

# Nonparametric Bayesian Prior Inducing Deep Network for Automatic Detection of Cognitive Status

Edmond Q. Wu<sup>ID</sup>, Dewen Hu<sup>ID</sup>, Ping-Yu Deng, Zhiri Tang, Yulian Cao, *Member, IEEE*, Wen-Ming Zhang<sup>ID</sup>, *Member, IEEE*, Li-Min Zhu<sup>ID</sup>, *Member, IEEE*, and He Ren

**Abstract**—Pilots’ brain fatigue status recognition faces two important issues. They are how to extract brain cognitive features and how to identify these fatigue characteristics. In this article, a gamma deep belief network is proposed to extract multilayer deep representations of high-dimensional cognitive data. The Dirichlet distributed connection weight vector is upsampled layer by layer in each iteration, and then the hidden units of the gamma distribution are downsampled. An effective upper and lower Gibbs sampler is formed to realize the automatic reasoning of the network structure. In order to extract the 3-D instantaneous time-frequency distribution spectrum of electroencephalogram (EEG) signals and avoid signal modal aliasing, this article also proposes a smoothed pseudo affine Wigner–Ville distribution method. Finally, experimental results show that our model achieves satisfactory results in terms of both recognition accuracy and stability.

Manuscript received October 11, 2019; revised January 5, 2020; accepted February 25, 2020. This work was supported in part by the National Natural Science Foundation for Distinguished Young Scholars of China under Grant 11625208, in part by National Key Research and Development Program of China 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, in part by the Suzhou Industrial Technology Innovation Special Project under Grant SS201704, and in part by the Jiangxi Province Key Research and Development Program under Grant 20192BBE50065. This article was recommended by Associate Editor P. Chen. (*Corresponding authors:* Edmond Q. Wu; Ping-Yu Deng; Zhiri Tang)

Edmond Q. Wu is with the Key Laboratory of System Control and Information Processing, Ministry of Education, 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).

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).

Ping-Yu Deng is with the Science and Technology on Avionics Integration Laboratory, China National Aeronautical Radio Electronics Research Institute, Shanghai, China (e-mail: deng\_pinguo@careri.com).

Zhiri Tang is with the Department of Computer Science, City University of Hong Kong, Hong Kong (e-mail: gerin.tang@my.cityu.edu.hk).

Yulian Cao is with the Department of Automation, University of New South Wales, Sydney, NSW 2052, Australia (e-mail: yulian.cao@unsw.edu.au).

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

He Ren is with the Department of Automation, COMAC Shanghai Aircraft Design and Research Institute, Shanghai 200232, China.

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

Digital Object Identifier 10.1109/TCYB.2020.2977267

**Index Terms**—Electroencephalogram (EEG) signals, gamma deep belief network (DBN), pilots’ fatigue, Wigner–Ville distribution (WVD).

## I. INTRODUCTION

### A. Motivation

Maintaining a high concentration for cognitive behavior for a long time will lead to brain fatigue, which is usually accompanied by a decline in decision-making ability, attention, and response speed [1]. During the flight, factors such as lack of sleep, long flight time, and bad weather may cause pilots’ brain fatigue, which may lead to operational errors or misjudgment [2], which is a potential danger during flight. Therefore, it is of great significance to give early warning to pilots’ fatigue in an objective and accurate way.

Obviously, identifying the cognitive state of the pilot’s brain is important for flight safety. It provides technical support for the pilot workload assessment. We also know that existing models have many limitations in assessing the cognitive state of the brain. For example, the parameters of the model need to be manually given in advance, and the network level needs to be set in advance. This is obviously not suitable for the data-driven automatic modeling mechanism. Obviously, multidimensional data should have a better mapping network which can better approximate high-dimensional data. However, the priors of the parameters and the network level of the multilayer mapping network are unknown. Obviously, it is very important to study this kind of network structure and parameters of automatic reasoning technology. In this article, an automatic reasoning scheme for the cognitive state of the pilots’ brain can prove the rationality of our proposed deep multilayer network based on nonparametric prior drive.

### B. Related work

Brain states can be divided into two categories: 1) awake and 2) fatigue. When a person feels tired, his fatigue status will gradually deepen over time. The assessment of fatigue level mainly includes subjective evaluation and objective evaluation. Subjective evaluation methods infer workload or brain fatigue status through subject experience, including NASA’s task load index (TLX) [3], Karolinska sleepiness scale (KSS) [4], and so on. Objective evaluation methods can assess the state

of fatigue caused by personal behavior [5] with long-term closed eyes, yawning or biological signals, such as blood pressure [6] and electrocardiogram [7].

Subjective assessment methods rely on individual subjective feedback and may distract the driver. Validating the questionnaire takes time and validation. In the objective evaluation methods, there may be privacy issues when recording driver's actions or facial expressions. The brain-computer interface (BCI), as a passive assessment method, has been shown to be effective in assessing the brain cognitive status [8], [9]. Kwak and Lee [31] proposed the ear-electroencephalogram (EEG)-based advanced decoding accuracy and SSVEP BCI solution. They designed a holistic approach to improve the ear-EEG decoding accuracy of the SSVEP paradigm. It can be assumed that when participants perform motion imaging (MI) task, the channels associated with motion MI should contain similar behavior. This assumption can be used to select channels with more relevant information [32]. A novel correlation-based time window selection (CTWS) algorithm was also proposed to select MI-based BCI [33]. In their method, the level of the reference signal is selected based on correlation analysis and performance evaluation. The experimental results greatly improve classification accuracy.

Brain dynamics-based assessment methods include computer tomography (CT), electroencephalography (EEG), and magnetic resonance imaging (MRI). Compared with other data extraction methods, the significant advantage of EEG is that it can be measured in real time, and certain changes in brain activity have been shown to be related to fatigue status [13], [14]. The dynamics of the brain's central and occipital lobe were found [15]. The experimental results show that the correlation between attention and fatigue can be determined based on the response time in simulation experiment. Lal and Craig [16] discovered that the power of alpha rhythm is one of the important features of drowsiness. Therefore, EEG is considered as a reliable fatigue detection method [13] and has been widely used in the field of fatigue detection [10]–[12]. Then, in this article, we will evaluate pilot fatigue based on changes in the EEG signal.

The cognitive of brain fatigue can be divided into two categories: 1) empirical analysis and judgment based on quantitative indicators and 2) brain fatigue feature learning and recognition based on the deep-learning network.

The degree of fatigue inferred from the changes in EEG signals has confirmed empirical analysis methods for quantitative indicators. For example, Papadelis *et al.* [17] studied the relative rhythm ratio (RRR) of Shannon entropy and Kullback-Leibler (K-L) entropy of EEG signal, and it was found that the RRR of  $\alpha$  rhythm before driving significantly increased, while K-L entropy decreased significantly. He *et al.* [38] proposed a common Bayesian network (CBN) method to implement EEG classification of multiple types of MI. They use a Gaussian mixture model to model the relevant channels to build a Bayesian network. They experimented on two well-known BCI datasets to obtain better classification results. Jap *et al.* [18] established four fatigue detection indicators  $(\theta + \alpha)/\beta$ ,  $\alpha/\beta$ ,  $(\theta + \alpha)/(\alpha + \beta)$ , and  $\theta/\beta$ . Brain fatigue is

assessed by its tendency. The results show that  $\delta$  and  $\theta$  waves are more stable over time, the  $\alpha$  wave is slightly attenuated, and the  $\beta$  is significantly attenuated. For a description of the four main rhythms and their mental states, refer to Table I.

Empirical analysis and judgment methods rely on existing research results. However, EEG signals are complex and volatile and vary widely from individual to individual. It is difficult to find the internal dynamics of brain cognition through human judgment. Deep-learning networks can better solve the above problems and have been widely used for feature learning and recognition of EEG signals [35]–[37]. Cai *et al.* [19] analyzed the EEG signal through the deep belief network (DBN) for depression detection and compared it with the  $K$ -nearest neighbor (KNN) and the artificial neural network (ANN). The results show that the recognition accuracy of DBN can reach 78.24%. Chen *et al.* [27] proposed an emotional learning model based on deep convolution neural network (DCNN) which effectively improved the accuracy and stability of EEG recognition. Chai *et al.* [29] found that Bayesian neural networks (BNNs) have better classification accuracy for mental fatigue testing. Two deep-learning-based frameworks were proposed by Zhang *et al.* [34], which have novel spatiotemporal representations of the original EEG stream to accurately identify human intention. In their program, both convolutional and recurrent neural networks can effectively explore the preserved spatial-temporal information in a cascade or parallel manner.

EEG signals have been used for emotion recognition, car driver fatigue recognition, and disease diagnosis. Little research has been done in the field of pilot brain fatigue recognition. The pilot brain fatigue recognition is an integral part of ensuring flight safety. Then, this article proposed a learning algorithm that combines quantitative indicators and deep-learning networks to provide a reliable solution for pilot fatigue detection. Obviously, changes in EEG rhythm waves are highly correlated with brain activity. We can characterize changes in brain cognition through the rhythmic features. First, the spectrum of four rhythms is calculated by the proposed smoothed pseudo affine Wigner-Ville distribution (SPA WVD) instead of the fast Fourier transform (FFT). Since the delta wave shows a deep sleep state, the usual fatigue index does not take its effect into account. From the perspective of physiological mechanisms, the appearance of delta waves indicates that people are drowsy. Therefore, in this article, an indicator was added, which is the ratio of the two high-frequency rhythm waves to the two low-frequency rhythm spectrum areas. Then, five typical fatigue indicators are defined:  $\theta/\beta$ ,  $\alpha/\beta$ ,  $(\alpha + \theta)/\beta$ ,  $(\theta + \delta)/(\alpha + \beta)$ , and  $(\alpha + \theta)/(\alpha + \beta)$ , where each rhythm symbol represents its spectral area.

