Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Optimized EEG–fNIRS Based Mental Workload Detection Method for Practical Applications

Hongzuo Chu (hongzuochu@gmail.com)
China Astronaut Research and Training Center https://orcid.org/0000-0001-8320-0769
Yong Cao
China Astronaut Research and Training Center
Jin Jiang
China Astronaut Research and Training Center
Jiehong Yang
China Astronaut Research and Training Center
Mengyin Huang
China Astronaut Research and Training Center
Qijie Li
China Astronaut Research and Training Center
Changhua Jiang
China Astronaut Research and Training Center
Xuejun Jiao
China Astronaut Research and Training Center

Research

Keywords: EEG, fNIRS, Mental Workload, Man–Machine Systems

Posted Date: July 24th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-683529/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Optimized EEG–fNIRS based mental workload detection method for practical applications

Hongzuo Chu[1,2,¶], Yong Cao[1,¶], Jin Jiang[1], Jiehong Yang[1,2] , Mengyin Huang[1,2] , Qijie Li[1], Changhua Jiang[1,*], Xuejun Jiao[1,2,*]

1. National Key Laboratory of Human Factors Engineering, China Astronaut Research and Training Center, Beijing, China.

2. Space Engineering University, Beijing, China.

[¶]These authors contributed equally to this work.

*Corresponding author. E-mail: 13488680126@163.com, jxjisme@sina.com

Emails:

Hongzuo Chu(hongzuochu@gmail.com); Yong Cao(caoyong_93@163.com); Jin Jiang(2583056287@qq.com); Jiehong Yang(15651075331@163.com); Mengyin Huang(yyhuangmy@163.com); Qijie Li(335279620@qq.com); Changhua Jiang(13488680126@163.com) ; Xuejun Jiao(jxjisme@sina.com);

Abstract

Background: Mental workload is a critical consideration in complex man–machine systems design. Among various mental workload detection techniques, multimodal detection techniques integrating EEG and fNIRS signals have attracted considerable attention. However, existing EEG–fNIRS-based mental workload detection methods have certain defects, such as complex signal acquisition channels and low detection accuracy, which restrict their practical application.

Method: The signal acquisition configuration was optimized and a more accurate and convenient EEG–fNIRS-based mental workload detection method was constructed. A classical MATB task was conducted with 20 participating volunteers. Subjective scale data, 64-channel EEG data, and two-channel fNIRS data were collected.

Results: A higher number of EEG channels correspond to higher detection accuracy. However, there is no obvious improvement in accuracy once the number of EEG channels reaches 26, with a four-level mental workload detection accuracy of 78.25±4.71%. Partial results of physiological analysis verify the results of previous studies, such as that the θ power of EEG and concentration of O_2Hb in the prefrontal region increase while the concentration of HHb decreases with task difficulty. It was further observed, for the first time, that the energy of each band of EEG signals was significantly different in the occipital lobe region, and the power of β_1 and β_2 bands in the occipital region increased significantly with task difficulty. The changing range and the mean amplitude of O_2Hb in high-difficulty tasks were significantly higher compared with those in low-difficulty tasks.

Conclusions: The channel configuration of EEG–fNIRS-based mental workload detection was optimized to 26 EEG channels and two frontal fNIRS channels. A four-level mental workload detection accuracy of 78.25±4.71% was obtained, which is higher than previously reported results. The proposed configuration can promote the application of mental workload detection technology in military, driving, and other complex human–computer interaction systems.

Keywords: EEG, fNIRS, Mental Workload, Man–Machine Systems

1. Introduction

Mental workload has long been a factor of immense interest in designing and applying complex human–machine systems [1], and is increasingly recognized as a serious, worldwide public health concern. Only when the mental workload is in the appropriate range can high performance and operational reliability be maintained. An irregular mental workload state will impair a person's work performance, leading to task failure [2, 3], and might endanger people's health or safety in severe cases. In human–machine systems with high safety requirements, such as in the military, aviation, aerospace, driving, and other domains, human errors such as information acquisition and judgment decision error caused by excessive mental workload are often responsible for accidents [4, 5]. For example, soldiers in a battlefield must participate in warfare for a long time without interruption and must stay alert to respond to various systems. The harsh environment poses extraordinary physical and psychological challenges to them [6, 7]. Goodman et al. [8] reported that of 1094 USAF Unmanned Aerial Vehicle(UAV) operators stationed in the US, approximately 20% reported high fatigue, 11% reported high cynicism, and 3% reported reduced performance. This is because the long-term unsuitable mental workload state seriously endangers the operator's mental and physiological health. Therefore, it is crucial to measure the mental workload of the operator accurately.

In the past several decades, neurophysiological signal have played an important role in mental workload detection because of its objectivity and stability[9]. One of the major topics to be investigated in this field is Electroencephalogram (EEG) based mental workload detection method. For instance, Georgios et al. [10] carried out a study of EEG based mental workload detection, and the experiment was implemented with N-back and mental arithmetic (the most commonly used single cognitive resource tasks for studies of mental workload [11-13]). The binary classification accuracies of mental workload are 88% and 86% in the N-back task and mental arithmetic task respectively. Besides, there are still some researches on EEG for mental workload detection in complex tasks. Hongquan Qu et al. [14] carried out a three-level Multi-Task Attribute Battery (MATB)[15] task with 32-channel electroencephalogram (EEG) acquisition. Power spectrum density (PSD) was analyzed with independent components analysis (ICA) algorithm, and the average recognition accuracy reached 79.8%. 64-channel EEG data was recorded in a simulated flight experiment, and PSD , Phase Lag Index (PLI) connection features were analyzed and extracted, giving a recognition accuracy 82% of three-level mental workload [16]. For application of mental workload detection, the significant advantage of EEG is that it contains abundant information, but it also has the disadvantages of low spatial resolution and complex operation.

As a new neurophysiological signal acquisition technology, recently, functional near infrared spectroscopy (fNIRS) has become a research hotspot in this field with the advantages of high spatial resolution and portability[17]. Reported studies proposed that fNIRS performs well in mental workload detection of both single task and complex task. For example, Asgher et al.[18] observed the brain activities of the prefrontal cortex (PFC) region with fNIRS technology in a four-level mental arithmetic task, and obtained a recognition accuracy 89.31% with classification algorithm of long short-term memory (LSTM) . Siddiquee et al.[19] explored the response difference of brain activity measured by fNIRS in different areas of the PFC. N-back task was conducted in the experiment, and the results show that the blood oxygen of middle prefrontal position can significantly improve the recognition precision, with a highest binary classification accuracy 90%. In a study of actual driving environment[20], four-channel fNIRS system was adopted to monitor the real-time change of blood oxygen in PFC region. The recognition accuracy of three-levels driving tasks reached 82.71%. Besides, in a study of air traffic control instructions tasks in flight simulators, Gateau et al.[21] collect fNIRS signals to detect the mental load state of pilots in two different group, and the accuracy reached 80% by SVM.

