



Chinese Society of Aeronautics and Astronautics
& Beihang University

Chinese Journal of Aeronautics

cja@buaa.edu.cn
www.sciencedirect.com



Real-time evaluation method of flight mission load based on sensitivity analysis of physiological factors

Jun CHEN^{a,*}, Lei XUE^a, Jia RONG^b, Xudong GAO^a

^a School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710072, China

^b Department of Data Science and AI, Monash University, Clayton, VIC 3800, Australia

Received 31 May 2021; revised 22 June 2021; accepted 8 October 2021

Available online 23 November 2021

KEYWORDS

Eye movement;
Mission load;
Physiological data;
Sensitivity analysis;
Support Vector Machine (SVM)

Abstract As the complexity of flight missions continues to increase, sending a timely warning or providing assistance to pilots helps to reduce the probability of operational errors and flight accidents. Monitoring pilots' physiological data, real-time evaluation of mission load is a feasible technical way to achieve this. In this paper, a set of flight tasks including aircraft control, human-computer interaction and mental arithmetic tests are designed to simulate five mission loads at different flight difficulty levels. A sensitivity analysis method based on a comprehensive test is proposed to select a set of sensitive physiological factors. Then, based on the SVM hierarchical combination classification method, the pilot mission load real-time evaluation model is established. The test results show significant differences in EMG, respiration rate (abdomen), heart rate, blood oxygen saturation, pupil area, fixation duration, number of fixations, and saccades. The high accuracy obtained from experiments proved that the proposed real-time evaluation model is applicable to meet the requirements of real working environments. The findings can provide methodological references for mission load evaluation research in other fields.

© 2021 Chinese Society of Aeronautics and Astronautics. Production and hosting by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

New aviation technology has improved aircraft informatization and has also increased human-computer interaction in a complex confrontational combat environment. The flight mission load continues to grow, which might negatively impact

pilots' performance and even endangers aviation safety or cause flight accidents.¹⁻⁴ Therefore, we need to conduct real-time mission load evaluation for pilots and provide warnings or assistance to pilots with high mission loads in time. This work will help reduce the probability of operational errors and flight accidents and improve mission effectiveness.

In the field of Manned/Unmanned Aerial Vehicle collaboration, human task load level is affected by workload and cognitive load. When "overloaded" work for a long time, it will inevitably lead to a decline in situational awareness. In 2015, the U.S. Air Force Research Laboratory launched the Loyal Wingman Program to realize Manned/Unmanned Aerial Vehicle collaboration formation operations through real-time monitoring of the physiological state of the pilot.⁵ Therefore, it is

* Corresponding author.

E-mail address: junchen@nwpu.edu.cn (J. CHEN).

Peer review under responsibility of Editorial Committee of CJA.



Production and hosting by Elsevier

necessary to conduct real-time detection and monitoring of human task load status and integrate human status factors with the situational awareness of unmanned systems to achieve the effect of “human in the loop”.

Mission load refers to the amount of load that the human body bears per unit of time, including the intensity of the mission load and the impact of psychological factors and environmental conditions on the human body.⁶ At present, there are three main methods for evaluating mission load: subjective evaluation, mission performance, and physiological measurement.^{7–10} Compared with the subjective evaluation method and the mission performance method, the physiological measurement method has objectivity, real-time and continuity characteristics, and is more suitable for real-time evaluation of mission load.¹¹ Mission load research is widely used in driving fields, such as car drivers,^{12–15} marine pilots,¹⁶ aircraft pilots,^{17–21} occupant²² and so on. In this work, we focus on the flight mission load only.

Flight mission load evaluation based on the physiological measurement method generally uses three types: eye movement data, electroencephalogram (EEG) data and other physiological data (hereafter referred to as physiological data). The EEG signal is not suitable for real-time evaluation because of poor stability, high data dimension, poor interpretability, and significant interference to pilots during collection. Still, it can be used for offline comparative analysis. The eye movement and physiological signals have prominent characteristics and good interpretability. The measurement device can be embedded in the pilot’s helmet and combat clothing, which is less intrusive, so it is very suitable for real-time assessment of flight missions.^{18,19}

With the application and promotion of physiological sensors, some researchers have used physiological instruments to collect physiological data such as heart rate, respiration rate, pulse and electromyography (EMG).^{12,13,20} Based on the above data, the driver’s workload is evaluated offline,¹¹ and the correlation between physiological data and task difficulty is analyzed.¹² Some researchers used driving simulators to analyze the relationship between eye movement data and mission difficulty changes.^{14,15,18,21} Studies have shown that with increased information on road signs, the driver’s attention to the area of interest and the number of scans is increasing.¹⁴ Data such as fixation rate, duration of focus, rate of the saccade, saccade duration, and pupil area are sensitive to workload changes.^{15,21} Therefore, physiological data such as heart rate, respiration rate, pulse, EMG, fixation data, saccade data and pupil area can reflect changes in the mission load state-certain extent.

The above research work has strongly promoted the development of mission load evaluation technology. However, due to the various physiological factors involved, the selection generally relies mainly on common medical sense and subjective experience and lacks scientific and objective standards. Therefore, it is necessary to determine the most relevant factor set according to the specific mission type to improve the accuracy of the mission load evaluation model.

On the other hand, among the classification methods based on physiological data, Support Vector Machines (SVM) and Neural Networks (NN) have received extensive attention.^{23–26} Currently, the neural network model is a black-box model

based on reasoning, and reasoning cannot be explained. Its theory and algorithms still need to improve and enhance further. The support vector machine model has attracted many scholars because of its transparent derivation process, avoiding falling into optimal local values and good performance.²⁷ However, the SVM is a two-class classifier, and the SVM multi-class classifier lacks practicability due to complex calculations and difficulty in implementation. To achieve real-time evaluation, a method with fast classification speed and high classification accuracy is required. At present, some researchers have achieved multi-class classification by using multiple SVM two-class classifiers, that is, based on SVM hierarchical combination classifiers.²⁸

Therefore, the main contributions of this paper falling to solve the above two problems from the literature review.

