Feature Fusion-Based Capsule Network for Cross-Subject Mental Workload Classification

Yinhu Yu [1] and Junhua Li [1, 2]

[1] Wuyi University, Jiangmen 529020, China
[2] University of Essex, Colchester CO4 3SQ, UK
juhalee.bcmi@gmail.com; junhua.li@essex.ac.uk

Abstract. In a complex human-computer interaction system, estimating mental workload based on electroencephalogram (EEG) plays a vital role in the system adaption in accordance with users' mental state. Compared to within-subject classification, cross-subject classification is more challenging due to larger variation across subjects. In this paper, we targeted the cross-subject mental workload classification and attempted to improve the performance. A capsule network capturing structural relationships between features of power spectral density and brain connectivity was proposed. The comparison results showed that it achieved a cross-subject classification accuracy of 45.11%, which was superior to the compared methods (e.g., convolutional neural network and support vector machine). The results also demonstrated feature fusion positively contributed to the cross-subject workload classification. Our study could benefit the future development of a real-time workload detection system unspecific to subjects.

Keywords: Mental Workload Classification, Capsule Network, Feature Fusion, Cross-Subject, EEG, Brain Connectivity, Power Spectral Density

1 Introduction

With the prevalence of human-machine interactive systems, mental demand is dramatically increased to result in high mental workload. Excessive mental workload would quickly cause fatigue so that performance and accuracy are declined. In contrast, an extremely low mental workload would lead to inefficiency. Therefore, an appropriate level of mental workload should be maintained. In order to maintain the appropriate level of workload, we have to evaluate mental workload.

Traditional methods for evaluating mental workload include National Aeronautics and Space Administration-Task Load Index (NASA-TLX), subjective scale method, primary task performance method, and auxiliary task performance method. These methods rely on humans' self-feeling and the evaluation might be influenced by a few factors such as humans' emotions. Alternative ways based on physiological information have gradually become popular as they are objective for workload evaluation [1]. To date, electroencephalogram (EEG) [2], electrocardiogram (ECG) [3], eye movement [4], and functional near-infrared spectroscopy (fNIRS) [5] have

been used for mental workload evaluation. Among them, EEG is a good choice because of its low cost, high temporal resolution, and portability.

Machine learning methods such as k-Nearest Neighbors (k-NN) [6], Random Forest (RF) [7], and Support Vector Machines (SVM) [8] were utilised to classify mental workload levels based on EEG. More recently, deep learning has shown advantages in the classification of mental workload. Convolutional neural network (CNN) is one of the deep learning models, which has been widely utilised for diverse applications, including P300 feature detection [9], seizure detection [10], and mental workload classification [11]. CNN exhibits advantages compared to the traditional machine learning methods. For example, Asgher et al. used CNN to analyse and classify mental workload levels in the n-back tasks, which outperformed SVM [12]. Although CNN has been applied to diverse applications successfully, it is not good at capturing spatial relationships between features. A new model called capsule network was proposed to overcome this drawback and is able to capture spatial relationships [13]. In addition, it is worth noting that the majority of studies performed within-subject classification for the mental workload, leaving less studies for cross-subject classification. The cross-subject classification is more difficult because there is a larger variation across subjects.

Features extracted from the time domain, spectral domain, and spatial domain can be used to classify mental workload. In the time domain, the decrease of event-related potential P300 in amplitude has been discovered to be associated with the increase of mental workload [14, 15]. In frequency domain, several studies have illustrated the associations between EEG signal frequencies and mental workload [16- 22]. Band powers (including delta, theta, alpha, beta, and gamma bands) or their ratios have been used to evaluate the levels of mental workload. For instance, Ryu et al. found that the power in the alpha band was suppressed under the high mental workload conditions [18]. Moreover, the percentage of theta power at some brain regions was significantly increased with the increase of difficulty in the simulated air traffic control (ATC) task [19]. Besides, delta, beta, and gamma bands have also been reported to be related to mental workload [20-22]. In spatial domain, since the human brain has been considered to be a large-scale network composed of various brain regions [23], brain connectivity analysis may reveal the interactions between brain regions. For instance, brain connectivity has been found to be relevant to schizophrenia [24], attention-deficit/hyperactivity disorder (ADHD) [25], autism [26] and motor imagery (MI) [27]. For the mental workload studies, it has also been adopted [7, 28]. As shown in the assessment of driving drowsiness [29], functional connectivity can provide complementary information. It implies that the combination of features from different domains may benefit the classification.