The sigmoid belief network (SBN) connects the binary elements of adjacent layers through the sigmoid function to infer a deeper representation of multivariate 2-D vectors [20]. The DBN is a deep network obtained by replacing the top hidden layer of the SBN with an undirected restricted Boltzmann machine (RBM). It uses RBM to connect binary elements of adjacent layers. This type of learning mechanism simulates binary observations [21]. The advantage is that the

deep probabilistic models can be effectively learned through layer-by-layer training, and classification performance can be optimized by discriminating and fine-tuning algorithms [28], which can present better results under various classifications [19], [30]. Although DBNs can model a large data to achieve data classification. However, one of the difficulties that everyone faces is how to determine the structure of the network. Some scholars have tried to solve this problem. For example, Wang proposed a growing DBN with transfer learning. It automatically determines the network structure and improves the accuracy of the model [35]. Obviously, this article has achieved a good classification effect. However, it has not fundamentally solved the problem of automatically deriving a network structure of the generated model. Fang *et al.* [42] proposed a Bayesian thinking for nonlinear state estimation which shows the nature of state estimation from the perceptive of Bayesian update. Based on the notion of Bayesian state estimation, they present a general form for parameter inference. Based on these backgrounds, we try to solve the deep-learning network structure optimization problem with nonparametric methods.

### C. Our Model

Based on DBN, this article uses the gamma function to infer hidden units of DBN. The DBN needs to use binary units for inference. It needs to adjust the width of each layer (the number of hidden units) and the depth of network (layers). This type of current network can connect different layers through binary units, while it is difficult to represent the probability of neurons appearing. To improve this dilemma, we develop a new network of probabilistic neuron activation mechanisms. We use the Poisson function to control the occurrence of neurons, and use the gamma function to derive the structure of the DBN network. We call this network gamma DBN (GDBN).

In GDBN, we can use the regularized Poisson factorization method to determine the width of the first layer network and infer the connection hidden units and weights of the next layer by the shape parameters of the gamma distribution. The network depth is determined by *a priori* reasoning of the shape parameters based on the gamma distribution. In this way, the width and depth of the network can be obtained by nonparametric methods. Our model uses the Gibbs sampling to derive network parameters. Specifically, the upsampling obtains the prior connection weight that follows the Dirichlet distribution and the downsampling obtains the hidden units that follow the gamma distribution. By jointly training all hidden layers, when the first layer width is obtained via the Poisson factor analysis with  $L_1$  regularization, the width of each layer can be inferred in a greedy manner, our GDBN has greatly changed the network structure and learning mode of DBN, and promoted the main self-learning ability of DBN. The contributions of this article can be summarized as follows: 1) we propose a nonparametric learning method to discover potential network structures and 2) we develop an up-down Gibbs sampling method to infer the joint distribution of network structure parameters. Through GDBN and SPAWVD, the scheme of pilot brain cognition given in this work is shown in Fig. 1.

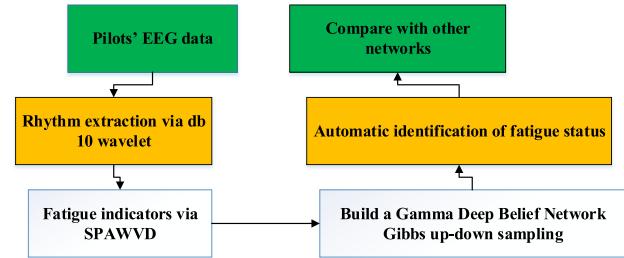


Fig. 1. Flowchart for inferring brain fatigue.

TABLE I  
RHYTHM WAVES AND THEIR COGNITIVE BEHAVIOR

Rhythm	Frequency	Characteristics
$\delta$	0~4Hz	Also known as sleep waves, common in sleep status
$\theta$	4~8Hz	Related to the inhibition status of the central nervous system, common in fatigue;
$\alpha$	8~13Hz	Common in the closed eye status, disappeared after stimulation;
$\beta$	13~30Hz	Often appear as an additional rhythm of $\alpha$ , under the stimulated status, $\alpha$ rhythm will be converted to $\beta$ rhythm.

The remainder of this article is structured as follows. Section II proposed five fatigue indicators based on SPAWVD, and Section III provides the establishment of GDBN. Experiments and simulations are arranged in Section IV. Section V gives the conclusion.

## II. FATIGUE INDICATORS

### A. Rhythm Waves

Rhythm waves can directly reflect changes in brain activity. Table I shows the cognitive behaviors corresponding to the four rhythm waves.

### B. Feature Learning

In order to transform rhythm wave into more important features, an improved Wigner–Ville distribution (WVD) [40] based on the Kaiser window is used to calculate the instantaneous spectral information of rhythm wave. By adding a sliding exponential window to suppress cross-term interference in the WVD distribution, the smoothed pseudo WVD (SPWVD) [41] is affinely smoothed. This new distribution is called SPAWVD.

For a given signal  $s(t)$ , the time–frequency distribution can be expressed in the general form

$$w(t, \omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \emptyset(\theta, \tau) \\ \times s\left(u + \frac{\tau}{2}\right) s^*\left(u - \frac{\tau}{2}\right) e^{-j\theta t - j\tau\omega - j\theta u} du d\tau d\theta \quad (1)$$

where  $\emptyset(\theta, \tau)$  is a kernel function and  $s^*(u)$  represents the conjugate of  $s(u)$ .

When  $\emptyset(\theta, \tau) = 1$ ,  $w(t, \omega)$  is the WVD

$$w(t, \omega) = \int_{-\infty}^{+\infty} s^*(t - \frac{\tau}{2})s(t + \frac{\tau}{2})d\tau. \quad (2)$$

The WVD can be considered as the energy density distribution of the signal in the time-frequency plane. For continuous signal  $s(t)$ , WVD can be expressed as

$$w(t, \omega) = \int s(t + \frac{\tau}{2})s^*(t - \frac{\tau}{2})e^{-j\omega\tau}d\tau. \quad (3)$$

Accordingly, for the discrete signal  $s(t)$ , WVD can be expressed as

$$w(t, \omega) = 2 \sum_{\tau=-\infty}^{\tau=\infty} s(t + \tau)s^*(t - \tau)e^{-j\omega\tau}. \quad (4)$$

Let  $n = 0, \dots, N - 1$ . The discrete form of the sampled signal  $s(n)$  is

$$w(l, k) = \frac{1}{N} \sum_{n=0}^{N-1} s(l+n)s^*(l-n)e^{-\frac{4j\pi nk}{N}} \quad (5)$$

where  $l$  is the time of discrete WVD, and  $k$  is its frequency.

To avoid aliasing of the sampled signal, it is recommended to use an analytical signal. The analytic signal consists of a real part and an imaginary part, and the imaginary part can be obtained by a Hilbert transform, and the analytical signal can be expressed as

$$s(t) = s_r(t) + jH(s_r(t)) \quad (6)$$

where

$$H(s_r(t)) = s_r(t) * h(t) \quad (7)$$

$$h(t) = \begin{cases} \frac{2\sin^2(\frac{\pi t}{2})}{\pi t} & t \neq 0 \\ 0 & t = 0 \end{cases} \quad (8)$$

where the symbol  $*$  represents convolution and  $H(\cdot)$  represents the Hilbert transform.

When  $s(t)$  is a discrete signal, (7) can be rewritten as

$$H(s_r(n)) = \sum_{m=-\infty}^{\infty} s_r(m)h(n-m). \quad (9)$$

To suppress cross-term interference, the Kaiser window function  $G(t, w)$  is proposed

$$G(t, w) = \frac{I_0 \left[ a\sqrt{1 - (t/\tau)^2} + a\sqrt{1 - (w/\tau)^2} \right]}{I_0(a)} \quad (10)$$

where  $a$  is the non-negative parameter that adjusts the shape of the window,  $\tau$  is the residence time of the Kaiser window, and  $I_0$  is the first type of zero-order modified Bessel function. The WVD with the Kaiser window function can be expressed as

$$w(t, w) = \frac{1}{2\pi} \iint w(t', \omega') G(t - t', \omega - \omega') dt' d\omega' \quad (11)$$

where  $G(t, w)$  is a Kaiser window function.

The application range of the window function is set to be  $\pm 2\sigma_t$  and  $\pm 2\sigma_\omega$ , and  $w$  and  $t$  are multiples of the time and frequency steps, respectively.  $\Delta t$  and  $\Delta\omega$  of time and

frequency resolution of Kaiser window function are related to  $\sigma_t = j\Delta t$  and  $\sigma_\omega = k\Delta\omega$ , where  $j$  and  $k$  are time-domain and frequency-domain constants, respectively. The sampled window function can be expressed as

$$G[p, q] = \frac{I_0 \left[ a\sqrt{1 - \left( \frac{2p}{N-1} - 1 \right)^2} + a\sqrt{1 - \left( \frac{2q}{N-1} - 1 \right)^2} \right]}{I_0(a)} \quad (12)$$

where  $p$  and  $q$  are integers between  $\pm 2j$  and  $\pm 2k$ .

By combining (11) and (12), we can obtain the sampling window function of WVD as follows:

$$W'(l, m) = \frac{\Delta t \Delta \omega}{2\pi} \sum_{p=l-j}^{l+j} \sum_{q=m-k}^{m+k} w(p, q)G(p-l, q-m). \quad (13)$$

By adding the smoothing function  $\Psi(t, m)$ , we obtain SPAWVD as follows:

$$\Omega(t, \beta) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Psi\left(\frac{s-t}{\beta}, \beta m\right) W'(l, m) dl dm. \quad (14)$$

Then, the instantaneous spectral energy of the four rhythm waves is calculated by SPAWVD and combined into five fatigue indicators

$$\begin{cases} p_\alpha(\omega)/p_\beta(\omega) \\ p_\theta(\omega)/p_\beta(\omega) \\ (p_\alpha(\omega) + p_\theta(\omega))/p_\beta(\omega) \\ (p_\theta(\omega) + p_\delta(\omega))/(p_\alpha(\omega) + p_\beta(\omega)) \\ (p_\alpha(\omega) + p_\theta(\omega))/(p_\alpha(\omega) + p_\beta(\omega)) \end{cases} \quad (15)$$

where  $p_X(\omega)$  is the energy spectral density of the  $X$ -wave SPAWVD.