Aforementioned studies of mental workload detection are based on single physiological signal (either EEG, or fNIRS). Either the number of grades or the recognition accuracy of mental workload detection was not ideal, probably due to the limited information of single physiological signal. Therefore, researchers began to focus on multi-physiological signal fusion detection methods. Liu, Y. et al. **[22]** carried out a study on mental workload detection of simple cognitive resource tasks by fusing 28-channel EEG and 16-channel fNIRS. 3-level workload was induced by N-back task in the experiment, and the result indicates recognition accuracy based on fusion of EEG and fNIRS was significantly greater compared with single signal of EEG or fNIRS. Similar result was also observed in mental workload detection of complex tasks, Sangtae et al. [23] proposed a multi-physiological signal based mental workload detection method, which collected 64-channel EEG signals, and eight-channel fNIRS signals of drivers simultaneously. The result indicates that the recognition accuracy based on multi-physiological signal was significantly greater than that of single physiological signal. What's more, reported literatures [11, 24] support the above viewpoint as well. In conclusion, existing studies suggest that the combination of multiple physiological signals can obtain better performance in mental workload detection compared with signal physiological signal. Nevertheless, there are still some limitation of the reported studies in the following aspects: the multiple physiological signals acquisition configuration was relatively complex[11, 23] , only three or even less different levels of mental workload was considered[25-27], and the recognition accuracy was not ideal enough.

For practical application, the following issues must be paid attention to in real-time monitoring of mental workload in complex man-machine system. First, monitoring equipment should be portable and easy to operate, with a simple channel configuration of signal acquisition as far as possible. In addition, the response speed of the monitoring model should be as high as possible; consequently, algorithms with a large amount of computation are not suitable for this task. Finally, the gradation of the mental workload should be more detailed and the detection accuracy should be higher. This study focus on these issues, and adopted

complex simulation tasks to mimic the actual task scenarios. First, optimized data acquisition system integrates a few channels of EEG acquisition devices and a portable near-infrared devices was adopted. In addition, this study considers the modelling performance of Only-EEG feature sets, Only-fNIRS feature sets, both EEG and fNIRS feature sets under various classifiers to obtain a optimization modeling method. Last but not least, the difficulty of experiment task was divided into four levels, both of EEG and fNIRS features were took into account in the detection model of mental workload for a better modeling performance. The results of this study can promote the application of mental workload detection technology in military, driving, and other complex human–computer interaction systems.

2. Results

2.1 Behavioral Data Analysis Results

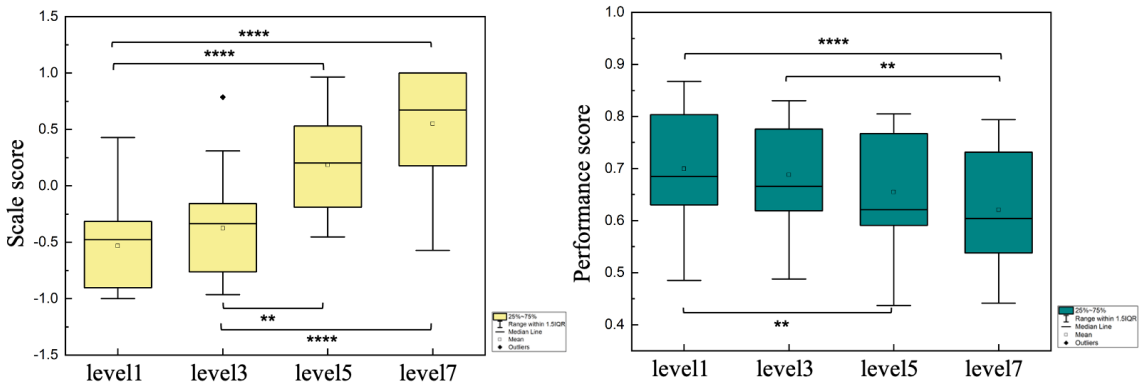

Fig 1. Subjective scale result and performance score result. The normalized subjective scale scores for each task difficulty are shown on the left, and the normalized performance scores for each difficulty are shown on the right. Note that the ** means $p < 0.05$, **** means $p<0.01$.

Shown in Figure 1, The influence of the task's difficulty level on the comprehensive scale and comprehensive performance of 20 subjects was analyzed. Figure 1 indicates that subjective scale scores increase while the task performance decrease with the increase of task difficulty. After the outliers were removed, one-way ANOVA analysis showed that the overall score of the scale had a major effect on the mental workload level ($F (3, 76) = 31.633$, $P<0.001$). Post-hoc analysis showed significant differences between load levels 1 and 5 ($P <0.01$), 1 and 7 ($P <0.01$), 3 and 5 ($P <0.05$), and 3 and 7 ($P <0.01$) ($P <0.05$). One-way ANOVA analysis of performance scores showed that the main effect of combined performance scores was also observed on the mental workload level, $F (3, 76) = 17.16$, $P<0.001$. Post-hoc analysis showed that there were significant differences in the overall performance between difficulty 1 and 5 ($p=0.001$), 1 and 7 ($p<0.001$), and 3 and 7 ($p=0.001$) ($p<0.05$).

2.2 Features Change Analysis Results

2.2.1 EEG Channels Select Result

In order to obtain the optimal EEG channel configuration for measuring mental workload, the importance of each channel was calculated by the algorithm mentioned in 2.5.2 above, and the results were sorted from high to low, as shown in Fig. 3.

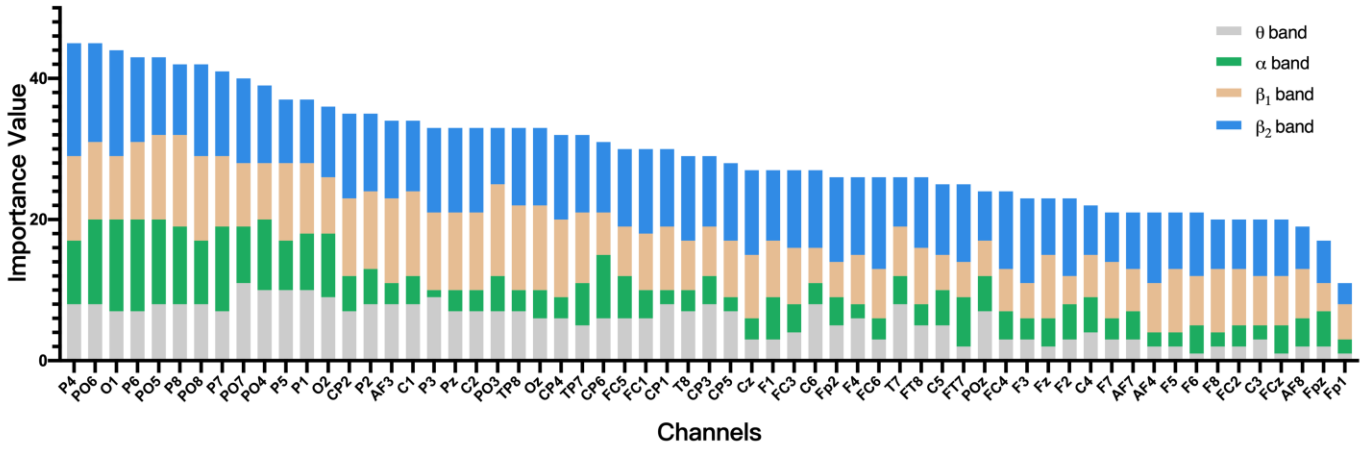

Fig 2. Channels importance ordering. The horizontal axis is channel, and the vertical axis is importance value of each channel, and the results were sorted from high to low. Different colors represent different frequency bands of EEG, and the height of the color block represents the features importance for a given channel.