In this study, we attempted to develop a feature fusion-based capsule network to capture structural relationships between features derived from the spectral domain and spatial domain for the cross-subject classification of mental workload. We compared it to other methods (i.e., k-NN; RF; SVM; and CNN) in terms of classification accuracy and showed the detailed results in this paper. Our study addresses the shortcomings mentioned above and provides a potential solution.

2 Methods

2.1 Experiment and Dataset

A total of seven subjects were recruited for the experiment. The subjects had not attended any EEG-related experiments or flight simulation experiments previously. The institutional ethics review committee of the National University of Singapore approved the experimental protocol. All subjects signed a consent form before the start of the experiment.

In the experiment, subjects experienced different levels of manipulation difficulty in controlling an aircraft by a joystick. Oculus Rift virtual reality headset was used to display virtual 3D aircraft. The subjects started with a low difficulty task and then performed the medium and high difficulty tasks, which corresponded to low, medium, and high levels of mental workload, respectively. Each task lasted 2 minutes, resulting in a total of 6 minutes for three tasks. And each subject repeated the tasks three times. Besides, 62 EEG channels were used to record EEG data at a sampling rate of 256Hz.

2.2 Feature Extraction and Fusion

The recorded data were preprocessed to mitigate artifacts and then divided into segments with a length of two-second. This resulted in 540 segments for each subject. Each segment (62 × 512) is a sample in the following classification evaluation. Short-time Fourier transform (STFT) with a one-second sliding time window and no overlapping was used to extract power features in five bands: delta (1-4Hz), theta (4-8Hz), alpha (8-12Hz), beta (12-30Hz), and gamma (30-45Hz). This resulted in 2 features for each frequency band and each EEG channel. We collected all features to form a matrix of 62 × 10 (62 channels × 5 bands*2).

Besides, we used phase locking value (PLV) to estimate phase synchronization between EEG channels. According to our previous study [7], the dominant frequency band for PLV is the gamma band. We, therefore, extracted PLV features from this band. PLV value ranges from 0 (reflecting no phase synchronization) to 1 (reflecting perfect phase synchronization) [30-32]. PLV value between channel k and channel l in the time span $t = \{t_1, t_2, \ldots, t_n\}$ can be calculated by

$$\mathrm{PLV}_{k,l} = \langle e^{j(\varphi_k(t)-\varphi_l(t))} \rangle \tag{1}$$

where $\langle \cdot \rangle$ represents the arithmetic mean over the time span, φ_k and φ_l are the signal phases in channels k and l. After estimating each pair of channels, we obtained a connectivity matrix of 62 × 62. Subsequently, we merged the band power matrix and connectivity matrix to form a larger feature matrix of 62 × 72. After that, the features were normalized to the range [0, 1] along with the channel dimension. For PLV features, elements on the diagonal were not included for the normalization because these elements represented self-connections.

2.3 Model Architecture

The model architecture is illustrated in Fig. 1. The proposed model consists of two convolutional layers, one primarycaps layer, and one digitcaps layer. The first convolutional layer has 8 convolution filters with the kernel size of 5 × 5 and the stride of 1. Rectified linear unit (ReLU) was used as an activation function. The settings of the second convolution layer were the same as the settings of the first layer except for the number of filters (16 for the second layer). The output size of the second layer was 16×54×64. This was followed by a primarycaps layer, where the number of filters was 32, the stride was 2, and the kernel size was 5 × 5. Each primary capsule was a vector with a depth of 4, of which the length and direction represent occurrence probability and associations to each workload level.