### III. GAMMA DEEP BELIEF NETWORK

In the current deep networks, binary hidden units are difficult to express the probability of neurons appearing. We use the Poisson factorization method to process the input to obtain the first layer of valid neurons. In fact, each neuron is a parameter of the Poisson distribution and these parameters obey the gamma distribution. The depth of the network structure is obtained *a priori* from the shape parameters of the gamma distribution. The width of the network structure is controlled by the Poisson factorization and the  $L_1$  regularization. A deep network usually uses binary units for simple processing inference, while deep networks need to adjust the width (the number of hidden units) and the depth (the number of layers) of the network, while our model uses non-negative real hidden units and automatically infers the width. The width of the first layer is determined by sparse  $L_1$  regularization, which forms *a priori* for the width inference of the subsequent layer.

The method of constructing a deep network structure reasoning in Fig. 2 is as follows: we use Poisson factorization to learn the latent variables of multidimensional input non-negative matrix and obtain the number of neurons in each layer network, that is, the width of the network. In the second layer network, the first layer obeys the gamma distribution hidden

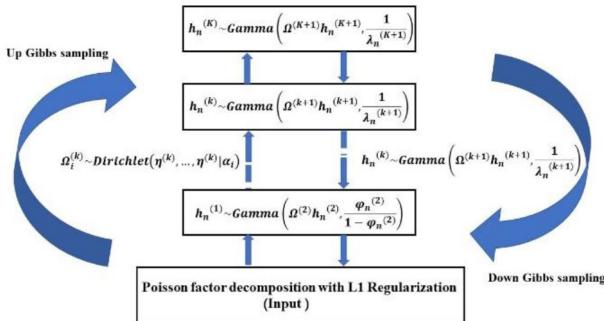


Fig. 2. Model construction mechanism.

unit matrix as the input of the second layer of Poisson factorization and obtains the width of the second layer network and the neurons and weights of the next layer network. According to the same principle, the shape parameters of the gamma hidden units of each layer are decomposed into the product of the weights and the gamma hidden units of the next layer. In this way, the structure and width of the network are inferred.

#### A. Model Definition

GDBN defines a learning mechanism for nonparametric prioritized network structures. Its bottom layer to top layer are observation layers, which are the first layer (hidden layer), ..., the  $K - 1$  layer (hidden layer), and the  $K$ th layer top layer. The bottom layer is the observation layer, which consists of the observation unit  $o_n^{(1)}$ . It can be expressed as the product of the weights and the hidden units of the next layer by a Poisson factor analysis [25]

$$o_n^{(1)} \sim \text{Poisson}(\Omega^{(1)} h_n^{(1)}) \quad (16)$$

where  $\Omega^{(1)}$  is the connection weight,  $h_n^{(1)}$  is the first hidden layer, and  $x \sim \text{Poisson}(\lambda)$  represents that  $x$  obeys the Poisson distribution with the parameter  $\lambda$ , and the probability function is

$$P(x = k) = \frac{\lambda^k}{k!} e^{-\lambda}. \quad (17)$$

From the observation layer to the top layer, the first, ...,  $k, \dots, K$  layer hidden layer can be expressed as

$$\begin{cases} h_n^{(1)} \sim \text{Gamma}(\Omega^{(2)} h_n^{(2)}, \frac{\varphi_n^{(2)}}{1-\varphi_n^{(2)}}) \\ \vdots \\ h_n^{(k)} \sim \text{Gamma}(\Omega^{(k+1)} h_n^{(k+1)}, \frac{1}{\lambda_n^{(k+1)}}) \\ \vdots \\ h_n^{(K)} \sim \text{Gamma}(\Omega^{(K+1)} h_n^{(K+1)}, \frac{1}{\lambda_n^{(K+1)}}) \end{cases} \quad (18)$$

where  $\Omega^{(k)}$  is the connection weight,  $h_n^{(k)}$  is the hidden layer of the  $k$ th layer, and  $\varphi_n^{(2)}$  is the probability parameter, which satisfies

$$\frac{\varphi_n^{(2)}}{1-\varphi_n^{(2)}} = \frac{1}{\lambda_n^{(2)}} \quad (19)$$

$x \sim \text{Gamma}(\alpha, [1/\lambda])$  indicates that  $x$  obeys a gamma distribution with shape parameter  $\alpha$  and scale parameter  $(1/\lambda)$ . Its

expectation and variance are  $(\alpha/\lambda)$  and  $(\alpha/\lambda^2)$ , respectively. The probability density function of input variable is

$$f(x; \alpha, \lambda) = \begin{cases} \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} & x > 0 \\ 0 & x = 0. \end{cases} \quad (20)$$

Similarly, the  $K$ th layer can be expressed as

$$\begin{cases} h_n^{(K)} \sim \text{Gamma}(\alpha, \frac{1}{\lambda_n^{(K+1)}}) \\ \alpha = (\alpha_1, \alpha_2, \dots, \alpha_{M_K})' \end{cases} \quad (21)$$

For the top layer  $h_n^{(K)}$ ,  $\alpha$  is the shared gamma distribution shape parameter and  $(1/\lambda_n^{(K+1)})$  is its scale parameter. We expect the correlations of input  $o_n^{(1)}$  to be captured by weight  $\Omega^{(1)}$ , and the neurons of  $h_n^{(k)}$  to be captured by the product of  $\Omega^{(k+1)}$  and  $h_n^{(k+1)}$ .

To limit the network complexity and facilitate the parameter inference, each column of  $\Omega^{(k)} \in \mathbb{R}^{M_{k-1} \times M_k}$  is regularized based on the  $L_1$  paradigm. When  $k \in 1, 2, \dots, K - 1$ , we have

$$\begin{cases} \Omega_i^{(k)} \sim \text{Dirichlet}(\eta^{(k)}, \dots, \eta^{(k)} | \alpha_i) \\ \alpha_i \sim \text{Gamma}(\frac{\gamma_0}{M_K}, \frac{1}{\lambda_0}) \end{cases} \quad (22)$$

where  $\text{Dirichlet}(\vec{p} | \vec{\alpha})$  denotes a Dirichlet distribution whose  $\vec{p}$  obeys the parameter  $\vec{\alpha} = (\alpha_1, \dots, \alpha_K)$ , and its probability density function is

$$\text{Diri}(\vec{p} | \vec{\alpha}) = \frac{1}{B(\vec{\alpha})} \prod_{k=1}^K p_k^{\alpha_k-1} \quad (23)$$

$B(\vec{\alpha})$  represents the normalized constant of the Dirichlet distribution

$$B(\vec{\alpha}) = \int \prod_{k=1}^K p_k^{\alpha_k-1} d\vec{p}. \quad (24)$$

In (22),  $\omega_i^{(k)} \in \mathbb{R}^{M_{k-1}}$  is the  $i$ th column of  $\Omega^{(k)}$ , and  $\lambda_0$  and  $\gamma_0$  obey the gamma distribution

$$\begin{cases} \lambda_0 \sim \text{Gamma}(a_0, \frac{1}{b_0}) \\ \gamma_0 \sim \text{Gamma}(c_0, \frac{1}{d_0}) \end{cases} \quad (25)$$

For  $k = 3, \dots, T + 1$ , we have

$$\begin{cases} \lambda_n^{(t)} \sim \text{Gamma}(a_0, \frac{1}{b_0}) \\ \varphi_n^{(2)} \sim \text{Beta}(c_0, d_0) \end{cases} \quad (26)$$

where  $x \sim \text{Beta}(\alpha, \beta)$  indicates that  $x$  obeys the beta distribution of  $\alpha$  and  $\beta$ , and its probability density function is

$$f(x; \alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}. \quad (27)$$

Then, the relationship between the hidden units  $h_1^{(k)}, h_2^{(k)}, \dots, h_{M_k}^{(k)}$  in line  $M_k$  can be represented by column vector  $\omega_i^{(k)}$  ( $i = 1, 2, \dots, M_K$ ) of  $\Omega^{(k+1)}$ .

Since the conjugate prior to the shape parameter of the gamma distribution in the network is unknown, it is difficult to calculate the condition posterior derivation network structure. Then, the calculation can be simplified based on the data enhancement algorithm in [31]. In a single-layer network,

the hidden unit of each layer is independent of *a priori*, and the deep network with  $K \geq 2$  can capture the correlation between hidden units. For  $k = 1, 2, \dots, K$ , we have

$$\begin{cases} \varphi_n^{(1)} = 1 + \frac{1}{e} \\ \varphi_n^{(k+1)} = \frac{-\ln(1-\varphi_n^{(k)})}{\lambda_n^{(k+1)} - \ln(1-\varphi_n^{(k)})}. \end{cases} \quad (28)$$

Then,  $u_n^{(t)}$  can be represented by a Poisson probability of the product of  $\Omega^{(k)}$  and  $h_n^{(k)}$  at the  $t$ th layer, when  $k = 1$  is the observation layer and  $k = 2, 3, \dots, K$  is the hidden layer

$$u_n^{(k)} \sim \text{Poisson}\left(-\Omega^{(k)} h_n^{(k)} \ln\left(1 - \varphi_n^{(k)}\right)\right). \quad (29)$$

The above equation holds for  $k = 1$ . When  $k = 2, 3, \dots, K$ , we have

$$\begin{cases} u_{mn}^{(k)} = \sum_{i=1}^{M_k} u_{mni}^{(k)} \\ u_{mni}^{(k)} \sim \text{Poisson}\left(-\omega_{mi}^{(k)} h_{in}^{(k)} \ln\left(1 - \varphi_n^{(k)}\right)\right) \end{cases} \quad (30)$$

where  $m \in \{1, \dots, M_{k-1}\}$ . Let  $q_{in}^{(k)(k+1)} := u_{ni}^{(k)} := \sum_{m=1}^{M_k} u_{mni}^{(k)}$  indicates the number of times that the factor  $i \in \{1, \dots, M_k\}$  in the  $t$ th layer appears in the observation unit  $n$ , and  $q_n^{(k)(k+1)} := (u_{n1}^{(k)}, \dots, u_{nM_k}^{(k)})'$ .