It can be seen in Fig. 2 that almost all channels have good performance in β_1 and β_2, which indicates that power of β_1 and β_2 are sensitive physiological characteristics of mental workload. For θ and α bands, only channel PO5 and P4 were observed better performance. What's more, Fig. 3 reveals that the sensitive channels on mental workload of EEG are mainly comprised of occipital lobe (P4, PO6, O1, P6, PO5, P8, and OZ), the frontal (AF3) and sports area (C1, C2).

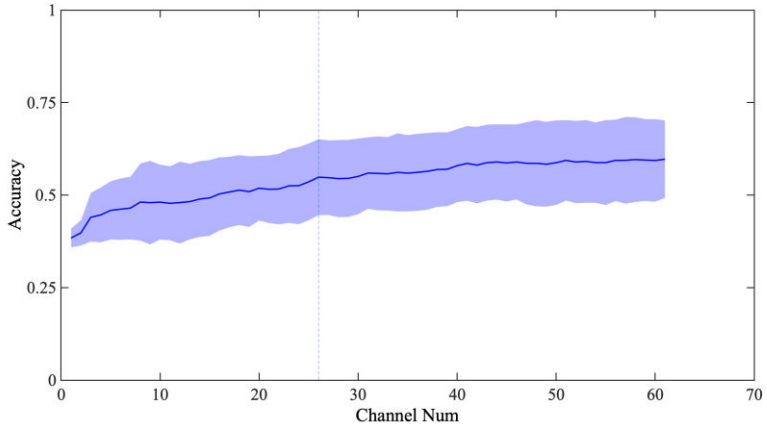

Fig 3. The accuracy varies with the number of channels. The figure shows the change curve of accuracy based on SVM under different number of channels. The horizontal axis is the number of channels and the vertical axis is the accuracy. The selection of channels is based on the sorting shown in Figure 2.

Furthermore, classification model based on SVM classifier was performed to explore the influence of channel number on classification accuracy. The channel was added to the model one by one based on the ranking order in Fig.3, and the modelling performance varying with the number of channels are shown in Fig. 3. It can be seen that the classification accuracy of the model increases with the increase of the number of channels, but the rising trend slows down when the number of channels reaches 26.

Therefore, we selected 26 channels (P4, PO6, O1, P6, PO5, P8, PO8, P7, PO7, PO4, P5, P1, O2, CP2, P2, AF3, C1, P3, Pz, C2, PO3, TP8, Oz, CP4, TP7, CP6) for subsequent research.

2.2.2 EEG Features Change Analysis Results

ANOVA was performed to analysis the difference of power response in θ, α, β_1 and β_2 band under various mental workload levels. The statistical analysis results of 62 scalp electrodes was demonstrated as scalp maps in Fig4, and the result was corrected by FDR, since hundreds of comparisons were implemented simultaneously. After removing the outliers, ANOVA results showed that the differences PSD response were mainly concentrated in all regions of θ band, prefrontal and occipital regions of α band, and occipital regions of β_1 and β_2 band.

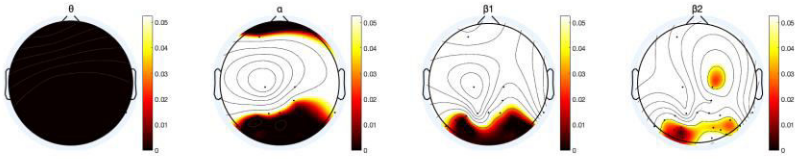

Fig 4. ANOVA of each channel comparing four difficulty level. The figure shows ANOVA results in four frequency bands. All data are processed by Z-score first and outliers are removed. the result was corrected by FDR. White represents p>0.05, deep brown indicates p<0.01, and light brown indicates 0.01<p<0.05.

Further, in order to find out the change trend of PSD with task difficulty, the difference of power in θ, α, β_1 and β_2 band between highest load (Level 7) and lowest load (Level 1) were calculated. The average results among 20 subjects of 62 scalp electrodes was demonstrated as scalp maps in Fig 6 (a). Fig 6(b) indicates that power of θ band in prefrontal increased with the increase of task load, while power of α band in the right hemisphere and occipital region decreased with the increase of task difficulty. In addition, the power of β_1 and β_2 bands in the occipital region increased with the increase of task difficulty.

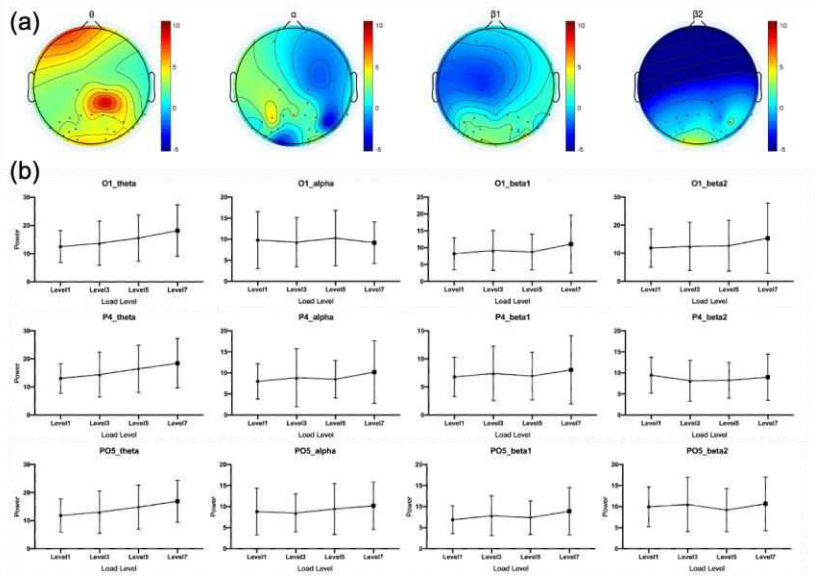

Fig 5. Power of different EEG band. Fig 5 (a) shows the difference of power in θ, α, β_1 and β_2 band between highest load (Level 7) and lowest load (Level 1), and the result of 62 electrodes was averaged among 20 subjects in the scale map. Fig 6 (b) shows the change trend of PSD in θ, α, β_1 and β_2 band of specific channel (O1, P4 and PO5) with task difficulty. The x-axis represents task difficulty level and the y-axis represents the power amplitude.

