# JOINT SEMI-SUPERVISED FEATURE AUTO-WEIGHTING AND CLASSIFICATION MODEL FOR EEG-BASED CROSS-SUBJECT SLEEP QUALITY EVALUATION

*Yong Peng*[1,2,*], *Qingxi Li*[1], *Wanzeng Kong*[1], *Jianhai Zhang*[1], *Bao-Liang Lu*[3], *Andrzej Cichocki*[1,4]

[1]School of Computer Science and Technology, Hangzhou Dianzi University, China
[2]Provincial Key Lab. for Computer Information Processing Technology, Soochow University, China
[3]Department of Computer Science and Engineering, Shanghai Jiao Tong University, China
[4]Skolkovo Institute of Science and Technology, Moscow 143026, Russia
yongpeng@hdu.edu.cn

**ABSTRACT**

Measuring the sleep quality is important or even crucial for people who are engaged in dangerous jobs such as the high-speed train drivers. Since the scalp EEG data are generated by the neural activities of the brain cortex, it is collected from subjects with different hours of sleep time (4 hours, 6 hours and 8 hours) to conduct sleep quality evaluation. To suppress the cross-subject variances of EEG data, in this paper, we propose a joint feature auto-weighting and semi-supervised classification model, termed GRLSR, which is formulated by introducing an auto-weighting variable into the least square regression to adaptively and quantitatively measure the importance of each dimension of the feature. Once the model is solved, besides the measurement results, we can use the auto-weighting variable to 1) analyze the importance of each frequency band in sleep quality expression and 2) identify the capacity of different channels connecting to the sleep effect. Therefore, the proposed GRLSR is a pure data-driven computing model for EEG-based cross-subject sleep quality evaluation. Experimental results show its effectiveness.

***Index Terms***— Sleep quality evaluation, EEG, Feature auto-weighting, Semi-supervised learning, Classification

## 1. INTRODUCTION

Sleep is a normal physiological need to keep us in healthy status and sufficient sleep can make us concentrate on daily work much easier; however, if having insufficient sleep, we will feel tired and have low work efficiency. Therefore, it is of great necessity to develop reliable sleep quality evaluation approaches especially for people engaged in dangerous jobs, leading to a cutting-edge research topic across multiple disciplines such as information sciences, neuroscience, health care, and medicine. For the sleep quality evaluation, the subjective methods such as Pittsgurgh Sleep Quality Index [1], Epworth Sleep Scale [2] and Sleep Diaries [3] are time-consuming, laborious and untrustworthy and cannot satisfy the demands of accurate and efficient evaluation. Since the EEG data is the scalp response of neural activities of the cerebral cortex and can be collected by wearable acquisition devices, it has been widely used as an effective media and objective way to diverse researches such as emotion recognition [4, 5], motor imagery, and driving fatigue estimation [6, 7]. Recently, using machine learning approaches to automatically and fast EEG-based sleep quality evaluation has been a hotspot [8–11]. The $K$-nearest neighbor, support vector machine and extreme learning machines (ELM) [12, 13] are widely used to differentiate different sleep levels.

In this paper, we also focus the topic on EEG-based cross-subject sleep quality evaluation using machine learning approach. When compared with existing methods, the main innovation of our proposed GRLSR is the introduction of the auto-weighting variable based on which we can perform more delicate research besides the recognition accuracy of different levels of sleep quality. Specifically, based on the learned auto-weighting variable, we can 1) investigate the importance of each frequency band in distinguishing different sleep quality levels and 2) identify the capacity of different channels connecting to the sleep effect.

The remainder of the paper is organized as follows. Section 2 gives the model formulation and optimization of GRLSR. In section 3, we describe the paradigm of EEG data collection and the experiments of evaluating the effectiveness of GRLSR. Section 4 concludes the whole paper.