By marginalizing  $\Omega^{(k)}$ , we obtain

$$q_n^{(k)(k+1)} \sim \text{Poisson}\left(-h_n^{(k)} \ln\left(1 - \varphi_n^{(k)}\right)\right). \quad (31)$$

Then, we marginalized the above Poisson probability and obtained the gamma distribution  $\theta_j^{(t)}$ . Then, we have

$$q_n^{(k)(k+1)} \sim NB\left(\Omega^{(k+1)} h_n^{(k+1)}, \varphi_n^{(k+1)}\right) \quad (32)$$

where  $x \sim NB(r, p)$  denotes that  $x$  obeys the negative binomial distribution of parameters  $r$  and  $p$ , and its probability density function is

$$f(x; r, p) = \binom{k+r-1}{r-1} p^r (1-p)^k \quad (33)$$

where  $p$  is the probability of a successful Bernoulli experiment,  $k$  is the number of failures, and  $r$  is the number of successes.

The  $(k+1)$ th layer can be derived from the  $k$ th layer hidden variable  $u_{mn}^{(k)}$  via (26) and (28)

$$\begin{aligned} & \left\{ \left(u_{mn1}^{(k)}, \dots, u_{mnM_k}^{(k)}\right) \mid u_{mn}^{(k)}, \emptyset_{m:}, h_n^{(k)} \right\} \\ & \sim \text{Multi}\left(u_{mn}^{(k)}, \frac{\omega_{m1}^{(k)} h_{1n}^{(k)}}{\sum_{i=1}^{M_k} \omega_{mi}^{(k)} h_{in}^{(k)}}, \dots, \frac{\omega_{mi}^{(k)} h_{in}^{(k)}}{\sum_{i=1}^{M_k} \omega_{mi}^{(k)} h_{in}^{(k)}}\right) \end{aligned} \quad (34)$$

$$\begin{aligned} & \left\{ u_{in}^{(k+1)} \mid q_{in}^{(k)(k+1)}, \omega_{i:}^{(k+1)}, h_n^{(k+1)} \right\} \\ & \sim \text{CRT}\left(q_{in}^{(k)(k+1)}, \omega_{i:}^{(k+1)}, h_n^{(k+1)}\right) \end{aligned} \quad (35)$$

where Multi is a multidistribution and CRT is a Chinese restaurant distribution [26].

### B. Up-Down Gibbs Sampling

Gibbs sampling is difficult to directly sample when it approximates a sample sequence from the multicomponent probability distribution. Since the traditional Gibbs sampling is not suitable for GDBN, the Gibbs up-down algorithm is proposed to estimate the hidden variables, and all layers of the network

are jointly trained in each iteration to sample a layer of the network. It is shown in Fig. 3. The upsampling the connection weights follows Dirichlet distribution. The downsampling implicit units obey the gamma distribution. We perform the following iterative sampling on each layer in the network.

**Sampling  $\mathbf{u}_{mni}^{(k)}$ :** Via (34),  $u_{mni}^{(k)}$  of all layers is sampled. For the first layer of hidden layers, the observation unit  $u_{mn}^{(1)}$  can be regarded as a sequence of the  $m$ th feature in the  $n$ th state. We can assign  $\{m_{nj}\}_{j=1, \dots, u_n^{(1)}}$  to the hidden factor one by one, and marginalize  $\Omega^{(1)}$  and  $h_n^{(1)}$ . For  $i \in \{1, \dots, M_{1 \max}\}$ , we have

$$P(l_{nj} = i | -) \propto \frac{\delta^{(1)} + u_{m_{nj} \cdot i}^{(1)-nj}}{M \delta^{(1)} + u_{\cdot \cdot i}^{(1)-nj}} \left( u_{ni}^{(1)-nj} + \omega_i^{(2)} h_n^{(2)} \right) \quad (36)$$

where  $l_{nj}$  is the feature tag of  $l_{nj}$ , and the symbol “.” indicates the summation of the corresponding tag, such as  $u_{ni}^{(1)} = \sum_m x_{mni}^{(1)} x^{-nj}$  means not considered the sequence of feature  $j$  in the  $n$ th state. To simplify the model, a truncation step is added, that is, if  $K=1$ , the number of hidden units is limited to  $M_{1 \max}$ , and then we have  $\alpha_i \sim \text{Gamma}([\gamma_0/M_{1 \max}], [1/\lambda_0])$ .

**Sampling  $\boldsymbol{\omega}_i^{(k)}$ :**

$$\left( \omega_i^{(k)} | - \right) = \text{Dirichlet}\left(\eta^{(k)} + u_{1 \cdot i}^{(k)}, \dots, \eta^{(k)} + u_{M_{k-1} \cdot k}^{(k)}\right). \quad (37)$$

**Sampling  $\mathbf{u}_{mn}^{(k+1)}$ :** Sampling  $u_n^{(k+1)}$  by (35), replacing  $\Omega^{(k+1)} h_n^{(K+1)}$  with  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_{M_K})'$ .

**Sampling  $\mathbf{h}_n^{(k)}$ :** Via (29) and the conjugate of gamma distribution to the Poisson distribution, we can sample  $h_n^{(k)}$

$$\begin{aligned} \left( h_n^{(k)} | - \right) &= \text{Gamma}\left(\Omega^{(k+1)} h_n^{(k+1)} + q_n^{(k)(k+1)}, \right. \\ &\quad \left. \left(\lambda_n^{(k+1)} - \ln\left(1 - \varphi_n^{(k)}\right)\right)^{-1}\right). \end{aligned} \quad (38)$$

**Sampling  $\boldsymbol{\alpha}$ :**

$$\begin{aligned} \left( \alpha_m | - \right) &= \text{Gamma}\left(\frac{\gamma_0}{M_K} + u_{m \cdot}^{(K+1)}, \right. \\ &\quad \left. \left(\lambda_0 - \sum_n \ln\left(1 - \varphi_n^{(K+1)}\right)\right)^{-1}\right). \end{aligned} \quad (39)$$

**Sampling  $\boldsymbol{\lambda}_n^{(k)}$ :** First, sample  $\varphi_n^{(2)}$

$$\left( \varphi_n^{(2)} | - \right) = \text{Beta}\left(c_0 + q_{\cdot n}^{(1)(2)}, d_0 + h_{\cdot n}^{(2)}\right) \quad (40)$$

via (28), we have

$$\left( \lambda_n^{(k)} | - \right) = \text{Gamma}\left(a_0 + h_{\cdot n}^{(k)}, \left(b_0 + h_{\cdot n}^{(k-1)}\right)^{-1}\right) \quad (41)$$

where  $h_{\cdot n}^{(k)} = \sum_{i=1}^{M_k} h_{in}^{(k)}$  ( $k = 1, 2, \dots, K$ ) and  $h_{\cdot n}^{(K+1)} = \alpha$ .

## IV. EXPERIMENT

In the experimental part, we use two examples to verify the validity of our model. One is the 20 newsgroups dataset, and the other is the dataset that we collected experimentally.

**Example 1 (Classification in the 20 Newsgroups Data Set):** In this article, we consider only two types of task that distinguishes the newsgroups. We will also remove standard stop words. Terms that appear at least five times are considered. The

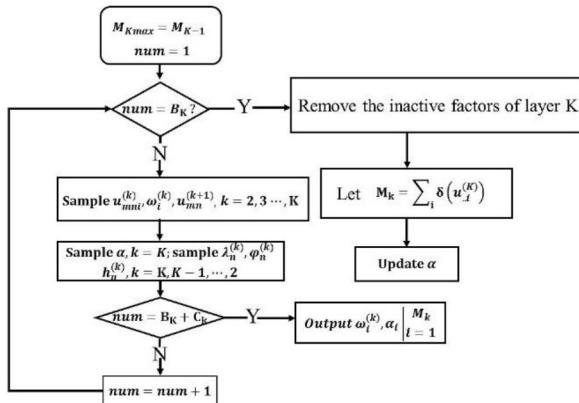


Fig. 3. Up-down Gibbs sampling.

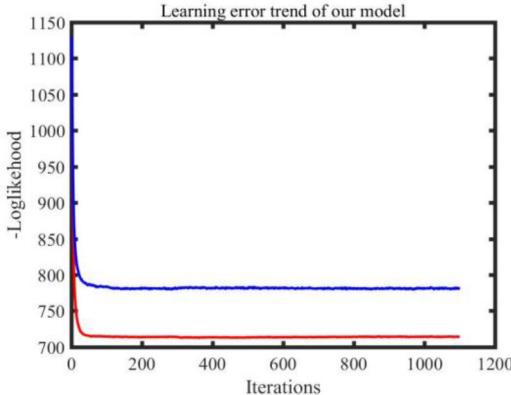


Fig. 4. Inferred network structure.

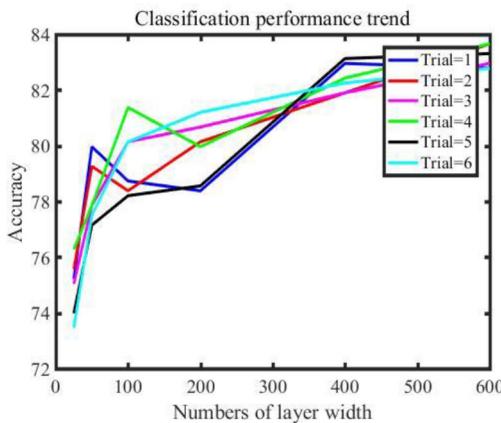


Fig. 5. Classification accuracy for 20 newsgroups.

classification accuracy based on six independent experiments is reported. The experiment set the upper limit of the first layer width to  $\{25, 50, 100, 200, 400, 600\}$ ,  $B_k = C_k = 1000$ , and  $\eta^{(k)} = 0.01$ . We use the up-down Gibbs sampling algorithm to train a network with  $K \in \{1, 2, 3, \dots, 8\}$  layers.