Specifically, the change trend of PSD in θ, α, β_1 and β_2 band of three typical channels (O1, P4 and PO5) were analyzed, as shown in Fig 5 (a). According to the data shown in Fig 5(b), three major conclusions can be obtained: Power of θ band in O1, P4 and PO5, showed a good correlation with task difficulty and increased with the increase of load level, which was consistent with the previous research [28]; Power of β-band in O1 and PO5 also increased with the increase of task load; The α-band energy of O1 channel is negatively correlated with the load, which is consistent with the study [12].

2.2.3 fNIRS Features Change Analysis Results

Next, the changes of O_2Hb and HHb during the change of mental workload were analyzed. The average of 20 subjects under the same load level obtained the results as shown in Figure 6, which shows the variation trend of O_2Hb and HHb amplitude in the left and right prefrontal regions with a time window of 3s. After the outliers were removed, one-way ANOVA analysis showed that the main effect of O_2Hb on the left and right side was observed at the load level, $F(3, 596) = 24339.950$, $P<0.001$, $F(3, 596) = 5499.275$, $P<0.001$. The main effect of HHb on the left and right side was also observed on the load level, $F(3, 596) = 4455.428$, $P<0.001$, $F(3, 596) = 2370.904$, $P<0.001$. Post-hoc results showed significant differences among all grades ($P<0.01$).

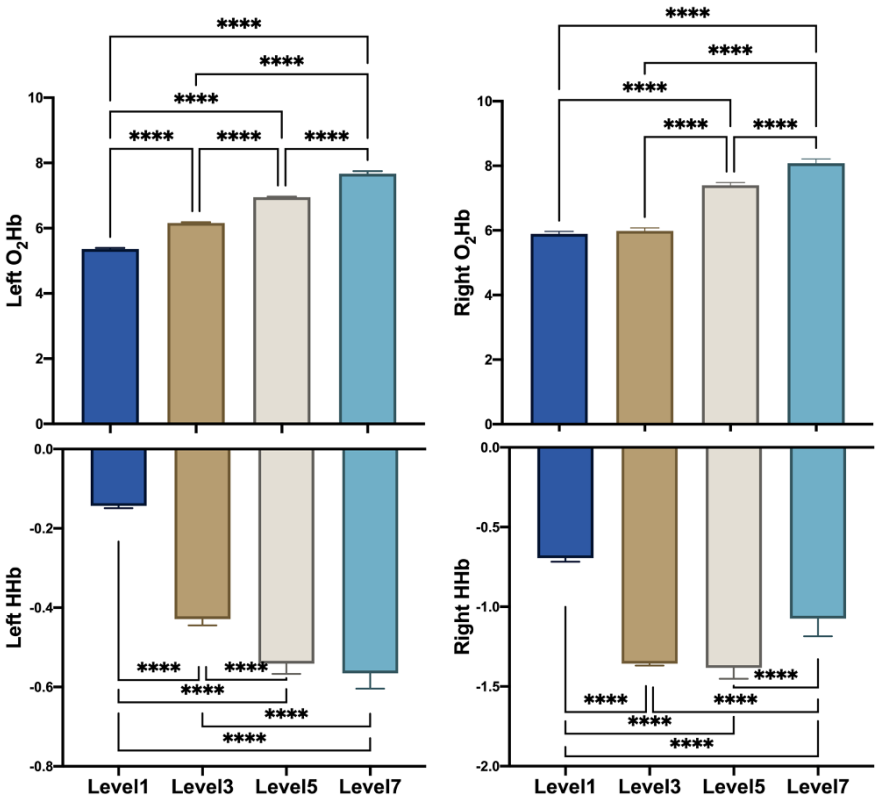

Fig 6. Changes of blood oxygen under different load levels. The left and right images represent the left and right sides of the forehead, with O_2HB changes on the upper side and HHB changes on the lower side. Note that the ** means $p < 0.05$, **** means $p<0.01$.

After the brain enters the working state, local neuronal activations increase metabolic rate, leading to increased blood flow and volume [29]. At the beginning of the task, PFC oxygenation locally increases, and the higher the load of the task, the more active the brain becomes, which also leads to the increase of O_2Hb in the PFC region [30]. As can be seen from the pattern shown in the figure, during the period of Level 1 to Level 7 tasks, the concentration of O_2Hb increases with the difficulty of tasks, and the content of O_2Hb at Level 7 is significantly higher than that of other tasks with load levels. We also note that the change of O_2Hb is more dramatic under high workload levels. Especially during a Level 7 task, the variation range and average amplitude of O_2Hb are significantly improved compared with Level 1. The change of HHb and O_2Hb showed an opposite trend. The content of HHb decreased with the difficulty of the task, and the change range of HHb was significantly lower than that of O_2Hb.

2.3 Classification Results

In order to study the classification effect of EEG, fNIRS, and EEG-fNIRS feature sets, three classifiers, namely, SVM, RF, and DT, were performed to establish the recognition model. Table 1 presents the minimum, maximum, average and mean square deviation of classification accuracies of different feature sets among 20 subjects. For Only-EEG feature set, the mean of classification accuracies of SVM, RF and DT are 52.43%, 55.47% and 46.05% respectively, with a highest recognition accuracy 55.47% in RF classifiers. For Only-fNIRS feature set, the mean of classification accuracies of SVM, RF, and DT are 67.71%, 69.20% and 62.70% respectively, with a highest recognition accuracy 69.20% in RF classifiers. For EEG-fNIRS feature set, the mean of classification accuracies of SVM, RF, and DT are 73.85%, 78.26% and 67.38% respectively, with a highest recognition accuracy 78.26% in RF classifiers.

Table 1. Classification results of three feature sets

Feature Set	Classifier	Results			
		Min	Mean	Max	Std.
Only-EEG	SVM	29.19	52.433	79.81	11.75
	RF	32.54	**55.473**	77.40	11.99
	DT	25.93	46.051	72.12	12.22
Only-fNIRS	SVM	54.21	67.712	83.33	8.20
	RF	52.88	**69.204**	80.88	7.87
	DT	50.48	62.703	76.92	7.15
EEG-fNIRS	SVM	62.02	73.853	85.71	6.14
	RF	70.67	**78.255**	88.21	4.71
	DT	54.81	67.384	81.00	7.79

After removing outliers, a two-factor method was used to analyze the effects of various classifiers and datasets on the accuracy. The results showed that the main effect of accuracy was observed on the feature set and the classifier, $F(2,171) = 92.539$, $P < 0.01$,

F (2,171) = 15.253, P<0.001. No interaction effect F (4, 171) = 0.317, P>0.05 was observed. To sum up, RF classifier performed better in three feature sets compared with SVM and DT, and EEG-fNIRS feature set provided better performance than both only-EEG feature set and only-fNIRS feature set, with a highest four-level recognition accuracy 78.26±4.71%.