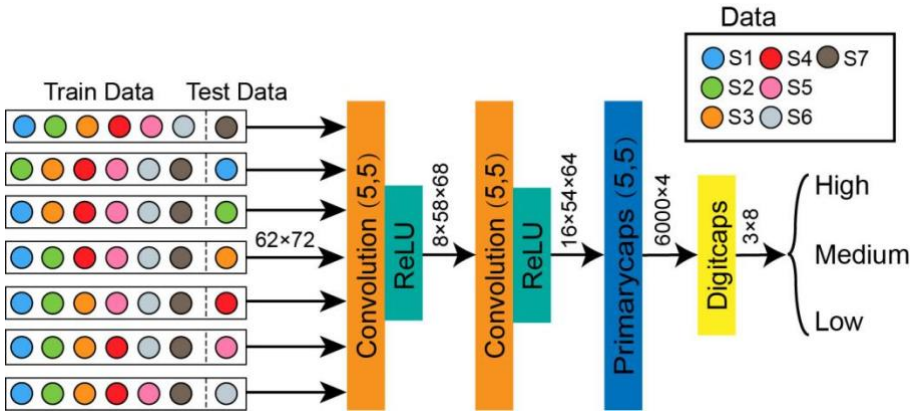

Fig. 1. The proposed model architecture. Colorful dots stand for subjects. The leave-one-subject-out was used to evaluate the model performance. The sizes of the outputs of each layer are shown in the figure.

The detailed operation process of the primarycaps layer is as follows. First, the layer used 32 filters to extract deeper features from the output of the upper layer. The features matrixes of 25 × 30 were achieved by 32 filters. Subsequently, we grouped the features with 4 as a unit to (32 / 4) * 25 * 30 primary capsules to encode the probability and low-level features related to mental workload level. We set three capsules with depth 8 in the digitcaps layer because there are three classes in our study. Capsules' length represents the probability of each mental workload level. Dynamic routing was used to train capsule layers.

2.4 Dynamic Routing

The dynamic routing mechanism [13,33] is as follows. In the first step, the *i*-th primary capsule u_i is transformed into a high-level mental workload "predicted vector" $\hat{u}_{j|i}$ through the weight matrix W_{ij} (j=1,2,3) by

$$\hat{u}_{j|i}=W_{ij}\,u_i \tag{2}$$

It represents the relative relationship between low-level mental workload features and high-level mental workload features.

In the second step, the "predicted vector" $\hat{u}_{j|i}$ is weighted and summed to obtain s_j as follows

$$s_j=\Sigma_i c_{ij}\hat{u}_{j|i} \tag{3}$$

where c_{ij} is the coupling coefficient between the *i*-th primary capsule and the *j*-th mental workload capsule, representing the probability that the *i*-th low-level primary capsule is connected to the *j*-th high-level mental workload capsule. The sum of all coupling coefficients is 1. The coupling coefficient $c_{ij}$is calculated by

$$c_{ij}=\frac{\exp(b_{ij})}{\Sigma_k \exp(b_{ik})} \tag{4}$$

where b_{ij} is the log prior probability of the *i*-th primary capsule connected to the *j*-th mental workload capsule.

In the third step, the nonlinear function is used to compress s_j to obtain the vector output v_j of the *j*-th mental workload capsule by

$$v_j=\frac{\|s_j\|^2}{1+\|s_j\|^2}\frac{s_j}{\|s_j\|} \tag{5}$$

This operation can ensure that the length of the mental workload capsule vector is between 0 and 1. We initialized log prior probability b_{ij} by zeros and updated them in the routing process by

$$b_{ij}\leftarrow b_{ij}+\hat{u}_{j|i}\cdot v_j \tag{6}$$

where · stands for the scalar product. Iteration is stopped until the predefined maximum number of the iteration is reached. This iteration process can increase weights for the features closely associated with one mental workload level while decreasing the weights for the other features.

2.5 Loss Function

The margin loss and the reconstruction loss were used as the optimization objective of the model. The margin loss is calculated by

$$L_k=T_k \max(0,m^+-\|v_k\|)^2+\lambda(1-T_k)\max(0,\|v_k\|-m^-)^2 \tag{7}$$

where T_k is an indicator of the class. When the mental workload of class *k* is present, T_k is equal to 1 (otherwise T_k = 0). m^+ and m^- are set as 0.9 and 0.1, respectively. λ is a coefficient for adjusting the proportion of the loss for absent mental workload classes and is set as 0.5 in our case.