We provide the trend of the negative log-likelihood values of the joint distribution objective function, as shown in Fig. 4. The experiment set the upper limit of the first layer width to  $\{25, 50, 100, 200, 400, 600\}$ ,  $B_k = C_k = 1000$ , and  $\eta^{(k)} = 0.01$ . We use the up-down Gibbs sampling algorithm to train a network with  $K \in \{1, 2, 3, \dots, 8\}$  layers.



Fig. 6. Flight simulators.

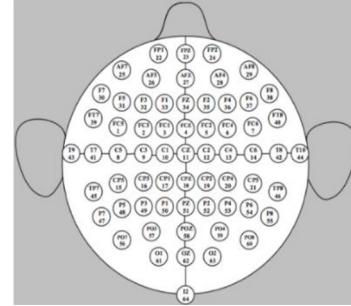


Fig. 7. International 10-20 system.

convergent. The experimental results also show that in the GDBN network, as the network level increases, the multilayer network significantly improves the quality of the unsupervised feature vector extraction.

*Example 2 (Detecting Cognitive State):* In this example, we provide a case of pilot's brain cognitive automatic detection through GDBN.

#### A. Data Collection

*Simulated Flight Equipment:* Experiments were performed on the C919 and ARJ-21 flight simulators, and the environment inside the simulator was consistent with the actual flight process. The operating system displays parameters such as the current attitude of the aircraft and engine speed. The exterior and interior of the flight simulator are shown in Fig. 6.

*Experimental Pilots:* Forty Chinese airline pilots participated in the simulated flight with an average flight time of 7173 h and an average age of 39 years. Prior to the experiment, all pilots participated in training on simulated aircraft. They are familiar with flight operations. Before the experiment, all pilots flew for 15–20 h per week. Their physiological indicators are normal. They rested the entire day before the experiment to ensure that the experiment was initially free of fatigue.

*EEG Data Acquisition:* EEG data acquisition uses an international 10-20 system, which includes 64 reference electrodes. The sampling frequency is 160 Hz. The 64 electrode distributions are shown in Fig. 7.

*Experimental Process:* The data collection process is divided into four phases. After the current phase of data collection, the pilots are required to complete drowsiness indicators (including Stanford Sleepiness Index and Karolinska Sleepiness Index). The time and process of each stage of the experiment are shown in Table II.

To ensure that the EEG signal is stable and to eliminate the interference at the beginning and end of the experiment,

TABLE II  
PROCESS OF EEG DATA COLLECTION

Stage	Time	Requirements
1	8: 30~9: 00	Sitting still
2	9: 00~10: 00	Normal simulated flight
3	10: 00~11: 00	Simulated flight in complex environments such as airflow disturbance
4	11: 00~12: 00	Simulate complex flight operations such as takeoff and landing

TABLE III  
PERIOD OF EEG SIGNAL SELECTION

Stage	Time	Cognitive Status
1	8: 40~8: 52	Non fatigue
2	9: 38~9: 50	Mild fatigue
3	10: 39~10: 51	Moderate fatigue
4	11: 38~11: 50	Severe fatigue

a 12-min signal in the middle of each experiment is selected. The selected signal time period and the corresponding status are shown in Table III. The experiment was conducted for two days, and a total of 40 samples were collected, each of which was a four-stage EEG signal. There were 36 samples in which the data at each stage of the label were consistent with the sleepiness scale. These samples were used as follow-up experimental samples. For each sample data, it is divided into independent sample signals of length 24 s. Each sample contains four EEG data for 12 min. At that time, we were able to obtain  $1080 \times 4$  independent sample signals.

### B. Feature Extraction

Before performing feature learning, we need to extract the rhythm of the EEG signal. The common methods are filters and wavelet transforms. We know that wavelet transform can only reduce low-frequency noise, while the wavelet packet can reduce noise for both high-frequency and low-frequency signals. Considering the interference in the EEG signal, we use wavelet packet transform to extract the rhythm of the EEG signal. How to choose an approximate wavelet packet function is very important. Princy *et al.* [39] proposed the insights that the best mother wavelet basis function in EEG signal decomposition is Daubechies. Here, we also show some visualization of typical wavelet functions, such as db3, db6, db10, and coiflet4, as shown in Fig. 8. Obviously, db10 is a good choice for decomposing EEG signals. We use db10 wavelet packet to extract  $\delta$ ,  $\theta$ ,  $\alpha$ , and  $\beta$  rhythm waves.

The time-domain signals, instantaneous amplitude, phase, instantaneous frequency, and energy spectrum of four rhythms are shown in Fig. 9. In order to prove the advantages of the SPAWVD algorithm, we use an SPWVD algorithm to extract the instantaneous frequency-domain information of EEG signals in the same position. These electrode positions are 35,

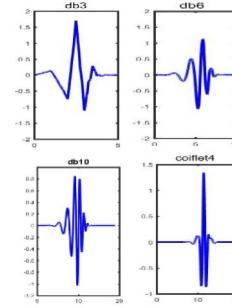


Fig. 8. Typical wavelets.

42, 54, and 62. The corresponding two time–frequency distributions are shown in Figs. 10 and 11. It can be seen that both SPAWVD and SPWVD can give instantaneous 3-D features of EEG signals, but the instantaneous local spectral information extracted by SPAWVD is more abundant and significant.

Based on the instantaneous spectrum of the rhythm acquired by SPAWVD, we construct five fatigue indicators  $\theta/\beta$ ,  $\alpha/\beta$ ,  $(\alpha + \theta)/\beta$ ,  $(\theta + \delta)/(\alpha + \beta)$ , and  $(\alpha + \theta)/(\alpha + \beta)$ . In order to select more important features, the fatigue status and nonfatigue status of each index are compared, as shown in Fig. 12. The horizontal axis represents the sample number and the vertical axis represents the power spectral density amplitude ratio in Fig. 12. It can be seen that the ratio of the five indicators is increasing from the nonfatigue status to the fatigue status, and the ratios of  $\theta/\beta$ ,  $(\alpha + \theta)/\beta$ , and  $(\alpha + \theta)/(\alpha + \beta)$  increase more. Therefore, they are input features of the cognitive state of the brain.

### C. Model Establishment and Recognition

For deep-learning networks, in the unsupervised features extraction, more layers can improve the performance of the network, and the increase in network depth will cause problems, such as slow training and overfitting. In addition, the GDBN can automatically infer the number of subsequent layers based on the width of the first layer (the number of hidden units), so the width of the first layer  $K_{1\max}$  is very important for the recognition accuracy of the network.

In this experiment, the network depths  $K = 1, 2, \dots, 5$ , and  $M_{1\max}$  are 25, 50, 100, 200, 400, 600, and 800, respectively, and the ratio of the training set to the test set is 3:1, the accuracy is shown in Figs. 13 and 14. It can be seen from Fig. 13 that for any network layer number  $K$ , as the first layer width  $M_{1\max}$  increases, the recognition accuracy rate increases, and after increasing to 200, it stabilizes and has only some fluctuations.

Fig. 14 shows the effect of the number of layers  $K$  on the recognition performance when  $M_{1\max}$  takes different values. It can be seen that increasing the number of layers has a better effects on improving the recognition accuracy. When the value of  $M_{1\max}$  is moderate, such as 100 or 200, the effect of increasing the number of layers on the accuracy is more obvious. When the value of  $M_{1\max}$  is large, the number of nodes per layer is large, the complexity and

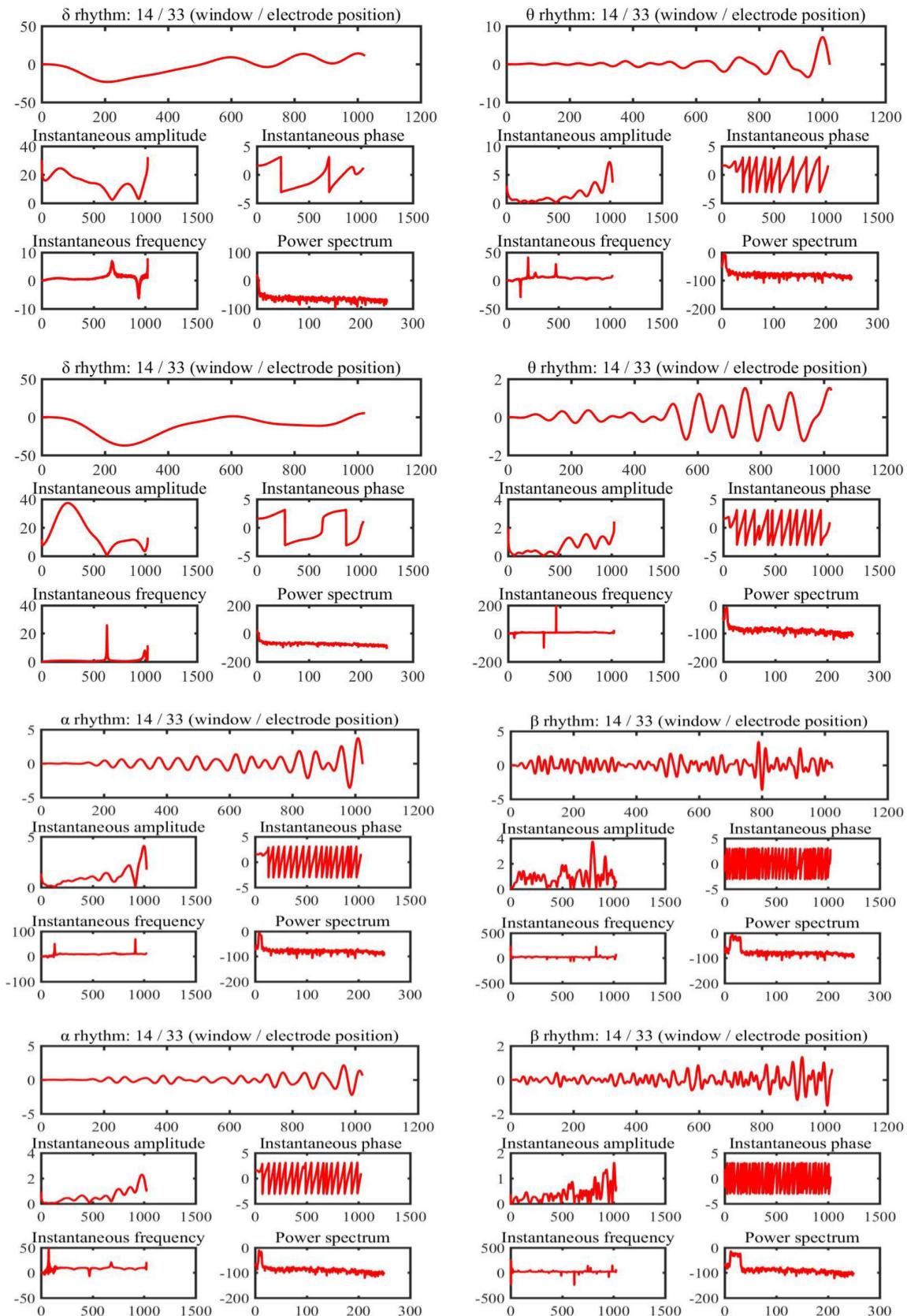


Fig. 9. Four rhythms and their instantaneous information.

prediction ability of the model increases, and the network depth increases. The improvement in model performance is limited.