## 2. EVALUATION METHOD

### 2.1. Model Formulation

Given labeled EEG samples $(\mathbf{X}_L \in \mathbb{R}^{d\times l}, \mathbf{Y}_L \in \mathbb{R}^{l\times c}) = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{l}$ from one or more subjects, and unlabeled EEG samples $\mathbf{X}_U = (\mathbf{x}_{l+1}, \cdots, \mathbf{x}_{l+u}) \in \mathbb{R}^{d\times u}$, $n = l + u$ from other subject(s), the cross-subject sleep quality evaluation aims to estimate $\mathbf{Y}_U \in \mathbb{R}^{u\times c} = (\mathbf{y}_{l+1}, \cdots, \mathbf{y}_{l+u})$ corresponding to $\mathbf{X}_U$, where $c = 3$ is the number of sleep levels. Here, $\mathbf{y}_i \in \mathbb{R}^c$ $(1 \le i \le l)$ is a binary vector in which $\mathbf{y}_i^j = 1$ if $\mathbf{x}_i$ belongs to the $j$-th class. Inspired by [14], we want to

This work was supported by NSFC (61971173), Zhejiang Science and Technology Program (2017C33049), China Postdoctoral Science Foundation (2017M620470), Ministry of Education and Science of Russian Federation (14.756.31.0001) and Provincial Key Laboratory for Computer Information Processing Technology, Soochow University (KJS1841).

measure the importance of the $d$ features in predicting $\mathbf{Y}_U$ which can be completed by introducing an auto-weighting variable $\boldsymbol{\theta} \in \mathbb{R}^d$ and $\theta_j > 0|_{j=1}^d$ quantitatively measures the importance of the $j$-th feature. This leads to the joint learning of the feature auto-weighting factor $\boldsymbol{\theta}$ and $\mathbf{Y}_U$ by solving

$$\begin{aligned} \min_{\mathbf{W},\mathbf{b},\boldsymbol{\theta},\mathbf{Y}_U} & \|\mathbf{X}^T\boldsymbol{\Theta}\mathbf{W} + \mathbf{1}\mathbf{b}^T - \mathbf{Y}\|_F^2 + \gamma\|\mathbf{W}\|_F^2 \\ s.t.\ & \boldsymbol{\theta} > \mathbf{0}, \mathbf{1}^T\boldsymbol{\theta} = 1, \mathbf{Y}_U \geq \mathbf{0}, \mathbf{Y}_U\mathbf{1} = \mathbf{1}, \end{aligned} \tag{1}$$

where $\mathbf{X} = [\mathbf{X}_L, \mathbf{X}_U]$, $\mathbf{Y} = [\mathbf{Y}_L; \mathbf{Y}_U]$, and $\boldsymbol{\Theta} \in \mathbb{R}^{d\times d}$ is a rescaled diagonal matrix with the $j$-th element $\Theta_{jj} = \sqrt{\theta_j}$. We denote this model as RLSR.

Additionally, by considering the local consistency of data [15], we include a graph regularizer $\mathcal{R}$ into (1) to formulate the proposed Graph regularized Rescaled Linear Regression (GRLSR) model objective as

$$\begin{aligned} \min_{\mathbf{W},\mathbf{b},\boldsymbol{\theta},\mathbf{Y}_U} & \|\mathbf{X}^T\boldsymbol{\Theta}\mathbf{W} + \mathbf{1}\mathbf{b}^T - \mathbf{Y}\|_F^2 + \gamma\|\mathbf{W}\|_F^2 + \alpha\mathcal{R} \\ s.t.\ & \boldsymbol{\theta} > \mathbf{0}, \mathbf{1}^T\boldsymbol{\theta} = 1, \mathbf{Y}_U \geq \mathbf{0}, \mathbf{Y}_U\mathbf{1} = \mathbf{1}, \end{aligned} \tag{2}$$

where $\mathcal{R} \triangleq \sum_{i,j=1}^n \|\mathbf{W}^T\boldsymbol{\Theta}\mathbf{x}_i - \mathbf{W}^T\boldsymbol{\Theta}\mathbf{x}_j\|_2^2 s_{ij}$ and its compact matrix form is $\mathrm{Tr}(\boldsymbol{\Theta}^T\mathbf{W}^T\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{W}\boldsymbol{\Theta})$. $\mathbf{L}$ is the graph Laplacian matrix and can be calculated by $\mathbf{L} = \mathbf{D} - \mathbf{S}$, where $\mathbf{D}$ is a diagonal degree matrix with its $i$-th diagonal element defined by $d_{ii} = \sum_j s_{ij}$. $\mathbf{S}$ is the graph affinity matrix to depict the local data manifold and can be defined as

$$s_{ij} = \begin{cases} \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right) & \text{if } \mathbf{x}_i \in \mathcal{N}_k(\mathbf{x}_j) \text{ or } \mathbf{x}_j \in \mathcal{N}_k(\mathbf{x}_i) \\ 0 & \text{otherwise} \end{cases}$$

where $\mathcal{N}_k(\mathbf{x}_j)$ means the $k$-nearest neighbors of $\mathbf{x}_j$.