A reconstruction loss was used additionally to encourage the mental workload capsules to encode the instantiation parameters of the input mental workload. The

reconstruction loss is calculated by mean square error (MSE). We scaled down the reconstruction loss by 0.00001 so that it did not dominate the margin loss during training. In the end, the total loss was the sum of the margin losses of all mental workload capsules and the reconstruction losses.

Model training was terminated when the maximum number of iterations (i.e., 200) was reached or the average loss was less than 10^{-5}. Moreover, we adopted a decaying learning rate. In other words, the learning rate was gradually reduced along with the iterations. This could help reduce the frequency of the fluctuation during the training, especially for the time around the minimum loss. The learning rate was changed after each iteration and was calculated by

$$lr = lr \times a^{epoch} \tag{8}$$

where lr represents the learning rate, a represents the base of the decaying learning rate, and $epoch$ represents the number of iterations until the current epoch.

3 Result

3.1 Methods Comparison

We performed a leave-one-subject-out scheme to evaluate the performance of the methods. Specifically, all data of a subject were used for testing while the data of the remaining subjects were used for training. This was repeated until every subject was in the testing set once. The final accuracies averaging across all subjects were reported in the format of mean ± standard deviation in this paper.

In this study, we not only compared different input features in the capsule network but also compared the capsule network with other mental workload classification methods (i.e., k-NN, SVM, RF, and CNN). CNN consists of convolutional layers, max-pooling layer, fully connected layer, and softmax. The input data were kept the same for all methods and the models were tuned to be as good as they can.

As shown in Fig. 2, the capsule network with the feature fusion of PLV and PSD achieved an average testing accuracy of 45.11 % ± 6.82 %, which was the best performance in the method comparison. The parameter settings of the model can be found in Table 1.

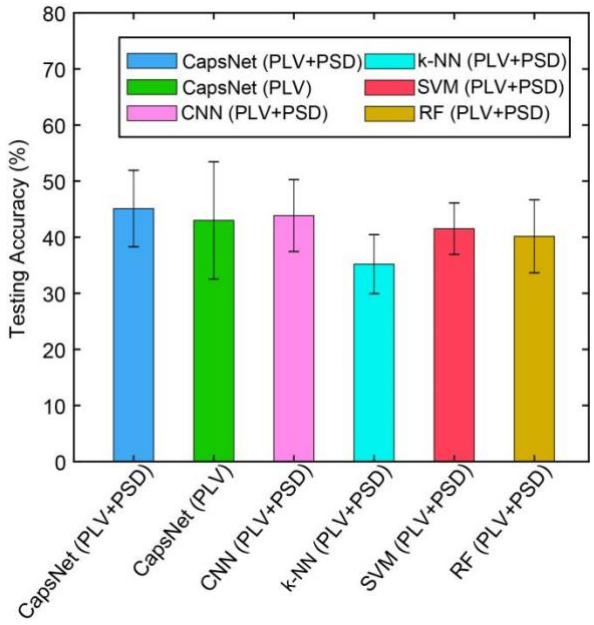

Fig. 2. Method comparisons in terms of testing accuracy

Table 1. Parameters of the capsule network model

	Name of The Parameter	Value
Training Settings	Initial Learning Rate	0.001
	Base of Decaying Learning Rate	0.9
	Weight of Reconstruction Loss	0.00001
	Maximum No. of Epochs	200
	Batch Size	20
Convolution Layer	Kernel Size	5×5
	Padding	0
	Stride	1
Convolution Layer	Kernel Size	5×5
	Padding	0
	Stride	1
Capsule Layers 1	Kernel Size	5×5
	Padding	0
	Stride	2
	Vector Length	4
Capsule Layers 2	Routing No.	3
	Vector Length	8

The second-highest testing accuracy was 43.86% ± 6.41%, which was achieved by CNN in the case of feature fusion of PLV and PSD. The methods k-NN, SVM, and RF achieved accuracies of 35.21% ± 5.25%, 41.53% ± 4.59%, and 40.16% ± 6.50%, respectively (see Fig. 2). The detail of testing accuracies for each subject can be found in Table 2. The results showed that deep learning models outperformed the traditional methods. It implies that deep learning models have a high capacity to learn essential information from EEG data.

