitoring, CNC machining and 3-D measurement of complex shaped parts, and control, sensing, and instrumentation for micro/nano manufacturing.

Dr. Zhu is currently an Associate Editor of the IEEE/ASME TRANSACTIONS ON MECHATRONICS, and the Editorial Board Members of the *Proceedings of the Institution of Mechanical Engineer, Part B: Journal of Engineering Manufacture*, and the *International Journal of Intelligent Robotics and Applications*.

He Ren received the Ph.D. degree in aerospace engineering from Northwestern Polytechnical University, XiŠan, China, in 1987 and also the Ph.D. degree from National 1000 Talent Program, Shanghai Engineering Research Center of Civil Aircraft Health Monitoring, COMAC, Shanghai, China.

He is currently a Professor with National 1000 Talent Program, Shanghai Engineering Research Center of Civil Aircraft Health Monitoring, COMAC. His research interests are maintenance engineering analysis, diagnosis, prognosis, and health management.