Below are some explanations of GRLSR on cross-subject sleep quality evaluation.

*First*. In GRLSR, the sleep quality level of target subject samples to be solved, $\mathbf{Y}_U$, is closely coupled with the other variables. This is superior to most existing semi-supervised models which optimize the model variables first and then perform label estimation to unlabeled samples [16].

*Second*. Once the GRLSR objective is solved, we obtain the learned auto-weighting variable $\boldsymbol{\theta}$ besides the evaluation results of sleep quality $\mathbf{Y}_U$. $\boldsymbol{\theta}$ offers us an quantitatively way to investigate the sleep quality evaluation from diverse aspects. To be specific, it can be used to a) analyze the importance of each frequency band connecting to sleep; b) identify the importance of different EEG channels in depicting the connection between brain areas and sleep. Moreover, from machine learning and pattern recognition perspective, it can be used to do feature selection and ranking.

### 2.2. Optimization

Below we derive an equivalent form of (2) which lays the foundation of its optimization.

Substitute $\boldsymbol{\Theta}\mathbf{W}$ with $\widetilde{\mathbf{W}}$ and then (2) can be written as

$$\begin{aligned} \min_{\widetilde{\mathbf{W}},\mathbf{b},\boldsymbol{\theta},\mathbf{Y}_U} & \|\mathbf{X}^T\widetilde{\mathbf{W}} + \mathbf{1}\mathbf{b}^T - \mathbf{Y}\|_F^2 + \gamma\sum_{j=1}^d \frac{\|\tilde{\mathbf{w}}^j\|_2^2}{\theta_j} \\ & + \alpha\mathrm{Tr}(\widetilde{\mathbf{W}}^T\mathbf{X}\mathbf{L}\mathbf{X}^T\widetilde{\mathbf{W}}), \\ & s.t.\ \boldsymbol{\theta} > \mathbf{0}, \mathbf{1}^T\boldsymbol{\theta} = 1, \mathbf{Y}_U \geq \mathbf{0}, \mathbf{Y}_U\mathbf{1} = \mathbf{1}. \end{aligned} \tag{3}$$

By fixing $\widetilde{\mathbf{W}}$ and $\mathbf{Y}$, $\boldsymbol{\theta}$ can be obtained by solving

$$\min_{\boldsymbol{\theta}>\mathbf{0},\boldsymbol{\theta}^T\mathbf{1}=1} \sum_{j=1}^d \left\|\tilde{\mathbf{w}}^j\right\|_2^2 / \theta_j. \tag{4}$$

Then we can get the solution to $\boldsymbol{\theta}$ as $\theta_j = \frac{\|\tilde{\mathbf{w}}^j\|_2}{\sum_{j'=1}^d \|\tilde{\mathbf{w}}^{j'}\|_2}$. As a result, we can first optimize $\widetilde{\mathbf{W}}$ and $\mathbf{b}$ by solving

$$\begin{aligned} \min_{\widetilde{\mathbf{W}},\mathbf{b},\mathbf{Y}_U\geq 0,\mathbf{Y}_U\mathbf{1}=\mathbf{1}} & \left\|\mathbf{X}^T\widetilde{\mathbf{W}} + \mathbf{1}\mathbf{b}^T - \mathbf{Y}\right\|_F^2 + \gamma\|\widetilde{\mathbf{W}}\|_{2,1}^2 \\ & + \alpha\mathrm{Tr}(\widetilde{\mathbf{W}}^T\mathbf{X}\mathbf{L}\mathbf{X}^T\widetilde{\mathbf{W}}). \end{aligned} \tag{5}$$

and then get the solution to $\boldsymbol{\theta}$ based on obtained $\widetilde{\mathbf{W}}$. Below we give the derivation in detail.