Table 2. Comparison of testing accuracies under different methods

Methods (%)	S1	S2	S3	S4	S5	S6	S7	Mean ± Standard Deviation
CapsNet (PLV+PSD)	57.04	43.15	41.11	47.78	34.81	47.04	44.81	45.11±6.82
CapsNet (PLV)	64.07	41.30	38.33	44.44	36.11	31.48	45.37	43.01±10.46
CNN (PLV+PSD)	50.00	43.89	33.15	44.81	38.33	45.19	51.67	43.86±6.41
k-NN (PLV+PSD)	27.78	40.74	32.96	42.78	33.70	31.67	36.85	35.21±5.25
SVM (PLV+PSD)	39.07	47.96	40.37	36.85	36.67	47.04	42.78	41.53±4.59
RF (PLV+PSD)	51.48	38.15	35.37	38.70	33.33	37.41	46.67	40.16±6.50

Better performance in the capsule network compared to CNN might be due to the capability of capturing structural relationships between features in the capsule network. We noticed that the standard deviation was smaller and the mean was higher in the case of feature fusion compared to the single feature category of PLV. This might be because the different kinds of features complement each other to improve the robustness so that there is a relatively robust performance across subjects.

In terms of the average training accuracy, the capsule network achieved the training accuracy of 98.72% ± 0.42%, while k-NN, SVM, RF, and CNN performed accuracies of 88.81% ± 0.63%, 100%, 100%, and 96.91% ± 0.79% (see Fig.3). The respective training accuracies for each subject are listed in Table 3. It was worth noting that SVM and RF had the highest training accuracy but the lower testing accuracy. It reflected that the overfitting was obvious in these two methods for the cross-subject mental workload classification. In the case of the same input features, in addition to SVM and RF, the training accuracy of the capsule network was also relatively high (see Fig. 3). However, the capsule network achieved a better testing accuracy. Taken together, the capsule network less suffers from overfitting. In this study, we observed that feature fusion of PLV and PSD was better than single category of features in both training accuracy and testing accuracy, showing the spectral features and connectivity features are complementary.

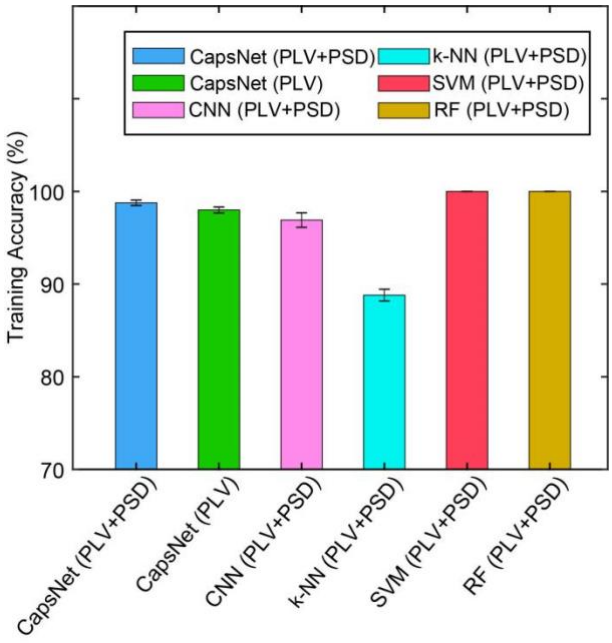

Fig. 3. Comparison of training accuracies under different methods

Table 3. Comparison of training accuracies under different methods

Methods (%)	S1	S2	S3	S4	S5	S6	S7	Mean± Standard Deviation
CapsNet (PLV+PSD)	98.95	98.61	97.96	98.46	99.04	98.80	99.20	98.72±0.42
CapsNet (PLV)	97.84	97.35	96.08	97.01	97.01	97.53	97.50	97.19±0.57
CNN (PLV+PSD)	96.39	96.42	96.42	96.24	97.93	98.15	96.85	96.91±0.79
k-NN (PLV+PSD)	88.30	89.32	87.62	88.95	89.38	89.01	89.04	88.81±0.63
SVM (PLV+PSD)	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00±0.00
RF (PLV+PSD)	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00±0.00