- (1) Comprehensive test method is proposed to select a limited number of highly significant physiological factors (problem 1)
- (2) SVM hierarchical combination classification model is designed to establish the flight mission load real-time evaluation (problem 2)

The rest of the paper is organized as follows: [Section 2](#) introduces the experimental design of the flight mission. [Section 3](#) is an introduction to the real-time evaluation method of flight mission load based on sensitivity analysis of physiological factors. [Section 4](#) analyzes and discussed the obtained experimental results. [Section 5](#) evaluates the proposed model effects. [Section 6](#) concludes the paper with limitations and future work.

2. Flight mission experimental design

This section is an introduction to the flight mission experimental design. It mainly includes the type of tasks, the mission difficulty, the experimental subject and the experimental equipment.

2.1. Flight tasks

Under normal circumstances, the flight tasks that pilots need to complete include control of the aircraft, interaction with the aircraft cockpit, and cognitive psychological activities. Among them, the control of the aircraft consists of making the aircraft fly smoothly and making the aircraft fly according to the mission plan or the designated waypoints. Interaction with the aircraft cockpit means that in a real battlefield confrontation environment, the pilot must always observe and understand possible enemy threats and eliminate enemy threats through certain operations or avoid enemy threats through certain maneuvers. Finally, cognitive psychological activities mean that when the pilots in our formation encounter a very urgent situation during the mission, they need to communicate and interact with other pilots. At the same time, the pilots themselves will use their own experience to judge the current battlefield situation and plan their following actions.

Based on the above description, three types of flight tasks are designed in this article:

- (1) Designated route tasks: Three different routes are designed in this article, corresponding to the mission planning routes in three different levels of complex environments, to simulate the pilots' workload caused by the real battlefield flight environment as much as possible.
- (2) Interference elimination tasks: The tasks of eliminating threats and completing maneuvers are designed in this article to simulate launching missiles to destroy enemy entities and to escape the enemy's attack range through maneuvers.
- (3) Mathematical calculation tasks: This article is designed to perform real-time mathematical calculations when the pilot needs to bear an extremely high load to simulate the communication and thinking between the pilot and other pilots in the formation during an emergency.

In the experiment designed above, designated route tasks and completing maneuvers tasks (one type of the interference elimination tasks) are to increase the intensity of the mission load. Eliminating threats tasks (another type of the interference elimination tasks) and mathematical calculation tasks are to increase the impact of psychological factors of the pilots.

The three types of flight tasks are described in detail below.

2.1.1. Designated route tasks

Three challenging flight routes are designed in the experiment: straight route, simple serpentine route, and complex serpentine route, as shown in Fig. 1. During the flight, the subject must start from the red plane marked in the picture, follow the route specified on the map, and try to ensure that the aircraft is always within the yellow line.

2.1.2. Interference elimination tasks

In the interference elimination task, the subject must complete the specified maneuvers or eliminate the threats that appear on the screen during the flight, as shown in Fig. 2.

The first type of interference elimination task is to complete the designated maneuver. Specifically, during the experiment, an instruction of roll to left or right will randomly appear on the screen. When the instruction appears, the subject must follow the instruction to complete the corresponding maneuver.

The second type of interference elimination task is to eliminate the threat. Three different threats are designed for the entire flight mission, corresponding to three different buttons on the flight control joystick. When a specific threat appears on the screen during the experiment, the subject needs to press the corresponding button to eliminate the threat.

During the experiment, the two interference elimination tasks will not appear at the same time.

2.1.3. Mathematical calculation tasks

In the mathematical calculation task, the subject must answer simple mathematical calculation questions raised by the examiner during the flight. Each group of experiments takes 60 s. For example, starting from the 5th second of the investigation, the examiner asks a simple two-digit addition and subtraction question every 10 s. Each experiment lasts 60 s, so the subject has to answer 5 questions in total, as shown in Fig. 3.

2.2. Mission difficult levels

Based on the above three types of flight tasks, five different battlefield situations can be simulated to obtain five different mission difficulties (as shown in Table 1):

Mission difficulty I: The pilot has just taken off from the base and has not yet arrived at the battlefield, so the pilot flies the plane to the battlefield straightly.

Mission Difficulty II: The pilot arrives at the battlefield by flying the plane, has not yet found the enemy and is on alert.

Mission difficulty III: The pilot flies the plane to the battlefield and is captured by the enemy's distant reconnaissance equipment. The pilot completes maneuvers to escape the reconnaissance.

Mission difficulty IV: The pilot flies the plane to the enemy's hinterland and encounters the enemy fighter plane. At this time, our side has a greater advantage. On the one hand, the pilot needs to complete maneuvers to defend against enemy threats, and on the other hand, he needs to wait for opportunities to attack the enemy.

Mission difficulty V: The pilot flies the plane to the enemy's hinterland, encounters and is surrounded by enemy fighters. On the one hand, the pilot needs to complete maneuvers to defend against enemy threats, and on the other hand, he needs to communicate with our other pilots, request support, and actively plan to take measures to combat enemy threats.

Different flight mission difficulties have different flight mission loads on the subjects. The mission load levels corresponding to mission difficulty I-V, in turn, include Very Low mission load (VL), Low mission load (L), Medium mission load (M), Higher mission load (H) and Very High mission load (VH).

2.3. Human participant in experiments

To avoid the influence caused by the level of operation skills, one subject and multiple measurements are adopted in this article. The selected subject is a male graduate student of the school, with good physical quality, normal hearing and vision, and performing simulated flight operations proficiently.