■ *Optimize* $\mathbf{b}$ *with other variables fixed.* The objective with respect to $\mathbf{b}$ is

$$\min_{\mathbf{b}} \|\mathbf{X}^T\widetilde{\mathbf{W}} + \mathbf{1}\mathbf{b}^T - \mathbf{Y}\|_F^2. \tag{6}$$

Taking its derivative with respect to $\mathbf{b}$ and setting it to zero, we get the solution to $\mathbf{b}$ as

$$\mathbf{b} = \frac{1}{n}(\mathbf{Y}^T\mathbf{1} - \widetilde{\mathbf{W}}^T\mathbf{X}\mathbf{1}). \tag{7}$$

■ *Update* $\widetilde{\mathbf{W}}$ *with other variables fixed.* By fixing $\mathbf{b}$ and $\mathbf{Y}_U$, the objective function of $\widetilde{\mathbf{W}}$ becomes

$$\min_{\widetilde{\mathbf{W}}} \|\mathbf{X}^T\widetilde{\mathbf{W}} + \mathbf{1}\mathbf{b}^T - \mathbf{Y}\|_F^2 + \gamma\|\widetilde{\mathbf{W}}\|_{2,1}^2 + \alpha\mathrm{Tr}(\widetilde{\mathbf{W}}^T\mathbf{X}\mathbf{L}\mathbf{X}^T\widetilde{\mathbf{W}}).$$

To avoid the non-differentiable problem caused by the possible existence of zero $\ell_2$-norm of rows in $\widetilde{\mathbf{W}}$, we replace $\|\widetilde{\mathbf{W}}\|_{2,1}$ with $\left(\sum_{j=1}^d \sqrt{\|\tilde{\mathbf{w}}^j\|_2^2 + \varepsilon}\right)^2$ where $\varepsilon$ is a small enough positive constant. We rewrite the above objective as

$$\begin{aligned} \min_{\widetilde{\mathbf{W}}} & \|\mathbf{X}^T\widetilde{\mathbf{W}} + \mathbf{1}\mathbf{b}^T - \mathbf{Y}\|_F^2 + \gamma(\sum_{j=1}^d \sqrt{\|\tilde{\mathbf{w}}^j\|_2^2 + \varepsilon})^2 \\ & + \alpha\mathrm{Tr}(\widetilde{\mathbf{W}}^T\mathbf{X}\mathbf{L}\mathbf{X}^T\widetilde{\mathbf{W}}). \end{aligned} \tag{8}$$

Based on the derivation in [17], taking the derivative of objective (8) with respect to $\widetilde{\mathbf{W}}$ and setting it to zero, we have

$$\widetilde{\mathbf{W}} = (\mathbf{X}\mathbf{X}^T + \gamma\mathbf{Q} + \alpha\mathbf{X}\mathbf{L}\mathbf{X}^T)^{-1}(\mathbf{X}\mathbf{1}\mathbf{b}^T - \mathbf{X}\mathbf{Y}). \tag{9}$$

By substituting $\mathbf{b}$ with (7), we have

$$\widetilde{\mathbf{W}} = (\mathbf{X}\mathbf{H}\mathbf{X}^T + \gamma\mathbf{Q} + \alpha\mathbf{X}\mathbf{L}\mathbf{X}^T)^{-1}\mathbf{X}\mathbf{H}\mathbf{Y}, \tag{10}$$

where $\mathbf{H} = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T$, and $\mathbf{Q}$ is a diagonal matrix with the $i$-th diagonal element defined as $q_{jj} = \frac{\sum_{v=1}^{d}\sqrt{\|\tilde{\mathbf{w}}^v\|_2^2+\varepsilon}}{\sqrt{\|\tilde{\mathbf{w}}^j\|_2^2+\varepsilon}}$. Note that $\mathbf{Q}$ is unknown and dependent on $\widetilde{\mathbf{W}}$, we can iteratively update $\mathbf{Q}$ and $\widetilde{\mathbf{W}}$.