4 Conclusion

In this study, we targeted the difficulty of the cross-subject mental workload classification. A feature fusion-based capsule network was proposed, which captured structural relationships between features of power spectral density and brain connectivity. We demonstrated that the feature fusion-based capsule network

achieved the best performance in the cross-subject mental workload classification in terms of testing accuracy. This study suggests that the feature fusion-based capsule network is relatively robust to the large variation across subjects and could be a good candidate way for the classification with large variations.

Although the feature fusion-based capsule network achieved the best performance in the cross-subject mental workload classification, the accuracy is not very adequate to make practical usage efficient. In the future, the accuracy should be further enhanced. We also noticed the training accuracies were much higher than the testing accuracies, implying the issue of model overfitting. A further study is required to mitigate this issue. Finally, it would be better to have a larger sample size for validating the performance of models.

References

1. Radüntz T.: Dual Frequency Head Maps: A New Method for Indexing Mental Workload Continuously during Execution of Cognitive Tasks. Frontiers in Physiology 8, 1019(2017).
2. Bernhardt K. A., Poltavski D., Petros T., et al.: The effects of dynamic workload and experience on commercially available EEG cognitive state metrics in a high-fidelity air traffic control environment. Applied Ergonomics 77, 83-91(2019).
3. Qu H., Gao X., Pang L.: Classification of Mental Workload Based on Multiple Features of ECG Signals. Informatics in Medicine Unlocked 24(8), 100575 (2021).
4. Yang Y., Chen Y., Wu C., et al.: Effect of Highway Directional Signs on Driver Mental Workload and Behavior using Eye Movement and Brain Wave. Accident Analysis & Prevention 146, 105705 (2020).
5. Shimizu T., Hirose S., Obara H., et al.: Measurement of Frontal Cortex Brain Activity Attributable to the Driving Workload and Increased Attention. SAE International Journal of Passenger Cars-Mechanical Systems 2(1), 736-744 (2009).
6. Ko L. W., Chikara R. K., Lee Y. C., Lin W. C.: Exploration of User's Mental State Changes during Performing Brain-Computer Interface. Sensors 20(11), 3169 (2020).
7. Pei Z., Wang H., Bezerianos A., Li J.: EEG-Based Multi-Class Workload Identification Using Feature Fusion and Selection. IEEE Transactions on Instrumentation and Measurement 99, 1-1 (2020).
8. Lim W. L., Sourina O., Liu Y., Wang L.: EEG-based mental workload recognition related to multitasking. In 2015 10th International Conference on Information, Communications and Signal Processing (ICICS), pp. 1-4. IEEE, Singapore (2015).
9. Cecotti H., Gräser A.: Convolutional neural networks for P300 detection with application to brain-computer Interfaces. IEEE Transactions on Pattern Analysis & Machine Intelligence 33(3), 433–445 (2011).
10. Page A., Shea C., Mohsenin T.: Wearable seizure detection using convolutional neural networks with transfer learning. In 2016 IEEE International Symposium on Circuits and Systems (ISCAS), pp. 1086-1089. IEEE, Montreal (2016).
11. Zhang J., Li S., Wang R.: Pattern Recognition of Momentary Mental Workload Based on Multi-Channel Electrophysiological Data and Ensemble Convolutional Neural Networks. Frontiers in Neuroscience 11, 310 (2017).
12. Asgher U., Khalil K., Ayaz Y., Ahmad R., Khan M. J.: Classification of Mental Workload (MWL) using Support Vector Machines (SVM) and Convolutional Neural Networks