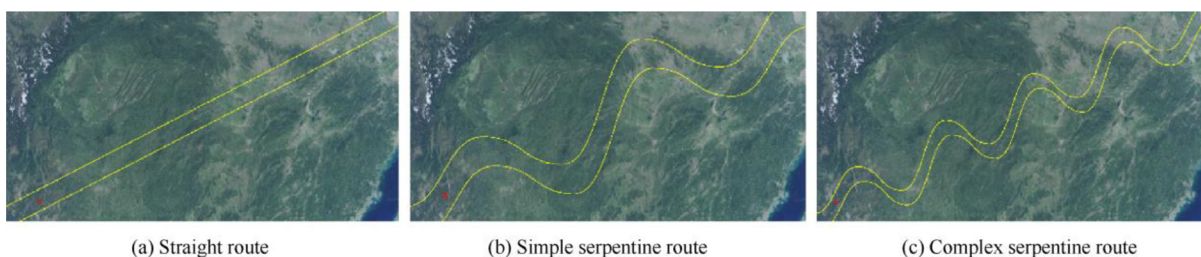


Fig. 1 Three flight routes of different difficulty.

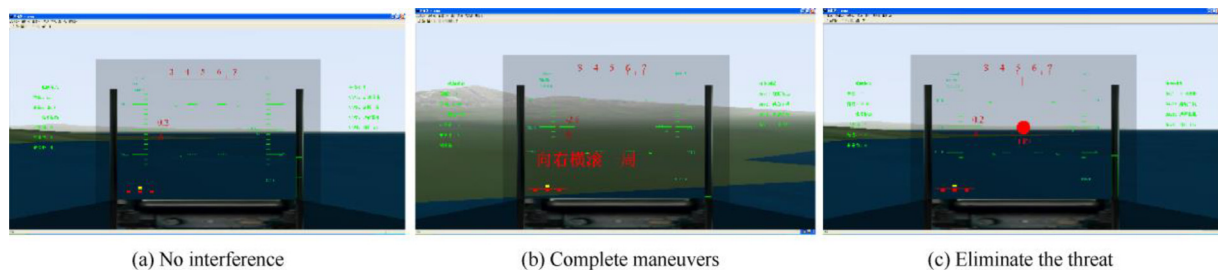


Fig. 2 Two different flight interference task.

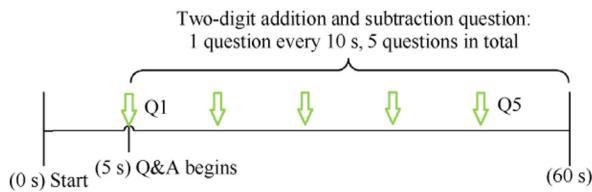


Fig. 3 Mathematical calculation task.

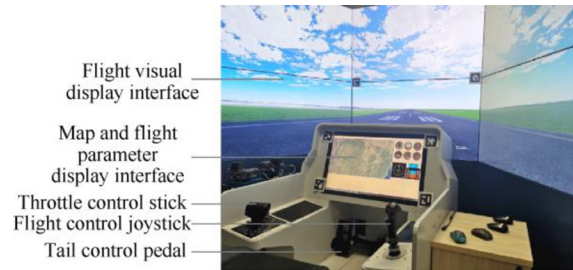


Fig. 4 Flight experiment system.

Difficulty level	Experimental design		
	Designated route mission	Interference elimination mission	Mathematical calculation mission
I	Straight route		
II	Simple serpentine route		
III	Simple serpentine route	Complete maneuvers	
IV	Complex serpentine route	Complete maneuvers, Eliminate the threat	
V	Complex serpentine route	Complete maneuvers, Eliminate the threat	Mathematical calculation

parameter information. The throttle control stick is used to control the speed of the aircraft. The flight control joystick is used to control the attitude of the aircraft. The tail control pedal can control the tail wing of the aircraft to realize a small range of attitude adjustment.

(2) Data collection system

The data collection system includes the Dikablis 3.0 eye tracker, BioRadio physiological instrument, and D-Lab data processing software, which complete the real-time collection, preprocessing and forwarding of eye movement physiological data. The primary data types and their definitions are shown in Table 2 and Table 3.

Among them, the Dikablis 3.0 eye tracker uses an infrared camera to record the subject's eyeballs and then uses software to calculate the current person's gaze position. The sampling frequency of the eye tracker is 60 Hz, and the gaze tracking accuracy is 0.1°-0.3°.

2.4. Experimental equipment

The equipment used in the experiment mainly includes a flight experiment system and a data collection system.

(1) Flight experiment system

The flight experiment system is composed of an aircraft cockpit, a three-dimensional virtual view, and flight and mission software, as shown in Fig. 4. It provides functions such as real-time aircraft control, human-computer interaction and mission operation.

In the flight simulation system used in our previous work,²⁹ six borderless Liquid-Crystal Display (LCD) screens form the flight visual display interface of the cockpit. The splicing of the six screens has a certain angle, creating a semi-enclosed shape, which can increase the immersion and realism during simulated flight. The touchable 30 in (1 in = 2.54 m) display is used for the lower display of the cockpit, displaying map and flight

Table 2 Eye movement data and its definition.

Eye movement data	Definition
Fixation duration (ms)	The duration of fixating at a certain information for more than 1.5 s is defined as fixation, and the fixation duration is the length of time for one fixation.
Number of fixations (times)	How many fixations are formed in a certain period of time.
Saccade duration (ms)	The length of time it takes for the line of sight to transfer from one message to another.
Number of saccades (times)	The number of saccades that occur within a period of time.
Saccade angle (°)	The angle at which the line of sight transfers from one information to another.
Pupil area (pixel)	Describe the size of the pupil.

Table 3 Physiological data and its definition.