Figs. 15 and 16 show the recognition accuracy and iteration time of the recognition process at each network depth, respectively,  $\eta = 0.01, 0.05, 0.1$ , and  $0.5$ . Obviously, an increase in

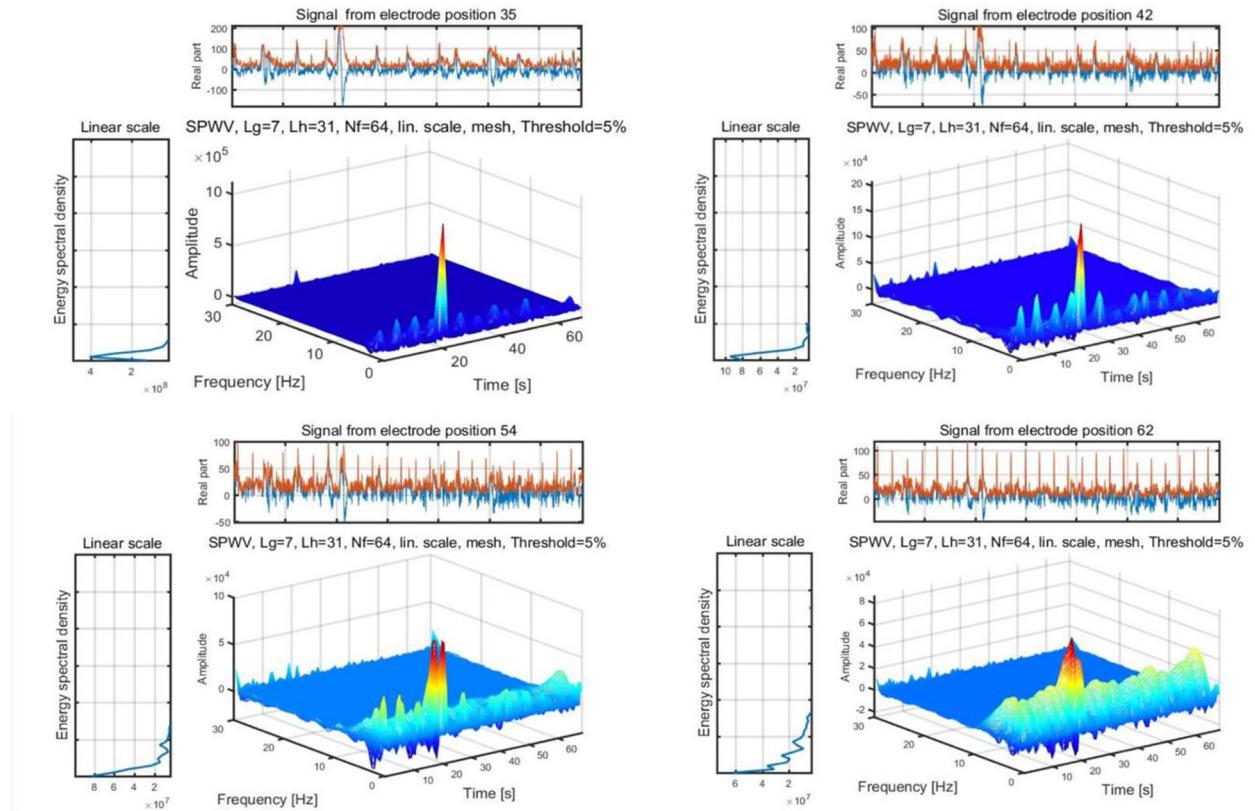


Fig. 10. Time–frequency distribution from SPWVD.

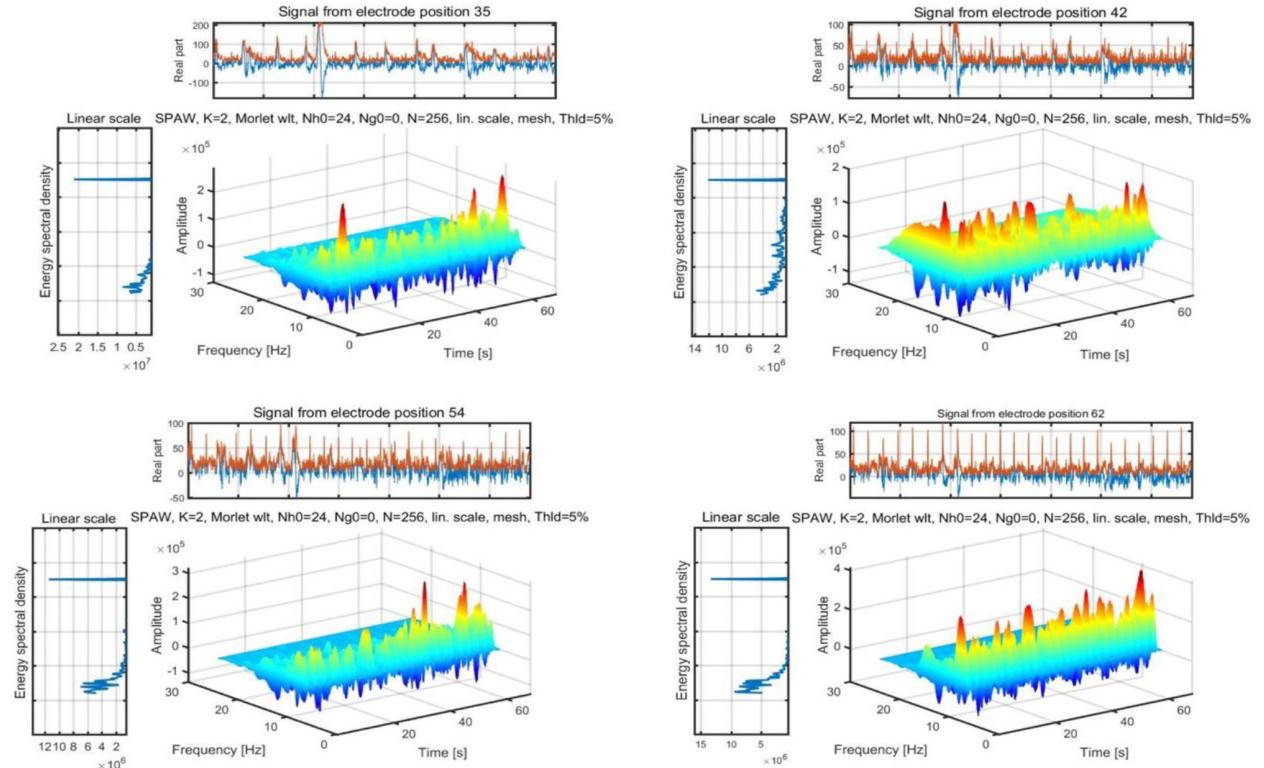


Fig. 11. Time–frequency distribution from SPAWVD.

the number of layers and a decrease in the learning rate  $\eta$  will lead to an increase in the time of each iteration. At the same time, as  $\eta$  increases, the recognition accuracy increases first

and then decreases. After weighing the model running cost and network performance, we take  $M\_1 \max = 200$ ,  $K = 4$ , and  $\eta = 0.05$  as a valid model learning parameters.