■ *Update $\mathbf{Y}_U$ with other variables fixed.* By fixing $\widetilde{\mathbf{W}}$ and $\mathbf{b}$, since the problem (5) is independent for each $i$ ($l+1 \le i \le l+u$), we can optimize each $\mathbf{y}_i$ by solving

$$\min_{\mathbf{y}_i \ge \mathbf{0}, \mathbf{y}_i^T\mathbf{1}=1} \|\widetilde{\mathbf{W}}^T\mathbf{x}_i + \mathbf{b} - \mathbf{y}_i\|_2^2. \tag{11}$$

This is a problem of Euclidean projection defined on the simplex [18]. Its corresponding Lagrangian function is

$$\mathcal{L} = \|\widetilde{\mathbf{W}}^T\mathbf{x}_i + \mathbf{b} - \mathbf{y}_i\|_2^2 + \eta(\mathbf{y}_i^T\mathbf{1} - 1) - \mathbf{y}_i^T\boldsymbol{\beta}_i, \tag{12}$$

where $\eta$ and $\boldsymbol{\beta}_i$ are the Lagrangian multipliers. It can be verified the optimal solution to $\mathbf{y}_i$ is

$$\mathbf{y}_i = (\widetilde{\mathbf{W}}^T\mathbf{x}_i + \mathbf{b} + \eta)_+, \tag{13}$$

where $(var)_+ = \max(0, var)$ and $\eta$ can be obtained via $\mathbf{y}_i^T\mathbf{1} = 1$. The detailed derivation can be found in [19].

**Notations**. We use $\alpha$, $\boldsymbol{\beta}_i$, $\gamma$, $\eta$ and $\boldsymbol{\theta}$ to represent the model parameters (variables) while the five frequency bands of EEG are denoted by *Delta*, *Theta*, *Alpha*, *Beta*, *Gamma*.

## 3. EXPERIMENTS AND RESULTS

### 3.1. EEG Data Preparation

The sleep EEG data set we used was collected on 14 healthy subjects. In the sleep deprivation experiment, subjects were strictly asked to sleep at night for respectively 4, 6 and 8 hours. The raw EEG data was recorded by the ESI NeuroScan system with a 62-channel electrode according to the international 10-20 system placement. The sampling frequency of EEG was down-sampled from 1000 to 200 Hz and then EEG signals were filtered to 1-50 Hz by Butterworth bandpass filter. We extracted the differential entropy (DE) feature [9] from five frequency bands, *i.e.*, *Delta* (1-3 Hz), *Theta* (4-7 Hz), *Alpha* (8-13 Hz), *Beta* (14-30 Hz), and *Gamma* (31-50 Hz), leading to a 310-dimension representation of each sample. Linear dynamic system was used to perform feature smoothing. For each subject, we selected 900 samples and there are the same number of samples in each class for the labeled subject. We adopted a subject-to-subject sleep quality evaluation strategy, resulting in 13 tasks in total. In each task, samples from subject 1 were set as labeled while samples from each of the other 13 subjects were as unlabeled.

Figure 1 gives the schematic diagram of why the proposed GRLSR framework can perform the critical frequency bands and channels identification. Since there is obvious correspondence between $\boldsymbol{\theta}$ and the feature vector, the learned $\boldsymbol{\theta}$ can effectively characterize the importance of each dimension of the DE feature by concatenating all five frequency bands.