Physiological data	Definition
Electromyography (EMG) (mv)	The superposition of motor unit action potentials in many muscle fibers in time and space.
Pulse (times)	Arterial pulsation that can be touched on the human body.
Heart rate(n/min)	The number of heart beats per minute.
Blood oxygen saturation (%)	The concentration of blood oxygen in the blood.
Respiration rate (chest) (mv)	Monitor the chest breathing rate through the voltage change of the breathing belt sensor.
Respiration rate (abdomen) (mv)	Monitor the abdominal breathing frequency through the voltage change of the breathing belt sensor.

The BioRadio physiological instrument is a physiological data collection device that integrates multiple physiological measurement sensors. It mainly includes EMG sensor (used to measure muscle conductivity), respiration rate sensor (used to measure the respiration rate of the abdomen and chest) and infrared blood oxygen sensor (used to measure heart rate, pulse and blood oxygen saturation). The supporting data processing software is Bioradio, which can display various physiological indicators in real-time and send the data to the D-lab via Transfer Control Protocol (TCP) to synchronize with the eye movement data.

2.5. Questionnaire

After each experiment, the subject must fill out the National Aeronautics and Space Administration Task Load Index

(NASA-TLX) questionnaire.³⁰ It includes six dimensions: mental demand, physical demand, temporal demand, performance, effort and frustration. Each dimension is divided into 21 levels, with values ranging from 0 to 20. The subject marks his level in each dimension according to his feeling and then calculates the score according to equation 1.

$$NASA - TLX = \frac{5 \times \sum_{i=1}^6 n_i}{6} \tag{1}$$

where n_i represents the value of the i th dimension, and the score of NASA-TLX ranges from 0 to 100, as shown in Fig. 5.

3. Real-time evaluation method of flight mission load based on sensitivity analysis of physiological factors

This section introduces the real-time evaluation method of flight mission load based on sensitivity analysis of physiological factors. It mainly consists of two parts. First, the factors with significant sensitivity are screened out through the comprehensive test method. Then the pilot mission load evaluation model is established based on the SVM hierarchical combination classification method.

3.1. Comprehensive test method for factor sensitivity

Before classifying the data, to improve the classification effect and reduce the impact of insensitive factors on the classification effect, sensitivity analysis of the factors is first required, that is, to select more sensitive factors to the difficulty of the experiment before classification.

By combining the different characteristics of multiple testing methods, a multi-parameter sensitivity analysis method of physiological factors based on comprehensive testing is designed in this article, as shown in Fig. 6.

Name	Task	Date
1. Mental demand How mentally demanding was the task?		
Very low		Very high
2. Physical demand How physically demanding was the task?		
Very low		Very high
3. Temporal demand How hurried or rushed was the pace of the task?		
Very low		Very high
4. Performance How successful were you in accomplishing what you were asked to do?		
Very low		Very high
5. Effort How hard did you have to work to accomplish your level of performance?		
Very low		Very high
6. Frustration How insecure, discouraged, irritated, stressed, and annoyed were you?		
Very low		Very high

Fig. 5 NASA-TLX questionnaire.

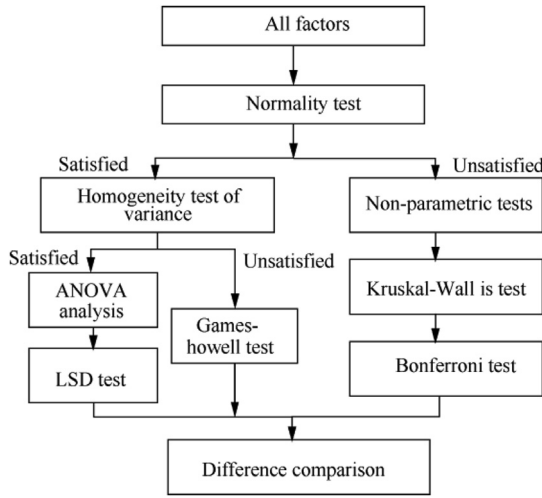


Fig. 6 Comprehensive test method for factor sensitivity.

The sensitivity analysis of factors is first necessary to determine what type of distribution the factor belongs to. Then choose different test methods according to the different distributions to judge whether there is a significant difference and the size of the difference between the difficulty of each mission of the factor.

To determine the type of factor distribution, the normality test must first be carried out. Among the commonly used normality test methods, the Shapiro-Wilk test is generally suitable for small sample tests. Therefore, the Shapiro-Wilk test is used in the standard test of eye movement and physiological factors. The test statistic W can be calculated by:

$$W = \frac{\left(\sum a_i y_{(i)}\right)^2}{\sum (y_i - \bar{y})^2} \quad (2)$$

where \bar{y} is the average value of the sample; $y_{(i)}$ is the smallest order statistic, $i = 1, 2, \dots, n$; a_i is determined by a specific formula. After calculating the statistic W , it can be judged whether the factor satisfies the normality test according to the designed significance level.

If the factor satisfies the normality test, it will be tested for the homogeneity of variance. The Lenene test method is used to test the homogeneity of variance. This method is more robust than other homogeneity tests, and the number of sample data in each group can be different. The test statistic W of this method can be obtained by:

$$W = \frac{(N - k) \sum_{i=1}^k N_i (\bar{Z}_i - \bar{Z}_{..})^2}{(k - 1) \sum_{i=1}^k \sum_{j=1}^{N_i} (Z_{ij} - \bar{Z}_i)^2} \quad (3)$$

where k is the number of sample groups, N_i is the data volume of the i th sample, N is the sum of the data volume of each sample, and Z_{ij} is the new variable value of the original data after the data conversion. \bar{Z}_i is the mean of the i th sample, and \bar{Z} is the total mean of all indicator data. After calculating the statistic W , it can be judged whether the factor satisfies the test of homogeneity of variance according to the designed significance level.