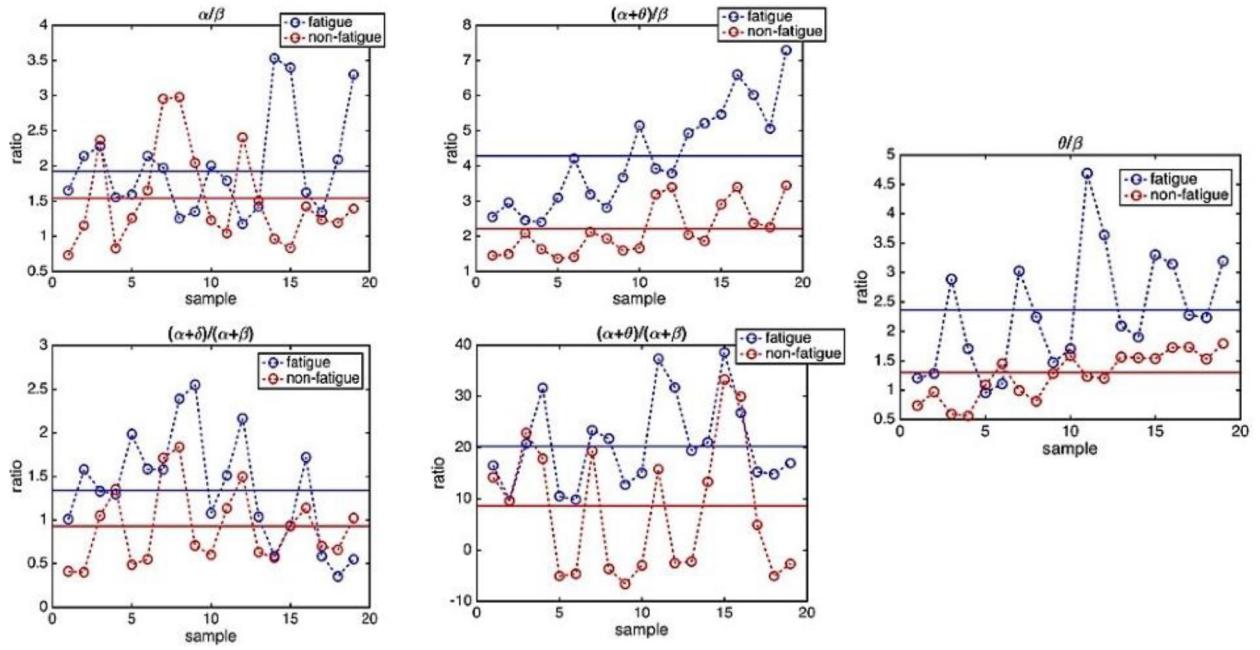
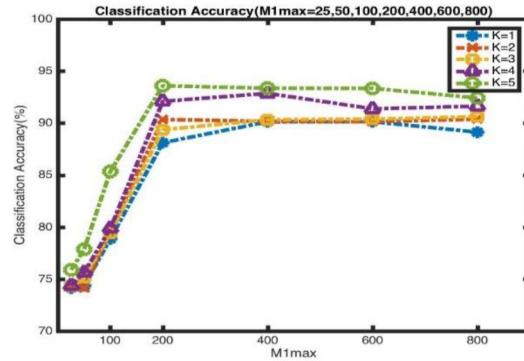
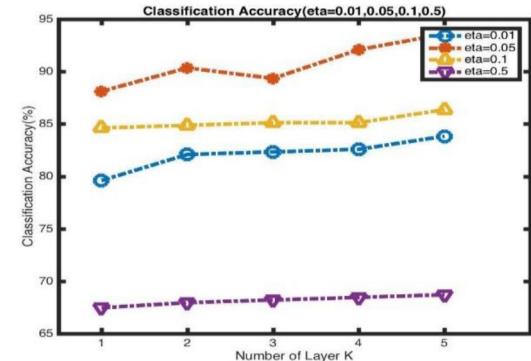
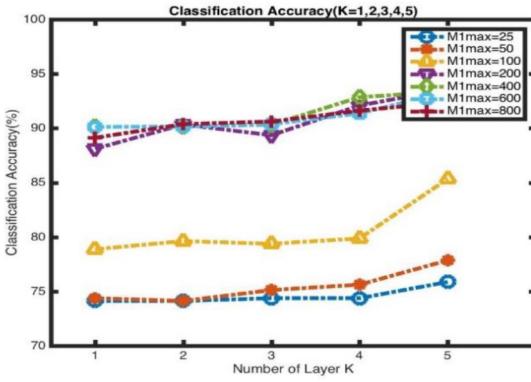
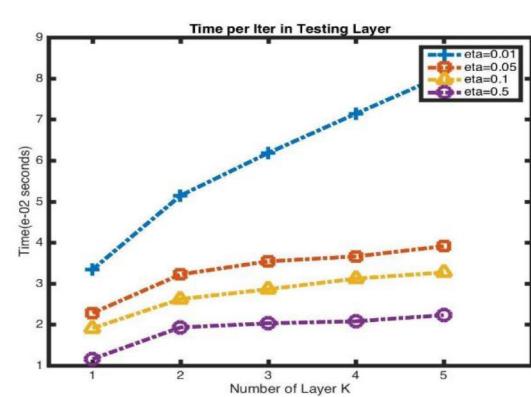


Fig. 12. Trend of cognitive status.

Fig. 13. Classification accuracy of different  $M_1 \text{ max}$ .Fig. 15. Classification accuracy of different  $\eta$ .Fig. 14. Classification accuracy of different  $K$ .Fig. 16. Iteration time of different  $\eta$ .

To further verify the recognition performance of the GDBN, different models, such as DBN, convolutional neural network (CNN), Gaussian process (GP), and support vector machine (SVM), are used for testing and comparison. In DBN, the network is initialized with a normal random distribution

whose threshold is initialized to zero. The maximum number of iterations of RBM is 200, the learning rate is 0.15, and the momentum parameter is 0.95. In CNN, the convolution kernel parameters are set as size =  $3 \times 3$ , stride = 1, and

TABLE IV  
CLASSIFICATION ACCURACY OF DIFFERENT MODELS

Classifiers	Accuracy (mean $\pm$ SD)	Testing time pre item(s)
GP [22]	0.8057 $\pm$ 0.084	0.0673
SVM [23]	0.8278 $\pm$ 0.063	0.0542
DBN[21]	0.8639 $\pm$ 0.026	<b>0.0301</b>
SDBN[24]	0.9053 $\pm$ 0.034	0.0388
CNN[27]	0.8585 $\pm$ 0.054	0.0523
GDBN	<b>0.9334<math>\pm</math>0.031</b>	0.0346

padding = 1. The number of convolution kernels at the convolutional layer 1 is 32, the number of convolution kernels at the convolutional layer 2 is 64, and the number of convolution kernels at the convolutional layer 3 is 128.

The downsampling layer uses the maximum pooling method. The size of the pooling is  $2 \times 2$ , and the full connection layer is set to connect directly to the downsampling layer. Other parameters of the CNN network are as follows: batch size = 64, the learning rate is 0.001, and the dropout of fully connected layer (FCL) is 0.5.

In this article, we processed 4230 samples at a 3:1 random ratio to form the corresponding training and test set. Four cross-validations were performed for each model to obtain recognition accuracy. As can be seen from Table IV, the GP model has the lowest accuracy and the accuracy fluctuates greatly. The recognition rate of DBN and its related models has less fluctuations and higher accuracy. In models with similar fluctuations in accuracy, the recognition rate of GDBN is higher. Obviously, our model has a good ability to learn and predict pilot fatigue.

#### D. Discussion

The gamma distribution density function has a very strong feature learning ability. However, its shape parameters have neither conjugate prior nor closed-form maximum-likelihood estimates. The prior reasoning of the shape parameters of the gamma distribution density function makes deep network with gamma hidden elements seem more attractive. We have found that the width of the first layer of the network can be controlled by the regularized Poisson factor, and the number of potential hidden units can be learned one layer at a time from the first layer to the highest hidden layer. Considering the upper and lower optimization network weighting strategy for deep learning, the Dirichlet distributed connection weight vector is upsampled layer by layer in each iteration, and then the hidden units of the gamma distribution are downsampled to form an effective up-down Gibbs sampler.

In this article, the GDBN was proposed to extract multilayer deep representations of high-dimensional input data. An effective up-down Gibbs sampler trains all its layers together and uses nonparametric learning to infer the network structure to achieve the automatic classification of high-dimensional data. Of course, when the width of the first layer is determined by

Poisson factor analysis, the prior of the subsequent layer can be inferred by the Dirichlet distribution. Through more online data input, the entire network can grow without restrictions. We also noticed that when the width of the first layer is limited, the depth of the network can be increased to match the growth of the data set. Of course, we also found that the network performance with limited width and infinite depth perform better than the shallower networks with a wider first layer.

#### V. CONCLUSION

Due to the complexity of EEG signals, fatigue judgment of artificial experience is only a qualitative evaluation. This article established a new DBN, further studied the fatigue characteristics, obtained the abstract features of the cognitive state of the brain, and achieved more accurate fatigue state inference and recognition. Therefore, the advantages of the proposed method in this article can be summarized as follows.

- 1) FFT can only provide average spectrum information for a period of time, but not instantaneous frequency-domain information of signal points. The instantaneous information of brain fatigue is more conducive to real-time brain fatigue monitoring. Therefore, this article proposes an instantaneous spectrum calculation method using SPAWVD instead of FFT.
- 2) In order to simulate the strong correlation and periodicity between the characteristics of each channel, a Poisson factorization is performed on the multidimensional input of the network, and the factor parameters obey the gamma distribution. The structure of GDBN is inferred by the shape parameters of the gamma distribution. It can infer the multilayer representation of feature vectors and capture the correlation between neurons and weights in all layers.
- 3) This article designs an up-down Gibbs sampling method to derive GDBN model parameters. We give the number of hidden cells in the first layer, and then infer the number of hidden cells layer by layer in a greedy layered manner. This learning mechanism avoids the need to adjust the number of hidden cells and the depth of each layer in the DBN and learns the cognitive features of the brain with the simplest possible network structure.

The experimental results show that GBDN has a better ability to infer pilot brain fatigue. The recognition accuracy reached 93.34%, which is higher than DBN and related networks. In addition, there is no significant increase in the iteration time of GBDN. Therefore, the GDBN has a good recognition performance and an effective learning efficiency.

#### REFERENCES

- [1] M. A. Boksem, T. F. Meijman, and M. M. Lorist, "Effects of mental fatigue on attention. An ERP study," *Cogn. Brain Res.*, vol. 25, no. 1, pp. 107–116, 2005.
- [2] J. Liu and L. Zhang, "Semantic network analysis of flight fatigue events," *China Safety Sci. J.*, vol. 26, no. 1, pp. 34–39, 2016.
- [3] W. P. Moroney, D. W. Biers, F. T. Eggemeier, and J. A. Mitchell, "A comparison of two scoring procedures with the NASA task load index in a simulated flight task," in *Proc. IEEE Nat. Aerosp. Electron. Conf.*, 1992, pp. 734–740.