What's more, according to table 1, we also observed that the standard deviations of recognition accuracy in EEG-fNIRS feature set was smaller than that of both only-EEG feature set and only-fNIRS feature set. In conclusion, EEG-fNIRS feature sets not only significantly improve the classification accuracy, but also make the model more stable and more robust, which is particularly important in practical applications.

3. Discussion

In order to promote the practical application of the mental workload status detection technology, this study conducted MATB to simulate the cognitive needs of operators in their daily work and used portable EEG acquisition equipment and fNIRS acquisition equipment to collect the physiological signals of subjects during the task. In this study, First, EEG signals from 64 channels were simplified to 26 channels, which significantly improved the convenience of operating the equipment. Second, variation of both EEG and fNIRS features with the task difficulty were analyzed, in order to provide physiological explanation for the variation of mental workload. Finally, the modeling performance of Only-EEG feature sets, Only-fNIRS feature sets, and EEG-fNIRS feature sets in four levels of mental workload monitoring was explored and compared.

In the analysis of operator behavioral data, we observed that job performance significantly decreased and subjective scale score increase with the increase of mental workload. This indicates that excessive mental workload will lead to insufficient cognitive resources for operators to maintain good job performance. We selected 26 channels with the highest degree of correlation with load for analysis from the results of channel screening. It can be seen that these 26 channels are mainly concentrated in the occipital lobe and parietal lobe, while a few channels are located in the frontal region. The occipital lobe region is generally considered as a visual region [31], which is highly correlated with basic cognitive functions such as visual search and visual attention [32]. The performance of the operation is highly correlated with the parietal lobe [33]. The prefrontal region is involved in various higher cognitive abilities, such as executive function and memory [34]. This indicates that the mental workload of operators during the task is mainly caused by vision and operation, especially as the operator needs to take into account all subsystems, so it evokes a significant response in the visual region.

Further analysis of the changes in features found that the θ band power in the frontal region increased significantly with the increase of task load, and all brain regions showed significant differences. It can be considered that θ band power is a sensitive physiological index of mental load. When the mental load was increased, α band power in the occipital lobe and other hemispheric regions decreased, which was consistent with the results in reported studies [11, 35, 36]. In general, the power of β_1 and β_2 bands increased obviously with the increase of the load in the occipital region. However, by analyzing the power spectrum energy

changes in specific channels, it was found that the power of β_2 band in O1 channel increased with the increase of task difficulty, while the power of β_2 band in P4 channel showed a downward trend. This shows that channel location has a substantial effect on features.

From the perspective of modeling performance, three different machine learning models were used to analyze the classification results under different feature sets. The results point out that the multimodal dataset is better than the unimodal dataset, which validates the results of [37-39]; moreover, the detection method established in this study is better in terms of classification accuracy than studies such as [11, 40] in the multimodal study. The comparison of the number of categories from load classification is higher than [18, 26, 41] and the robustness of the model is better. The reasons for the better results in this study are as follows: 1. the task selection, which is more relevant to the actual scenario, and the task difficulty setting is more reasonable, which effectively induces different levels of brain load; 2. the channel screening was conducted before modeling, and some redundant EEG channels were eliminated before fusion with fNIRS features. This feature combination overcomes the defects of low EEG spatial resolution and low fNIRS temporal resolution on the one hand, and may generate other key interaction information to help improve the accuracy of EEG load recognition on the other hand. The feature data can be further mined later and relevant experiments can be designed to verify the present conjecture.

Finally, it should be noted that there are still some limitations in our study. For one thing, data acquisition configuration of EEG contains 26 electrodes, which means there is still an optimized space. For another, only an offline experiment was implemented in this study, and the experimental task is not a real application task scenario. For further exploration, we hope to reduce the number of EEG channels further and improve our measurement method by monitoring the mental workload status of pilots in real-time during flight missions.

4. Conclusions

In conclusion, this study was to constructed a more accurate and convenient EEG–fNIRS-based mental workload detection method by optimizing the signal acquisition configuration. The result suggested 26 EEG channels and two frontal fNIRS channels is enough for a four-level mental workload detection accuracy of 78.25±4.71%, which is higher than previously reported studies. The results of this study can promote the application of mental workload detection technology in military, driving, and other complex human–computer interaction systems.

5. Method

5.1 Participants

Twenty volunteers participated in our experiment, all from China Astronaut Research and Training Center, with an average age of 25.6±2.24, normal or corrected to normal vision, without any neurological disease or history of neurological disease, and in a

stable mental state before the experiment. In order to avoid the influence of gender differences and hand dominance on the results, all the subjects were male and right-handed. After familiarizing themselves with all the experimental procedures and requirements, the subjects signed the informed consent form.

During the experiment, the subjects were asked to sit in front of a 23-inch LCD monitor with their eyes about 70–90cm away from the screen. The volunteers were asked to carefully read the instructions given on the screen before each sub-task and complete the experimental tasks as required. They were also asked to focus on the process of the experiment task execution as far as possible to get the best results, to ensure that the experimental data were real and effective, and to cooperate with data collection work.

5.2 Experimental Design

This experiment uses the MATB task to trigger the mental workload of subjects. This task is composed of three sub-tasks, namely, system monitoring, tracking, and resource management. It is a complex task in-volving three cognitive resources: attention, operation, and reasoning. The cognitive resources involved in the actual task are the same as those involved in the fundamental tasks performed by the operator, which can well simulate the real-world task environment of the subjects.

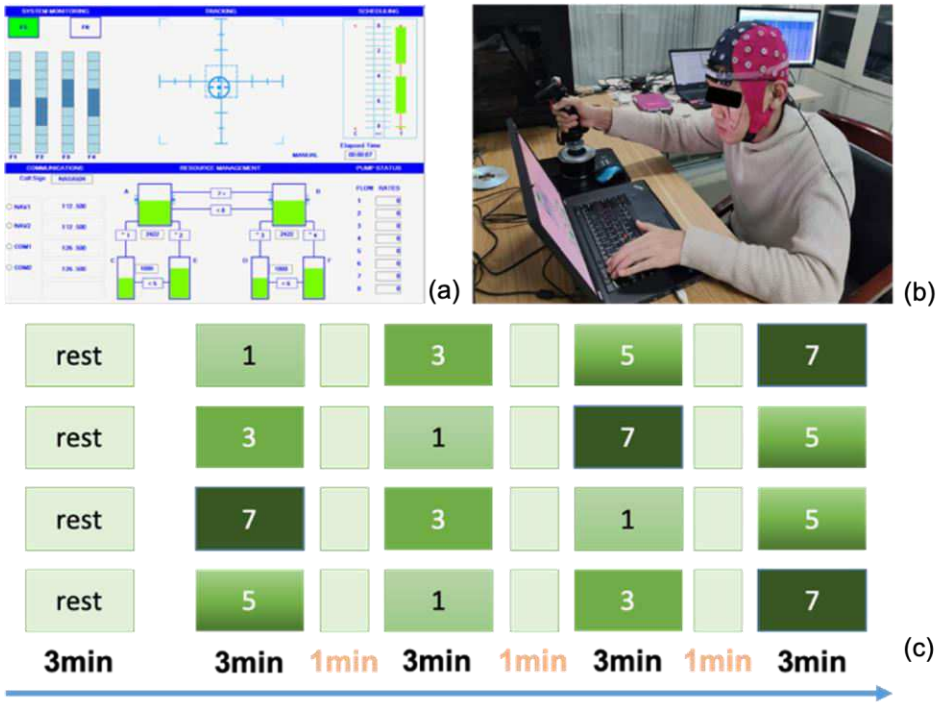

Fig 7. Stimuli and experimental procedure. Fig 7 (a) show the stimuli of the experiment paradigms. Fig 7 (b) shows a real scenario of the experiment. Fig 7 (c) demonstrate the experimental procedure. Note that the four different random stimuli sequence consisted of blocks repeated 4 times.