After the normality test and homogeneity of variance test, Analysis of Variance (ANOVA) is used to test the mean of the sample data to determine whether the five sets of sample data belong to the same distribution. The test statistic F is:

$$F = \frac{MSA}{MSE} \quad (4)$$

where MSA is the between-group variance, and MSE is the within-group variance. After calculating the statistic F , it can be judged whether the factor satisfies the ANOVA test according to the designed significance level.

For the factors that meet the above tests, they need to be compared afterward to determine which of the five mission difficulties have significant differences.

Here, Bonferroni Method is used to judge the difference between the difficulty of each mission of the indicator. Compared with other post-comparison methods, this method is more effective when the number of comparisons is small.

This method is based on Bonferroni's inequality, and its principle is expressed in mathematical language: assuming that each hypothesis m is tested at the α^* level and the probability of rejection is R_m , then the total probability of making the Type 1 Error is:

$$\alpha = \prod_{m=1}^M R_m \leq \sum_{m=1}^M R_m = M\alpha^* \quad (5)$$

According to the designed significance level, it can be judged whether there is a significant difference between the difficulty of each mission.

For the factors that do not meet the homogeneity of variance test, the Games-Howell method is used to compare them afterward. The Games-Howell method is suitable for pairwise comparison when the variance of the sample data is not uniform, and the sample size of each group can be equal or unequal. The test is based on Welch's correction of the degree of freedom of the t-test and uses the studentized range as a statistic. The test statistic of this method is:

$$t_w = \frac{\bar{x}_i - \bar{x}_j}{\sqrt{\frac{s_i^2}{n_i} + \frac{s_j^2}{n_j}}} \quad (6)$$

where s_i^2 and s_j^2 are the variances of the samples in the i and j groups respectively; n_i and n_j are the number of samples respectively; \bar{x}_i and \bar{x}_j are the mean values of the two samples. According to the designed significance level, it can be judged whether there is a significant difference between the difficulty of each mission.

For factors that do not meet the normality test, that is, when the distribution of the population is unknown, the use of parametric hypothesis testing will produce larger errors. At this time, the use of non-parametric hypothesis testing has great advantages. Since the five mission difficulties are designed in this article, non-parametric hypothesis tests are used for indicators that do not conform to the normal distribution. The specific method is the Kruskal-Wallis test with multiple independent samples. The test statistic of this method is:

$$K - W = \frac{12}{n(n+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(n+1) \quad (7)$$

where k is the number of sample groups, n is the total number of samples, n_i is the data volume of the i th sample; R_i is the sum of the ranks of the i th group of samples, and R_{ij} is the rank value of the j th sample value in the i th sample. Then, the Bonferroni Method is used to compare them afterward. After sensitivity testing of all physiological and eye movement factors, factors with significant differences can be screened out for classification.

3.2. Flight mission load evaluation model based on SVM hierarchical combination

This part is an introduction to the flight mission load evaluation model, which is based on the SVM hierarchical combination classification method.

3.2.1. Modeling principle

Support Vector Machine (SVM) is a two-class classification model. SVM finds the support vectors to reduce the computational complexity, finds an optimal interface between the sample sets of the two classes, separates the two classes, and keeps the hyperplane away from the “margin” are the largest.^{29,31} Its decision equation is:

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b \quad (8)$$

where $\mathbf{w} = [w_1, w_2, \dots, w_d]$ is the normal vector and b is the displacement term. x is a vector composed of eye movement and physiological data normalized by the classification factor used by the corresponding classifier. The normalization method is:

$$x_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (9)$$

3.2.2. SVM hierarchical combination classifier

SVM hierarchical combination classifier is used to achieve five classifications of mission load, and each layer of classifier uses different multi-parameter physiological data as classification factors. The final five-class model of mission load is shown in Fig. 7.

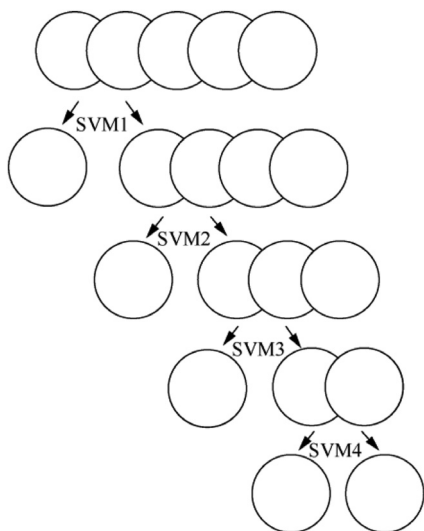


Fig. 7 SVM hierarchical combination classification model.

Among them, the factors used by each classifier are determined by the results of sensitivity analysis. The data after the normalization of each factor is used as a training sample, and the parameters \mathbf{w} and b of each classifier can be obtained through training so as to obtain the final evaluation model. Combining the obtained parameters, the level of mission load can be judged according to the relationship between the decision equation and 0.

3.2.3. Real-time analysis of the model

The factors with significant sensitivity can be selected through the factor sensitivity test on the eye movement data and physiological data collected in the experiment. Combining the SVM two classifiers in layers can build a mission load evaluation model based on sensitivity significant factors.

The real-time performance of the model is guaranteed by the following settings:

- (1) In actual use, through an eye tracker and a physiological instrument, the pilot's eye movement and physiological data can be collected in real-time. The sampling frequency of the eye tracker is 60 Hz, and the sampling frequency of the physiological instrument is 250 Hz.
- (2) The statistical period of the data is set to 10 s, which not only ensures sufficient data to analyze the mission load, but also ensures that the current mission load level can be obtained in real time.
- (3) The model is calculated by SVM, and the running time of SVM itself can meet the real-time requirements.

4. Results and analysis

This section is the display and analysis of the experimental results. It mainly includes experimental data collection and preprocessing, factor sensitivity test results and SVM hierarchical combination classification model.