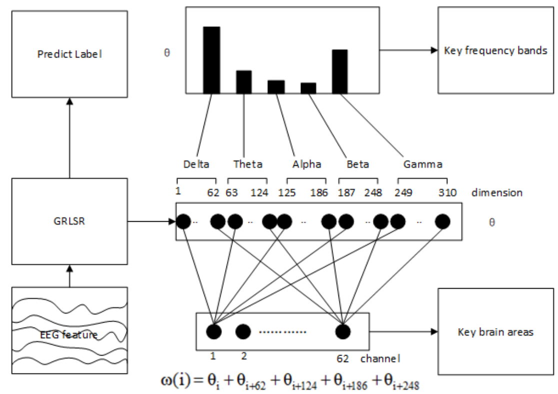


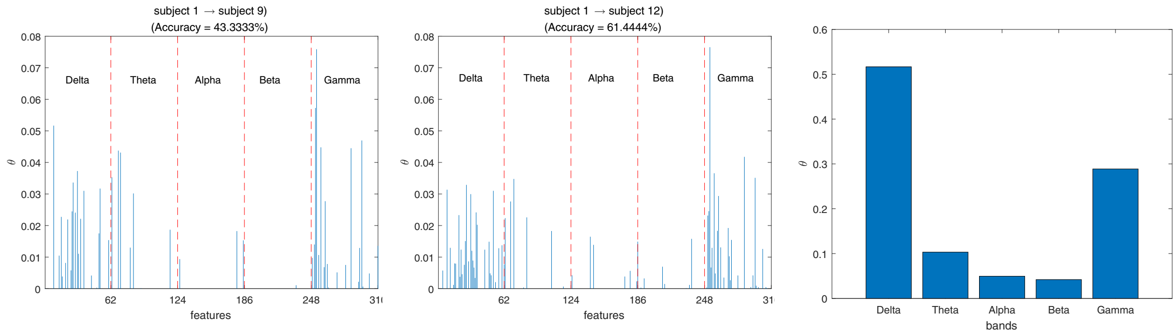

**Fig. 1**. The GRLSR-based critical frequency bands and channels identification.

### 3.2. Comparison on Evaluation Accuracy

We compare GRLSR with two supervised learning models (SVM and GELM [15]) and three semi-supervised learning models (TSVM [20], LapSVM, and ASL [21]). The parameters in GRLSR are searched from $[10^{-3}, 10^{-2}, \cdots, 10^{3}]$. The sleep quality evaluation results are shown in Table 1 in which the best results are in boldface. We can find that GRLSR outperforms the other models in most cases. As we can see, the pure supervised classification models, *i.e.* SVM, GELM, respectively achieved mean accuracies across 13 tasks as 32.88% and 54.71%. The semi-supervised models, *i.e.* TSVM, LapSVM, ASL, obtained mean accuracies of all 13 tasks as 52.57%, 48.33% and 44.91% respectively. The reason why GRLSR outperformed the other methods may be that the local consistency of data is considered. This is caused by the local invariance idea that similar inputs generate similar outputs. Then, we can conclude that 1) the graph regularization of efficiently exploiting the data manifold information is effective and 2) the utilization of unlabeled samples during the model training process is beneficial for the recognition of different levels of sleep quality.

**Table 1**. Accuracies(%) on subject-to-subject evaluation.

| | SVM | GELM | ASL | LapSVM | TSVM | RLSR | GRLSR |
|---|---|---|---|---|---|---|---|
| 1→2 | 34.22 | 59.11 | 45.56 | 53.22 | 45.89 | 37.55 | **66.67** |
| 1→3 | 26.55 | 51.56 | 42.22 | 55.22 | 43.78 | 44.44 | **66.67** |
| 1→4 | 25.22 | 35.00 | **39.67** | 37.22 | 38.11 | 30.11 | 30.67 |
| 1→5 | 33.33 | **66.67** | 40.44 | 38.89 | 45.56 | 44.11 | 47.00 |
| 1→6 | 33.33 | 60.00 | 42.67 | **66.67** | 35.89 | 39.88 | **66.67** |
| 1→7 | 28.55 | 49.33 | 41.11 | 33.33 | 34.22 | 39.66 | **51.67** |
| 1→8 | 23.33 | 47.44 | 48.00 | 34.33 | 41.67 | 54.55 | **58.56** |
| 1→9 | 37.00 | 45.33 | 40.67 | **55.22** | 45.78 | 43.33 | 43.33 |
| 1→10 | 45.22 | 52.33 | 45.33 | 63.56 | **68.44** | 66.66 | 66.67 |
| 1→11 | 14.11 | **66.67** | 40.67 | 45.67 | 36.89 | 49.22 | 50.33 |
| 1→12 | 33.33 | 66.67 | 45.11 | 42.56 | **100.0** | 58.22 | 61.44 |
| 1→13 | 35.11 | **67.56** | 50.78 | 35.78 | 66.89 | 36.55 | 35.78 |
| 1→14 | 58.22 | 43.67 | 61.56 | 66.67 | 80.33 | 81.77 | **88.89** |
| Avg. | 32.88 | 54.71 | 44.91 | 48.33 | 52.57 | 48.16 | **56.47** |




(a) $\boldsymbol{\theta}$ of task 1→9 (b) $\boldsymbol{\theta}$ of task 1→12 (c) average importance across tasks

**Fig. 2**. The $\boldsymbol{\theta}$ of tasks 1→9 (left) and 1→12 (middle), and the average importance of each band across all 13 tasks (right).