- [4] K. Kaida *et al.*, "Validation of the Karolinska sleepiness scale against performance and EEG variables," *Clin. Neurophysiol.*, vol. 117, no. 7, pp. 1574–1581, 2006.
- [5] S. Anumas and S. Kim, "Driver fatigue monitoring system using video face images & physiological information," in *Proc. 4th Biomed. Eng. Int. Conf.*, 2012, pp. 125–130.
- [6] K. Jiao, Z. Li, M. Chen, and C. Wang, "Synthetic effect analysis of heart rate variability and blood pressure variability on driving mental fatigue," *J. Biomed. Eng.*, vol. 22, no. 2, pp. 343–346, 2005.
- [7] A. Zhang, Z. Zhao, and Y. Wang, "Estimating VDT visual fatigue based on the features of ECG waveform," in *Proc. Int. Workshop Inf. Security Appl.*, 2009, pp. 446–449.
- [8] A. Nijholt *et al.*, "Brain–computer interfacing for intelligent systems," *IEEE Intell. Syst.*, vol. 23, no. 3, pp. 72–79, May/Jun. 2008.
- [9] T. M. Vaughan, "Brain–computer interface technology: A review of the second international meeting," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 2, pp. 94–109, Jul. 2003.
- [10] F.-C. Lin, L.-W. Ko, C.-H. Chuang, T.-P. Su, and C.-T. Lin, "Generalized EEG-based drowsiness prediction system by using a self-organizing neural fuzzy system," *IEEE Trans. Circuits Syst.*, vol. 59, no. 9, pp. 2044–2055, Sep. 2012.
- [11] M. Diykh, Y. Li, and P. Wen, "EEG sleep stages classification based on time domain features and structural graph similarity," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 11, pp. 1159–1168, Nov. 2016.
- [12] B. Zou, Y. Liu, M. Guo, and Y. Wang, "EEG-based assessment of stereoscopic 3D visual fatigue caused by vergence-accommodation conflict," *J. Display Technol.*, vol. 11, no. 12, pp. 1076–1083, 2015.
- [13] A. Craig, Y. Tran, N. Wijesuriya, and H. T. Nguyen, "Regional brain wave activity changes associated with fatigue," *Psychophysiology*, vol. 49, no. 4, pp. 574–582, 2012.
- [14] S. K. Lal, A. Craig, P. Boord, L. Kirkup, and H. Nguyen, "Development of an algorithm for an EEG-based driver fatigue countermeasure," *J. Safety Res.*, vol. 34, no. 3, pp. 321–328, 2003.
- [15] M. Azarnoush, A. M. Nasrabadi, M. R. Mohammadi, and M. Firoozabadi, "Investigation of mental fatigue through EEG signal processing based on nonlinear analysis: Symbolic dynamics," *Chaos Solitons Fractals*, vol. 44, no. 12, pp. 1054–1062, 2011.
- [16] S. K. Lal and A. Craig, "A critical review of the psychophysiology of driver fatigue," *Biol. Psychol.*, vol. 55, no. 3, pp. 173–194, 2001.
- [17] C. Papadelis *et al.*, "Indicators of sleepiness in an ambulatory EEG study of night driving," in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2006, pp. 6201–6204.
- [18] B. T. Jap, S. Lal, P. Fischer, and E. Bekiaris, "Using EEG spectral components to assess algorithms for detecting fatigue," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 2352–2359, 2009.
- [19] H. Cai, X. Sha, X. Han, S. Wei, and B. Hu, "Pervasive EEG diagnosis of depression using deep belief network with three-electrodes EEG collector," in *Proc. IEEE Int. Conf. Bioinformat. Biomed.*, 2016, pp. 1239–1246.
- [20] L. K. Saul, T. Jaakkola, and M. I. Jordan, "Mean field theory for sigmoid belief networks," *J. Artif. Intell. Res.*, vol. 4, no. 1, pp. 61–76, 1996.
- [21] G. E. Hinton, S. Osindero, and Y. W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.
- [22] E. Q. Wu *et al.*, "Fault diagnosis of rotating machinery using Gaussian process and EEMD-treelet," *Int. J. Adapt. Control Signal Process.*, vol. 33, no. 1, pp. 52–73, 2019.
- [23] Q. Wu and R. Law, "Complex system fault diagnosis based on a fuzzy robust wavelet support vector classifier and an adaptive Gaussian particle swarm optimization," *Inf. Sci.*, vol. 180, no. 23, pp. 4514–4528, 2010.
- [24] R. Chai *et al.*, "Improving EEG-based driver fatigue classification using sparse-deep belief networks," *Front. Neurosci.*, vol. 11, pp. 103–117, Mar. 2017.
- [25] M. Zhou, L. Hannah, D. B. Dunson, and L. Carin, "Beta-negative binomial process and Poisson factor analysis," in *Proc. Int. Conf. Artif. Intell. Stat.*, 2012, pp. 1462–1471.
- [26] D. Blackwell and J. B. MacQueen, "Ferguson distributions via Polya Urn schemes," *Ann. Stat.*, vol. 1, no. 2, pp. 353–355, 1973.
- [27] J. X. Chen *et al.*, "Accurate EEG-based emotion recognition on combined features using deep convolutional neural networks," *IEEE Access*, vol. 7, pp. 44317–44328, 2019.
- [28] Y. L. Cun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.
- [29] R. Chai *et al.*, "Classification of EEG based-mental fatigue using principal component analysis and Bayesian neural network," in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2016, pp. 4654–4657.
- [30] C. Zhang, K. C. Tan, H. Li, and G. S. Hong, "A cost-sensitive deep belief network for imbalanced classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 1, pp. 109–122, Jan. 2019.
- [31] 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.*, early access, doi: [10.1109/TCYB.2019.2924237](https://doi.org/10.1109/TCYB.2019.2924237).
- [32] J. Jin, Y. Y. Miao, I. Daly, C. Zuo, D. W. Hu, and A. Cichocki, "Correlation-based channel selection and regularized feature optimization for MI-based BCI," *Neural Netw.*, vol. 118, pp. 262–270, Oct. 2019.
- [33] J. K. Feng *et al.*, "Towards correlation-based time window selection method for motor imagery BCIs," *Neural Netw.*, vol. 102, pp. 87–95, Jun. 2018.
- [34] 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.*, early access, doi: [10.1109/TCYB.2019.2905157](https://doi.org/10.1109/TCYB.2019.2905157).
- [35] G. M. Wang, J. F. Qiao, J. Bi, W. J. Li, and M. C. Zhou, "TL-GDBN: Growing deep belief network with transfer learning," *IEEE Trans. Autom. Sci. Eng.*, vol. 16, no. 2, pp. 874–885, Apr. 2019.
- [36] D. P. Bertsekas, "Feature-based aggregation and deep reinforcement learning: A survey and some new implementations," *IEEE/CAA J. Automatica Sinica*, vol. 6, no. 1, pp. 1–31, Jan. 2019.
- [37] C. Liu, Y. Fu, J. Yang, X. Xiong, H. Sun, and Z. Yu, "Discrimination of motor imagery patterns by electroencephalogram phase synchronization combined with frequency band energy," *IEEE/CAA J. Automatica Sinica*, vol. 4, no. 3, pp. 551–557, Jul. 2017.
- [38] 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.
- [39] R. Princy, P. Thamarai, and B. Karthik, "De-noising EEG signal using wavelet transform," *Int. J. Adv. Res. Comput. Eng. Technol.*, vol. 4, no. 3, pp. 1070–1074, 2015.
- [40] J. M. O'Toole, M. Mesbah, and B. Boashash, "A new discrete analytic signal for reducing aliasing in the discrete Wigner–Ville distribution," *IEEE Trans. Signal Process.*, vol. 6, no. 11, pp. 5427–5434, Nov. 2008.
- [41] C. Yang, Z. Li, J. Yuan, and X. Zhang, "Fractional-order smoothed pseudo Wigner–Ville distribution and its applications in machinery fault diagnosis," in *Proc. Prognostics Syst. Health Manag. Conf. (PHM-Harbin)*, 2017, pp. 1–6.
- [42] H. Fang, N. Tian, Y. Wang, M. C. Zhou, and M. A. Haile, "Nonlinear Bayesian estimation: From Kalman filtering to a broader horizon," *IEEE/CAA J. Automatica Sinica*, vol. 5, no. 2, pp. 401–417, Mar. 2018.



**Edmond Q. Wu** is currently pursuing the Ph.D. degree with the Institute of Aerospace Science and Technology, Shanghai Jiao Tong University, Shanghai, China.

He is an Associate Professor with the Key Laboratory of System Control and Information Processing, Ministry of Education, Shanghai Jiao Tong University. His research interests include deep learning, cognitive modeling, and industrial information processing.



**Dewen Hu** 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., where he was promoted as a Professor in 1996. His research interests include image processing, system identification and control, neural networks, and cognitive science.

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



**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 currently a Professor with the Science and Technology on Avionics Integration Laboratory, China National Aeronautical Radio Electronics Research Institute, Shanghai, China, where he is the Deputy Director and the Chief Technology Officer. His research interests include avionics system design, analysis, and architecture.



**Wen-Ming Zhang** (Member, IEEE) received the Ph.D. degree in mechanical engineering from Shanghai Jiao Tong University, Shanghai, China, in 2006.

He is currently a Distinguished Professor with the State Key Laboratory of Mechanical System and Vibration, School of Mechanical Engineering, Shanghai Jiao Tong University. He has been involved in the dynamics and control for micro/nanoelectromechanical systems. His current research interests include nonlinear vibration and control, smart materials and structures, rotor dynamics, vibration energy harvesting, dynamics of fluid-structure interaction, and soft robotics.



**Zhiri Tang** received the bachelor's and master's degrees in microelectronics from Wuhan University, Wuhan, China, in 2017 and 2019, respectively. He is currently pursuing the Ph.D. degree with the Department of Computer Science, City University of Hong Kong, Hong Kong.

His research interests include machine learning and cognitive computing.



**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, Shanghai, China. 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 has been an Associate Editor of the IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING. He is currently an Associate Editor of the IEEE/ASME TRANSACTIONS ON MECHATRONICS.



**Yulian Cao** (Member, IEEE) received the Ph.D. degree in logistics engineering from the Wuhan University of Technology, Wuhan, China, in 2018.

She has been a Postdoctoral Fellow with the University of New South Wales, Sydney, Australia, since 2018. Her research involves evolutionary computation and swarm intelligence.



**He Ren** received the Ph.D. degree in aerospace engineering from Northwestern Polytechnical University, Xi'an, China, in 1987.

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