4.1. Experimental data collection and preprocessing

In the experiment, the subject needs to complete the investigation according to the designed flight mission. 30 sets of missions of each difficulty level, so there are 150 sets of experiments in total. During the experiment, the subject is required to wear the eye tracker and the physiological instrument to record his eye movement data and physiological data. After the experiment, the subject is required to fill out the NASA-LTX questionnaire according to their real feelings and calculate the score.

The raw data collected in the experiment contains a large number of null values. To convert the raw data into data that can be used for analysis, the null values must be cleared first. Then the characteristic values that can represent the characteristics of each indicator in each set of experiments must be extracted.

For the pulse signal, it needs to be filtered first. Then the number of waveforms appearing in each group of experimental data is counted as the pulse characteristic of the group of experiments. For the number of fixations and saccades, it is necessary to count the number of changes in the raw data between 0 and 1 and use this as the number of fixations and saccades for this group of experiments. Other factors are the

As shown in Fig. 10, for the factor of the number of fixations, the mission difficulty groups with significant differences are ($p < 0.05$):

- (1) Mission difficulty I and mission difficulty II, III, IV, V.
- (2) Mission difficulty II and mission difficulty I, III, V.
- (3) Mission difficulty III and mission difficulty I, II, IV, V.
- (4) Mission difficulty IV and mission difficulty I, III, V.
- (5) Mission difficulty V and mission difficulty I, II, III, IV.

Respiration rate (abdomen) is a factor that satisfies the normality test ($p = 0.244, p > 0.05$). Still, it does not fulfill the homogeneity of variance test ($p = 0$), so the difference between the experimental groups of this factor can be seen through the Games-Howell test, as shown in Fig. 11.

From Fig. 11, for the factor of respiration rate (abdomen), the mission difficulty groups with significant differences are ($p < 0.05$):

- (1) Mission difficulty I and mission difficulty III.
- (2) Mission difficulty III and mission difficulty I, IV.
- (3) Mission difficulty IV and mission difficulty III.

The rest of the factors did not satisfy the normality test ($p < 0.05$), so a non-parametric test (Kruskal-Wallis test for independent samples) is performed. There are significant differences between the experimental groups ($p < 0.05$), including EMG, blood oxygen saturation, heart rate, pupil area, fixations duration, and the number of saccades.

Through the Bonfroni method, the differences between the specific groups of the factors (in the order above) are shown in Fig. 12. The four factors of pulse, respiration rate (chest), saccade duration and saccade angle do not satisfy the normality test ($p < 0.05$), and there is no significant difference between the experimental groups in the non-parametric test ($p > 0.05$).

Therefore, four physiological factors such as EMG, respiration rate (abdomen), heart rate and blood oxygen saturation, and four eye movement factors such as pupil area, fixation duration, the number of fixations and the number of saccades have significant differences. Fig. 13 shows the normalized

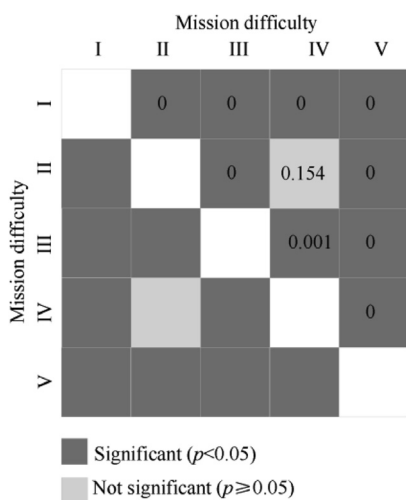


Fig. 10 Number of fixations test result.

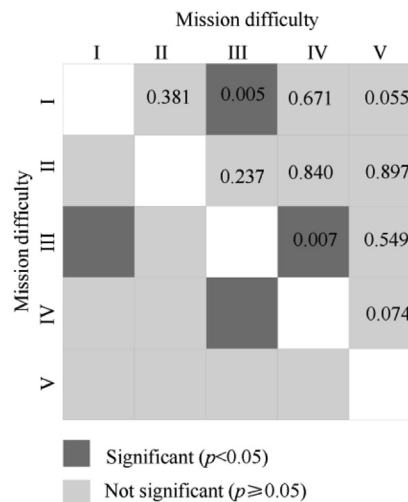


Fig. 11 Respiration rate (abdomen) test result.

mean changes of the experimental data for each mission difficulty of these factors with significant differences.

In Fig. 13, the EMG, heart rate, respiration rate (abdomen), fixation duration, number of fixations and number of saccades reached their peaks at mission difficulty III. Considering the impact of classification order on classification difficulty, Mission Difficulty III can be selected first to be separated from other mission difficulties. Then, the classification factors can be selected from the above six factors.

In the remaining mission difficulty, factors such as EMG, heart rate, blood oxygen saturation, pupil area and the number of saccades reached their peaks at mission difficulty IV. Therefore, mission difficulty IV can be selected to be separated from other mission difficulties in the second step.

The classification order of the remaining mission difficulties and the factor selection method are consistent with the above. According to the results of the sensitivity analysis, combined with the mean changes, and considering the impact of the classification order on the difficulty of classification, the final classification order and the factors used (Table 5) are as follows:

- Step 1. Separate III from I, II, IV, and V.
- Step 2. Separate IV from I, II, and V.
- Step 3. Separate I from II and V.
- Step 4. Separate II and V.

4.4. SVM hierarchical combination classification model

This part is a demonstration of the flight mission load evaluation model and its instructions.

4.4.1. Modeling

According to the classification order determined above, combined with the SVM hierarchical combination classifier structure, the final mission load five classification model is shown in Fig. 14.

And according to the factor sensitivity test results, the final factors for each classifier are shown in Table 6.