### 3.3. Identification of Critical Frequency Bands

In this experiment, we want to check the connection between different frequency bands and the sleep quality. Once the GRLSR is solved, we obtain the learned $\boldsymbol{\theta} \in \mathbb{R}^{310}$. Then, the importance of the *Delta* band can be measured by $\sum_{i=1}^{62} \theta_i$ and the *Theta* band is $\sum_{i=63}^{124} \theta_i$. And so on for each of the other bands. To visualize the learned $\boldsymbol{\theta}$, we randomly select the two ones corresponding to tasks 1→9 and 1→12 and respectively shown them in Figure 2(a) and 2(b), from which we can observe that most of non-zero values accumulate in the *Delta* and *Gamma* bands. Based on our understanding, the *Delta* band which usually happens in deep sleep may be more likely corresponding to the 8-hour sleep. Accordingly, the high frequency band, *Gamma*, may be more likely corresponding to the 4-hour sleep. The average importance of each frequency band is shown in Figure 2(c). Therefore, we conclude that there might exist critical frequency bands in sleep which may provide new insights to simplify the feature extraction for sleep quality evaluation.

### 3.4. Identification of Critical Channels

In this subsection, we want to check whether there exist some critical channels connecting to the sleep effect. Based on the learned $\boldsymbol{\theta} \in \mathbb{R}^{310}$, the importance of the $i$-th ($1 \leq i \leq 62$) channel, denoted as $\omega(i)$, can be calculated by $\omega(i) = \theta_i + \theta_{62+i} + \theta_{124+i} + \theta_{186+i} + \theta_{248+i}$. Then, we sorted the $\omega(i)|_{i=1}^{62}$ in descend order and recorded the indices of channels. For each number of selected channels, from 1 to 62, we ran the GRLSR and tuned its parameters to let it achieve the best performance. Then we show the comparison of GRLSR with channel selection and without channel selection on sleep quality evaluation in Figure 3. From this figure, we observe that the sleep quality evaluation performance is enhanced with a few selected channels. The number on the top of each blue bar in Figure 3 is the number of selected channels when GRLSR achieves the best performance.

In Figure 4, we visualize the selected six channels of task 1→4 which distribute mainly in the occipital lobe. From this result, we conclude that there might exist critical channels connecting to the sleep effect. This may be valuable for simplifying the EEG data collection and the customizing the wearable devices for sleep-related data acquisition.

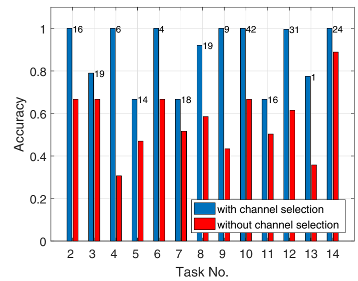



**Fig. 3**. Performance of GRLSR on features with and without channel selection.

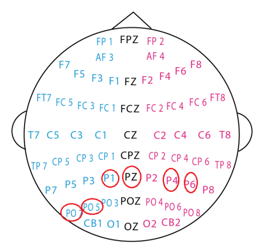


**Fig. 4**. The selected channels by GRLSR on task 1→4.

## 4. CONCLUSION

In this paper, we proposed a joint semi-supervised feature auto-weighting and classification model, termed GRLSR, for EEG-based cross-subject sleep quality evaluation. The main contribution of GRLSR is the introduction of the auto-weighting variable, which allows us to identify the critical frequency bands and channels connecting to the sleep effect. After completing the optimization of GRLSR, the importance of different frequency bands and channels can be automatically determined. Experimental results demonstrated the effectiveness of the proposed GRLSR model. This conclusions obtained by the proposed data-driven computing model may provide new insights to cognitive and neuroscientist who are conducting the sleep research.

## 5. REFERENCES


[1] Daniel J. Buysse, Charles F. Reynolds III, Timothy H. Monk, Susan R. Berman, and David J. Kupfer, "The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research," *Psychiatry Research*, vol. 28, no. 2, pp. 193–213, 1989.

[2] Murray W. Johns, "A new method for measuring daytime sleepiness: the Epworth sleepiness scale," *Sleep*, vol. 14, no. 6, pp. 540–545, 1991.

[3] Timothy H. Monk, Charles F. Reynolds III, David J. Kupfer, Daniel J. Buysse, Patricia A. Coble, Amy J. Hayes, Mary Ann Machen, Susan R. Petrie, and Angela M. Ritenour, "The pittsburgh sleep diary," *Journal of Sleep Research*, vol. 3, no. 2, pp. 111–120, 1994.