(CNN). In 2020 3rd International Conference on Computing Mathematics and Engineering Technologies, pp. 1-6. IEEE, Sukkur, Pakistan（2020).
13. Sabour S., Frosst N., Hinton G. E.: Dynamic routing between capsules. In Proc. 31st International Conference on Neural Information Processing Systems, pp. 3859-3869. Long Beach (2017).
14. Käthner I., Wriessnegger S. C., Müller-Putz G. R., et al.: Effects of mental workload and fatigue on the P300, alpha and theta band power during operation of an ERP (P300) brain-computer interface. Biological psychology 102, 118-129 (2014).
15. Pergher V., Wittevrongel B., Tournoy J., et al.: Mental workload of young and older adults gauged with ERPs and spectral power during N-Back task performance. Biological Psychology 146, 107726 (2019).
16. Mühl C., Jeunet C., Lotte F.: Eeg-based workload estimation across affective contexts. Frontiers in Neuroscience 8(8), 114 (2014).
17. Ke Y., Qi H., Zhang L., et al.: Towards an effective cross-task mental workload recognition model using electroencephalography based on feature selection and support vector machine regression. International Journal of Psychophysiology 98(2), 157–166 (2015).
18. Ryu K., Myung R.: Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic. International Journal of Industrial Ergonomics 35(11), 991-1009 (2005).
19. Brookings J. B., Wilson G. F., Swain C. R.: Psychophysiological responses to changes in workload during simulated air traffic control. Biological Psychology 42(3), 361–377 (1996).
20. Christensen J. C., Estepp J. R., Wilson G. F., Russell C. A.: The effects of day-to-day variability of physiological data on operator functional state classification. NeuroImage 59(1), 57-63 (2012).
21. Pesonen M., Hämäläinen H., Krause C. M.: Brain oscillatory 4–30 Hz responses during a visual n-back memory task with varying memory load. Brain Research 1138, 171-177 (2007).
22. Laine T. I., Bauer K. W., Lanning J. W., et al.: Selection of input features across subjects for classifying crewmember workload using artificial neural networks. IEEE Transactions on Systems Man and Cybernetics-Part A Systems and Humans 32(6), 691-704 (2002).
23. Bassett D. S., Sporns O.: Network neuroscience. Nature Neuroscience 20(3), 353-364 (2017).
24. Anticevic A., Repovs G., Krystal J. H., et al.: A broken filter: prefrontal functional connectivity abnormalities in schizophrenia during working memory interference. Schizophrenia Research 141(1), 8-14 (2012).
25. Mazaheri A., Coffeycorina S., Mangun G. R., et al.: Functional Disconnection of Frontal Cortex and Visual Cortex in Attention-Deficit/Hyperactivity Disorder. Biological Psychiatry 67(7), 617-623 (2010).
26. Jamal W., Das S., Oprescu I. A., et al.: Classification of autism spectrum disorder using supervised learning of brain connectivity measures extracted from synchrostates. Journal of Neural Engineering 11(4), 046019 (2014).
27. Wang H., Xu T., Tang C., et al.: Diverse Feature Blend Based on Filter-Bank Common Spatial Pattern and Brain Functional Connectivity for Multiple Motor Imagery Detection. IEEE Access 8, 155590-155601 (2020).
28. Kakkos I., Dimitrakopoulos G. N., SUN Y., et al.: EEG Fingerprints of Task-Independent Mental Workload Discrimination. IEEE Journal of Biomedical and Health Informatics 25(10), 3824-3833 (2021).

29. Harvy J., Thakor N., Bezerianos A., Li J.: Between-frequency topographical and dynamic high-order functional connectivity for driving drowsiness assessment. IEEE Transactions on Neural Systems and Rehabilitation Engineering 27(3), 358–367 (2019).
30. Tass P., Rosenblum M., Weule J., et al.: Detection of n: m phase locking from noisy data: application to magnetoencephalography. Physical Review Letters 81(15), 3291 (1998).
31. Celka P.: Statistical analysis of the phase-locking value," IEEE Signal Processing Letters 14(9), 577–580 (2007).
32. Aydore S., Pantazis D., Leahy R.M.: A note on the phase locking value and its properties. Neuroimage 74, 231–244 (2013).
33. Liu Y., Ding Y., Li C.: Multi-channel EEG-based Emotion Recognition via a Multi-level Features Guided Capsule Network. Computers in Biology and Medicine 123, 103927 (2020).