During the task, three sub-tasks appear on the screen simultaneously. As shown in Figure 7 (a), the upper left is the system monitoring task. When an abnormal state occurs, the subject will press the “F1–F6” button to respond, and the subject is required to press the corresponding button in the shortest possible time. On the upper right is the tracking task, in which subjects track a circular target by controlling a joystick and are asked to aim the crosshairs at the target as accurately as possible. The lower part is

the oil management task. The subjects control the opening and closing of the oil circuit by pressing "1–8", and they are required to ensure the oil volume of tank A/B remains within a specific range.

Experimental procedure shown in Figure 7 (c). The subjects completed a total of four blocks, each of which was randomly assigned four difficulty levels. Each block takes three minutes, and the four three-minute tasks are rated on a scale of difficulty from low to high: 1, 3, 5, and 7. There was a 1-minute rest period between two tasks and a 3-minute rest period before the task began. The actual experiment scene is shown in the Figure 7 (b). At the end of each task, the subjects were required to fill out the NASA-TLX scale. The four sequential tasks were performed in a random order among all subjects.

5.3 Data acquisition

The EGGO device of Ant Neuro was used to record the EEG signals of the subjects at 64 electrode positions as per international standards 10–20. The reference was set at CPZ and the eye movement signals were recorded from below the left eye. The impedance of all electrodes is kept below 5 KΩ and the sampling rate is 500 Hz.

The PORTALITE device of Artinis was used to record the changes in blood oxygen in the brain on both sides of the forehead. The sampling rate is set to 50 Hz.

5.4 Preprocessing and Feature extracting

After the collected EEG data were obtained, the average values of the two mastoid channels (M1, M2) were used to re-reference the signals. Then, a 0.5–45-Hz Butterworth band-pass filter and a 50-Hz notch filter were conducted to remove the interference of the DC component, high-frequency component, and power frequency. The sampling rate was reduced to 200 Hz to reduce the data memory size. Independent Component Analysis (ICA) was adopted to remove eye movement, motion artifacts, channel noise, and other interference. Finally, the data were segmented every 3 s with the beginning of stimulus as the mark. The above processing uses the MATLAB open-source toolkit EEGLAB.

For each epoch, Welch's method was used to extract the Power Spectral Density (PSD), and the power of θ (3–8 Hz), α (8–13 Hz), β_1 (13–20 Hz), and β_2 (20–30 Hz) were obtained. Then, seven-channel pairs (P8–P7, O2–O1, C2–C1, P4–P3, PZ–O1, PZ–O2, and O1–AF3) were selected from the left and right brain as well as the front and rear brain, and the energy difference of each channel pair in the θ, α, β_1, and β_2 bands was calculated as the new feature. A total of 276 (62 channels × 4 bands + 7 channels × 4 bands) EEG features were extracted.

The collected fNIRS signal contains considerable noise, including motion artifacts, physiological interference, and instrument noise. Preprocessing was performed using the Homer2 open-source toolkit to remove the motion artifacts. A 0.5-Hz low-pass filter was also applied to reduce instrumental and physiological noise. The processing steps are consistent with Foy et al. [42, 43]. Finally, the data were segmented every 3 s, also marked by the time the stimulus started.

For each epoch, 11 statistical features were extracted, i.e., mean value, standard deviation, mean square error, skewness, root mean square, peak value, peak factor, kurtosis, waveform factor, pulse factor, and margin factor. A neurovascular coupling feature of the frontal EEG and frontal oxygen signals was added to calculate the zero-lagged correlation between the amplitude of HHb or O_2Hb and the EEG frequency band power (in the four independent bands above). These HHb- or O_2Hb-based NVO features are represented as oxidative neurovascular coupling and deoxy-neurovascular coupling, respectively. A total of 48 (2 channels × 2 × 11 + 2 channels × 2) fNIRS features were extracted.

5.5 Data Analysis

5.5.1 Behavioral Data Analysis Methods

Behavioral data collected in the course of this experiment mainly includes the subjective scale score data filled out by the subjects and the task performance data generated during the task execution of the experimental platform. The subjects were asked to use the NASA-TLX scale to describe their subjective feelings of task load from six aspects: mental demand, physical demand, time demand, subjective evaluation of mission performance, effort, and frustration. To facilitate the analysis, the six dimensions of the subjective scale were normalized based on the maximum and minimum values and then averaged to obtain a comprehensive scale score. The MATB task used in this experiment consists of three sub-tasks: monitoring task, tracking task, and oil management task. Therefore, there are six performance indicators: the response time and response accuracy of the monitoring task; X- and Y-axis deviation of the tracking task; and fuel deviations of tank A and tank B for the oil management task. The performance data is normalized based on the maximum and minimum values and then averaged to obtain the comprehensive performance.

In order to analyze the influence of load level on tasks, we calculated Bonferroni corrected ANOVA for overall performance as well as for comprehensive scale scoring, taking mental load level as a factor, and conducted post-hoc comparison among various load levels. False Discovery Rate (FDR) correction was performed for multiple comparisons. The significance value was set at $P < 0.05$. By analyzing task behavior data, we can preliminarily summarize the law of the influence of task difficulty on mental load.

5.5.2 EEG Channels Selection Methods

Given that not all areas of the human brain are directly related to mental workload, the 64-channel whole-brain EEG is unnecessary, and the fewer the channels, the easier it is to use in practice. The goal of this study was to use as few EEG channels as possible without losing accuracy. In order to find the brain regions that have high correlation with mental workload, the Recursive Feature Elimination (RFE) algorithm based on SVM was used to filter the channels. RFE is a feature selection method with good performance and strong generalization ability[44]. The main idea is to select the best features by repeatedly building models (such as SVM), eliminating the selected features, and then repeating the above process on the remaining features until all the features are trawled. Considered to be physiological features closely related to mental load in reported studies [12, 28], PSD features of four frequency band including θ (3–8 Hz), α (8–13 Hz), β_1 (13–20 Hz), and β_2 (20–30 Hz) of 64-channel EEG signals were calculated

for screening. The top 100 features with the largest RFE results were selected of each subject for further analysis. Feature importance weight was defined as the number of people with this feature in the above-mentioned top 100 feature sets among 20 volunteers. Channel importance weight was define as the sum of feature weights of 4 frequency bands of the this channel.