[4] Viktor Rozgic, Shiv N. Vitaladevuni, and Rohit Prasad, "Robust EEG emotion classification using segment level decision fusion," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 1286–1290.

[5] Wei-Long Zheng, Jia-Yi Zhu, Yong Peng, and Bao-Liang Lu, "EEG-based emotion classification using deep belief networks," in *IEEE International Conference on Multimedia and Expo*, 2014, pp. 1–6.

[6] Wanzeng Kong, Zhanpeng Zhou, Bei Jiang, Fabio Babiloni, and Gianluca Borghini, "Assessment of driving fatigue based on intra/inter-region phase synchronization," *Neurocomputing*, vol. 219, pp. 474–482, 2017.

[7] Wei-Long Zheng and Bao-Liang Lu, "A multimodal approach to estimating vigilance using EEG and forehead EOG," *Journal of Neural Engineering*, vol. 14, no. 2, pp. 026017, 2017.

[8] Li-Li Wang, Wei-Long Zheng, Hai-Wei Ma, and Bao-Liang Lu, "Measuring sleep quality from EEG with machine learning approaches," in *International Joint Conference on Neural Networks*, 2016, pp. 905–912.

[9] Xing-Zan Zhang, Wei-Long Zheng, and Bao-Liang Lu, "EEG-based sleep quality evaluation with deep transfer learning," in *International Conference on Neural Information Processing*, 2017, pp. 543–552.

[10] Jia-Jun Tong, Yun Luo, Bo-Qun Ma, Wei-Long Zheng, Bao-Liang Lu, Xiao-Qi Song, and Shi-Wei Ma, "Sleep quality estimation with adversarial domain adaptation: From laboratory to real scenario," in *International Joint Conference on Neural Networks*, 2018, pp. 1–8.

[11] Martin Macas, Nela Grimova, Vaclav Gerla, Lenka Lhotska, and Elisaveta Saifutdinova, "Active learning for semiautomatic sleep staging and transitional EEG segments," in *IEEE International Conference on Bioinformatics and Biomedicine*, 2018, pp. 2621–2627.

[12] Yong Peng, Jia-Yi Zhu, Wei-Long Zheng, and Bao-Liang Lu, "EEG-based emotion recognition with manifold regularized extreme learning machine," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014, pp. 974–977.

[13] Yong Peng and Bao-Liang Lu, "Discriminative manifold extreme learning machine and applications to image and EEG signal classification," *Neurocomputing*, vol. 174, pp. 265–277, 2016.

[14] Xiaojun Chen, Guowen Yuan, Feiping Nie, and Joshua Zhexue Huang, "Semi-supervised feature selection via rescaled linear regression," in *International Joint Conference on Artificial Intelligence*, 2017, pp. 1525–1531.

[15] Yong Peng, Suhang Wang, Xianzhong Long, and Bao-Liang Lu, "Discriminative graph regularized extreme learning machine and its application to face recognition," *Neurocomputing*, vol. 149, pp. 340–353, 2015.

[16] Xiaojin Zhu, Zoubin Ghahramani, and Lafferty John D., "Semi-supervised learning using gaussian fields and harmonic functions," in *International Conference on Machine Learning*, 2003, pp. 912–919.

[17] Feiping Nie, Heng Huang, Xiao Cai, and Chris H. Ding, "Efficient and robust feature selection via joint $\ell 2,1$-norms minimization," in *Advances in Neural Information Processing Systems*, 2010, pp. 1813–1821.

[18] Anastasios Kyrillidis, Stephen Becker, Volkan Cevher, and Christoph Koch, "Sparse projections onto the simplex," in *International Conference on Machine Learning*, 2013, pp. 235–243.

[19] Feiping Nie, Sheng Yang, Rui Zhang, and Xuelong Li, "A general framework for auto-weighted feature selection via global redundancy minimization," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 5, pp. 2428–2438, 2018.

[20] Thorsten Joachims, "Transductive inference for text classification using support vector machines," in *International Conference on Machine Learning*, 1999, pp. 200–209.

[21] De Wang, Feiping Nie, and Heng Huang, "Large-scale adaptive semi-supervised learning via unified inductive and transductive model," in *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2014, pp. 482–491.