After the normalization of each factor, the data is used as a training sample, and the final evaluation model can be

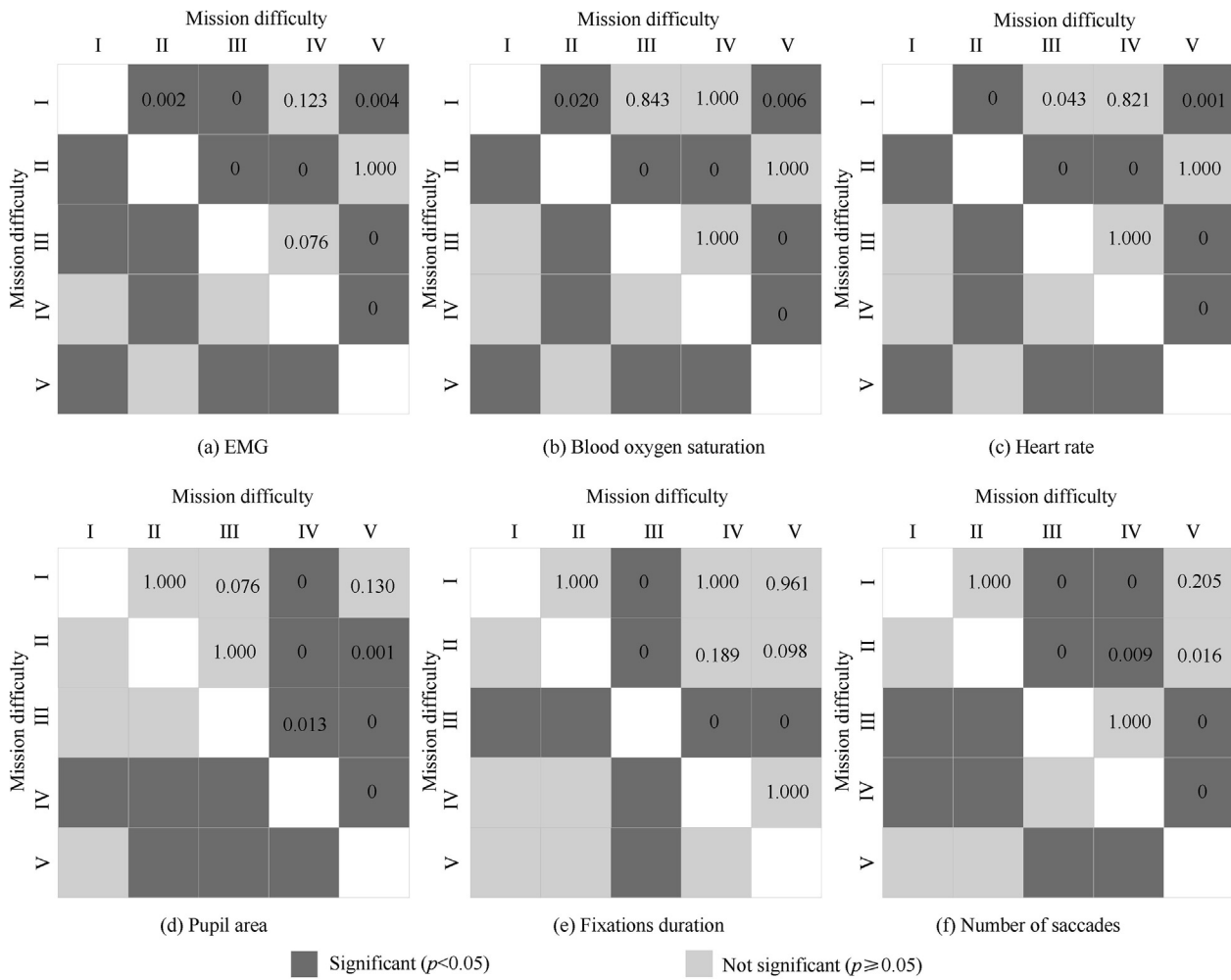


Fig. 12 Test results of other factors.

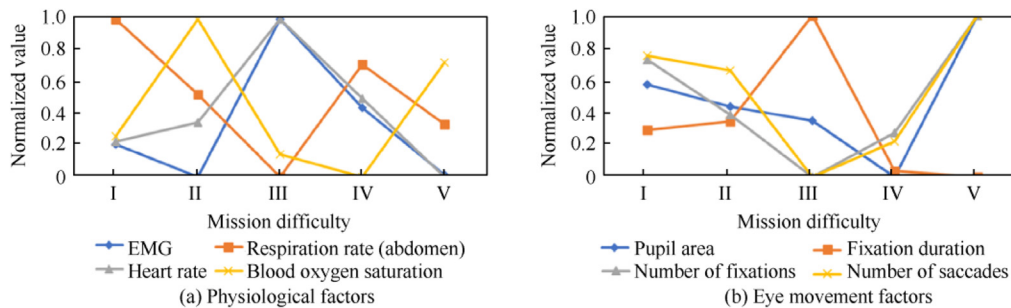


Fig. 13 Mean changes of physiological factors and eye movement factors.

obtained through training. The parameters w and b of each classifier obtained by training are (Table 7):

Combining the above parameters, the level of mission load can be judged according to the relationship between the decision equation and 0.