5.5.3 Classification Methods

In this study, three classification models, namely, Support Vector Machine (SVM), Decision Tree (DT), and Random Forest (RF), were adopted to model the extracted features. The grid search method was selected for different classifiers to obtain the optimal model parameters. Radial Basis Function (RBF) was applied as the kernel Function in SVM, with two important parameters of C (punish coefficient) and error tolerance, with the search space as [0.0001, 0.001, 0.01, 0.1, 1, 10, 20, 30]. Parameter Gamma represent the number of support vector, with a search space of [0.1, 0.2, 0.25, 0.4, 0.8, 1.6, 3.2, 6.4]. The DT algorithm has three parameters that need to be adjusted: the partition standard, the maximum depth, and the minimum sample number required to segment the internal nodes. 'Entropy' and 'Gini', are tried as the search space of the parameter 'partition standard. The search space of parameter 'maximum depth' and parameter 'Minimum sample number required to segment internal nodes is [10, 30, 60, 100] and [2, 5, 10, 15] respectively. RF is an ensemble classification model with good generalization. Reported studies[27, 45] have shown that random forest performs well in mental workload classification. The classifier mainly searches for the parameter "the number of Random Forest spanning trees", with a adjusting space [100, 200, 500]. In addition, in order to ensure the reliability of the classification results, all the classification algorithms are validated by five-fold cross-validation.

Abbreviations

EEG: electroencephalogram; ICA: independent component analysis; SVM: support vector machine; MATB : Multi-Task Attribute Battery; PSD: Power Spectral Density; PLI: Phase Lag Index; ERP: Event-related Potential; DT : Decision Tree; RF : Random Forest; FDR: False Discovery Rate.

Declarations

Ethics approval and consent to participate

Our research was approved by Ethnics committee of China Astronaut Research and Training Center

Consent for publication

Not applicable.

Data availability statement

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported in part by National Natural Science Foundation of China (NO. 81671861), Experimental Technology Foundation of National Key Laboratory of Human Factors Engineering (NO. SYFD061905，YYJJ190604).

Authors' contributions

Hongzuo Chu and Yong Cao carried out the design of the experiment. Jiehong Yang and Mengyin Huang implemented the experiment. Hongzuo Chu and Jin Jiang accomplished the data processing. Hongzuo Chu and Yong Cao wrote the manuscript. Qijie Li, Changhua Jiang, and Xuejun Jiao checked and modified the paper. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

Author details

1. National Key Laboratory of Human Factors Engineering, China Astronaut Research and Training Center, Beijing, China.

2. Space Engineering University, Beijing, China.

¶These authors contributed equally to this work.

*Corresponding author. E-mail: jxjisme@sina.com

References

[1] M. Kuwato and Y. Hirano, "Sense of coherence, occupational stressors, and mental health among Japanese high school teachers in Nagasaki prefecture: a multiple regression analysis," *BMC Public Health,* vol. 20, no. 1, p. 1355, 2020.

[2] R. N. Roy, S. Charbonnier, A. Campagne, and S. Bonnet, "Efficient mental workload estimation using task-independent EEG features," *Journal of neural engineering,* vol. 13, no. 2, p. 026019, 2016.

[3] E. M. Rantanen and J. H. Goldberg, "The effect of mental workload on the visual field size and shape," *Ergonomics,* vol. 42, no. 6, pp. 816-834, 1999.

[4] W. Chappelle, K. Mcdonald, and K. Mcmillan, "Important and Critical Psychological Attributes of USAF MQ-1 Predator and MQ-9 Reaper Pilots According to Subject Matter Experts," 2011.

[5] H. Mansikka, K. Virtanen, and D. Harris, "Dissociation Between Mental Workload, Performance, and Task Awareness in Pilots of High Performance Aircraft," *IEEE Transactions on Human-Machine Systems,* vol. PP, no. 1, pp. 1-9, 2018.

[6] B. El-Khodary and M. Samara, "The effect of exposure to war-traumatic events, stressful life events, and other variables on mental health of Palestinian children and adolescents in the 2012 Gaza War," *Lancet,* vol. 391, p. S6, 2018.

[7] V. G. Iannacchione *et al.*, "Validation of a Research Case Definition of Gulf War Illness in the 1991 US Military Population," *Neuroepidemiology,* vol. 37, no. 2, pp. 129-140, 2011.

[8] Chappelle *et al.*, "Assessment of Occupational Burnout in United States Air Force Predator/Reaper "Drone" Operators," *Military Psychology (American Psychological Association),* 2014.

[9] R. Li and Z. Liu, "Stress detection using deep neural networks," *BMC Medical Informatics and Decision Making,* vol. 20, no. Suppl 11, 2020.

[10] G. N. Dimitrakopoulos *et al.*, "Task-independent mental workload classification based upon common multiband EEG cortical connectivity," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 25, no. 11, pp. 1940-1949, 2017.

[11] Y. Liu, H. Ayaz, and P. A. Shewokis, "Multisubject "learning" for mental workload classification using concurrent EEG, fNIRS, and physiological measures," *Frontiers in human neuroscience,* vol. 11, p. 389, 2017.

[12] V. Pergher, B. Wittevrongel, J. Tournoy, B. Schoenmakers, and M. M. Van Hulle, "Mental workload of young and older adults gauged with ERPs and spectral power during N-Back task performance," *Biological psychology,* vol. 146, p. 107726, 2019.

[13] M. S. Caywood, D. M. Roberts, J. B. Colombe, H. S. Greenwald, and M. Z. Weiland, "Gaussian process regression for

predictive but interpretable machine learning models: An example of predicting mental workload across tasks," *Frontiers in human neuroscience,* vol. 10, p. 647, 2017.

[14] H. Qu *et al.*, "Mental Workload Classification Method Based on EEG Independent Component Features," *Applied Sciences,* vol. 10, no. 9, p. 3036, 2020.

[15] Y. Santiagoespada, R. R. Myer, K. A. Latorella, and J. Comstock, "The Multi-Attribute Task Battery II (MATB-II) Software for Human Performance and Workload Research: A User's Guide," *Santiago-Espada, Yamira,* 2011.

[16] I. Kakkos *et al.*, "Mental Workload Drives Different Reorganizations of Functional Cortical Connectivity Between 2D and 3D Simulated Flight Experiments," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 27, no. 9, pp. 1704-1713, 2019.

[17] S. H. Kohl, D. Mehler, M. Lührs, R. T. Thibault, and B. Sorger, "The Potential of Functional Near-Infrared Spectroscopy-Based Neurofeedback-A Systematic Review and Recommendations for Best Practice," *Frontiers in Neuroscience,* vol. 14, p. 594, 2020.

[18] U. Asgher *et al.*, "Enhanced accuracy for multiclass mental workload detection using long short-term memory for brain–computer interface," *Frontiers in neuroscience,* vol. 14, p. 584, 2020.

[19] M. R. Siddiquee, R. Atri, J. S. Marquez, S. Hasan, and O. Bai, "Sensor Location Optimization of Wireless Wearable fNIRS System for Cognitive Workload Monitoring Using a Data-Driven Approach for Improved Wearability," *Sensors,* vol. 20, no. 18, 2020.

[20] M. R. Islam *et al.*, "A Novel Mutual Information Based Feature Set for Drivers' Mental Workload Evaluation Using Machine Learning," *Brain Sciences,* vol. 10, no. 8, p. 551, 2020.

[21] T. Gateau, G. Durantin, F. Lancelot, S. Scannella, and F. Dehais, "Real-Time State Estimation in a Flight Simulator Using fNIRS," *Plos One,* vol. 10, no. 3, p. e0121279, 2015.

[22] Y. Liu, H. Ayaz, and P. A. Shewokis, "Mental workload classification with concurrent electroencephalography and functional near-infrared spectroscopy," *Brain-Computer Interfaces,* vol. 4, no. 3, pp. 175-185, 2017.

[23] A. Sangtae, N. Thien, J. Hyojung, J. G. Kim, and S. C. Jun, "Exploring Neuro-Physiological Correlates of Drivers' Mental Fatigue Caused by Sleep Deprivation Using Simultaneous EEG, ECG, and fNIRS Data," *Frontiers in Human Neuroscience,* vol. 10, no. 848, 2016.

[24] P. A. Cicalese *et al.*, "An EEG-fNIRS hybridization technique in the four-class classification of alzheimer's disease," *Journal of neuroscience methods,* vol. 336, p. 108618, 2020.

[25] A. De, "Prefrontal haemodynamics based classification of inter-individual working memory difference," *Electronics Letters,* 2020.

[26] A. S. Le, H. Aoki, F. Murase, and K. Ishida, "A novel method for classifying driver mental workload under naturalistic conditions with information from near-infrared spectroscopy," *Frontiers in human neuroscience,* vol. 12, p. 431, 2018.

[27] Z. Pei, H. Wang, A. Bezerianos, and J. Li, "EEG-Based Multiclass Workload Identification Using Feature Fusion and Selection," *IEEE Transactions on Instrumentation and Measurement,* vol. 70, pp. 1-8, 2020.

[28] C. Diaz-Piedra, M. V. Sebastián, and L. L. Di Stasi, "EEG theta power activity reflects workload among army combat drivers: an experimental study," *Brain sciences,* vol. 10, no. 4, p. 199, 2020.

[29] Y. Tong, K. P. Lindsey, L. M. Hocke, G. Vitaliano, D. Mintzopoulos, and B. Frederick, "Perfusion information extracted from resting state functional magnetic resonance imaging," *Journal of Cerebral Blood Flow & Metabolism Official Journal of the International Society of Cerebral Blood Flow & Metabolism,* p. 564, 2016.

[30] T. J. Huppert, R. D. Hoge, S. G. Diamond, M. A. Franceschini, and D. A. Boas, "A temporal comparison of BOLD, ASL, and NIRS hemodynamic responses to motor stimuli in adult humans," *Neuroimage,* vol. 29, no. 2, pp. 368-382, 2006.

[31] D. Liu, S. Duan, P. Wei, L. Chen, and J. Zhang, "Aberrant Brain Spontaneous Activity and Synchronization in Type 2 Diabetes Mellitus Patients: A Resting-State Functional MRI Study," *Frontiers in Aging Neuroscience,* vol. 12, p. 181, 2020.

[32] L. Cui *et al.*, "Tai Chi Chuan vs General Aerobic Exercise in Brain Plasticity: A Multimodal MRI Study," *Scientific Reports,* 2019.

[33] Ines *et al.*, "Externally induced frontoparietal synchronization modulates network dynamics and enhances working memory performance," *Elife,* 2017.

[34] Xiao-Feng, Jiao, Liu, Qi-Feng, Hua, and Yi-Jiao, "Relapsing-Remitting Multiple Sclerosis Is Associated With Regional Brain Activity Deficits in Motor- and Cognitive-Related Brain Areas," *Frontiers in neurology,* vol. 10, pp. 1136-1136, 2019.

[35] I. Amihai and M. Kozhevnikov, "Arousal vs. Relaxation: A Comparison of the Neurophysiological and Cognitive Correlates of Vajrayana and Theravada Meditative Practices," *PLoS ONE,* vol. 9, no. 7, p. e102990, 2014.

[36] R. Rojas and E. D. J. F. M. B. H. Abbass, "Electroencephalographic Workload Indicators During Teleoperation of an Unmanned Aerial Vehicle Shepherding a Swarm of Unmanned Ground Vehicles in Contested Environments," *Frontiers in Neuroscience,* 2020.

[37] P. Sirpal, A. Kassab, P. Pouliot, K. N. Dang, and F. Lesage, "fNIRS improves seizure detection in multimodal EEG-fNIRS

recordings," *Journal of Biomedical Optics,* vol. 24, no. 5, p. 1, 2019.
[38] W. Ellen, T. Jessica, C. S. Nam, and J. R. Franz, "Neuroimaging of Human Balance Control: A Systematic Review," *Frontiers in Human Neuroscience,* vol. 11, pp. 170-, 2017.
[39] M. U. Khan and M. Hasan, "Hybrid EEG-fNIRS BCI Fusion Using Multi-Resolution Singular Value Decomposition (MSVD)," *Frontiers in Human Neuroscience,* vol. 14, no. 599802, 2020.
[40] G. F. Wilson and C. A. Russell, "Operator Functional State Classification Using Multiple Psychophysiological Features in an Air Traffic Control Task," *Human Factors,* vol. 45, no. 3, pp. 381-389, 2003.
[41] H. Aghajani, M. Garbey, and A. Omurtag, "Measuring mental workload with EEG+ fNIRS," *Frontiers in human neuroscience,* vol. 11, p. 359, 2017.
[42] H. J. Foy and P. Chapman, "Mental workload is reflected in driver behaviour, physiology, eye movements and prefrontal cortex activation," *Applied ergonomics,* vol. 73, pp. 90-99, 2018.
[43] H. J. Foy, R. Patrick, C. Peter, and S. Manabu, "Prefrontal Cortex Activation and Young Driver Behaviour: A fNIRS Study," *Plos One,* vol. 11, no. 5, p. e0156512, 2016.
[44] C. Qi, Z. Meng, X. Liu, Q. Jin, and R. Su, "Decision Variants for the Automatic Determination of Optimal Feature Subset in RF-RFE," *Genes,* vol. 9, no. 6, p. 301, 2018.
[45] D. Novak, M. Mihelj, and M. Munih, "A survey of methods for data fusion and system adaptation using autonomic nervous system responses in physiological computing," *Interacting with Computers,* vol. 24, no. 3, pp. 154–172, 2012.