4.4.2. Model instructions

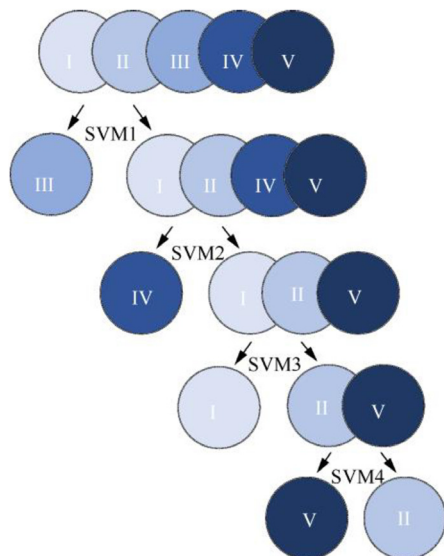
According to the established SVM hierarchical combination classifier, combined with the collected multi-modal physiolog-

ical data of the pilot, the current pilot's mission load level can be calculated through the decision equation. The specific process is as follows:

First, the collected data is normalized, and then the classifier process is combined with the factors of each classifier to calculate. The current mission load level of the pilot is judged according to the result of the decision equation: If the result of the decision equation of the SVM1 classifier is greater than 0, the pilot is in a Medium mission load (M) state, if it is less than

Table 5 Classification order and factors used.

Step	Factors
1	EMG, Respiration rate (abdomen), Fixation duration, Number of fixations
2	EMG, Heart rate, Pupil area, Number of saccades
3	EMG, Respiration rate (abdomen), Heart rate, Blood oxygen saturation
4	Pupil area, Fixation duration, Number of fixations, Number of saccades

**Fig. 14** Mission load five classification model.**Table 6** Factors of each classifier.

Classifier	Factors
SVM1	EMG, Respiration rate (abdomen), Fixation duration, Number of fixations
SVM2	EMG, Heart rate, Pupil area, Number of saccades
SVM3	EMG, Respiration rate (abdomen), Heart rate, Blood oxygen saturation
SVM4	Pupil area, Fixation duration, Number of fixations, Number of saccades

Table 7 Parameters of each classifier.

Classifier	Parameters
SVM1	$w = [3.811, -1.118, 1.320, -1.574]^T$, $b = -2.004$
SVM2	$w = [3.696, 0.609, -2.758, -2.521]^T$, $b = 0.193$
SVM3	$w = [5.683, 1.370, 1.090, -1.879]^T$, $b = -1.810$
SVM4	$w = [2.852, -0.724, 4.361, 0.934]^T$, $b = -3.417$

0, the second classification is performed. If the result of the decision equation of the SVM2 classifier is greater than 0, the pilot is in a Higher mission load (H) state. If it is less than 0, the third classification is performed. If the result of the decision equation of the SVM3 classifier is greater than 0, the pilot is in a Very Low mission load (VL) state. If it is less than 0, the fourth classification is performed. If the result of the decision equation of the SVM4 classifier is greater than 0, the pilot is in a Very High mission load (VH) state; otherwise, the pilot is in a Low mission load (L) state.

5. Model effect evaluation

After establishing the five-class evaluation model of pilot mission load, further evaluation of its classification effect is required. In this article, the model's ability to distinguish between different classes is analyzed using confusion matrix analysis.

According to the classification results, the confusion matrix of each classifier can be obtained, as shown in Fig. 15:

Fig. 15 shows the confusion matrices of the classification results obtained by SVM1 to SVM4. In this work, 150 sets of experiments were conducted, and 30 sets of difficulty for each mission. The class labels indicate whether a particular set is with the correct mission difficulty. In each classification, 30 groups are positive samples (label as 1), and the rest are negative samples (labeled as 0) because they do not have the specific mission difficulty. And after each classifier is classified, the positive samples of this classification will not participate in the subsequent classification process. That is, the total number of classifications will be reduced by 30. By reading these confusion matrices, all classifiers can classify almost all the negative samples, but SVM2 and SVM3 had misclassified samples for both classes. Some of the negative samples were predicted as positive because the experimental design was not distinguished enough. In general, it can be roughly seen from the confusion matrix that the classification effect of the four classifiers is good.

Through the confusion matrix, some evaluation indexes can further evaluate the classification effect of the classification model, including error, accuracy, precision, recall, F-value, ROC curve and AUC value. The specific explanation is as follows:

- (1) Error (ERR): the ratio of the number of incorrectly predicted samples to the number of all predicted samples, the expression is:

$$\text{ERR} = \frac{\text{FP} + \text{FN}}{\text{FP} + \text{FN} + \text{TP} + \text{TN}} \quad (10)$$

- (2) Accuracy (ACC): the ratio of the number of correctly predicted samples to the number of all predicted samples, the expression is:

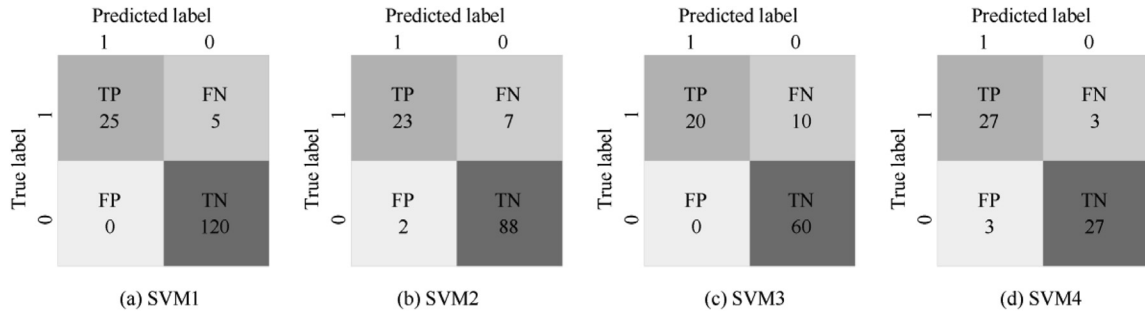


Fig. 15 Confusion matrix of each classifier.

$$ACC = \frac{TP + TN}{FP + FN + TP + TN} \quad (11)$$

- (3) Precision (P): the ratio of the number of samples predicted to be positive to the number of true positive samples, the expression is:

$$P = \frac{TP}{TP + FP} \quad (12)$$

- (4) Recall (R): the ratio of the number of correct positive samples to the number of all positive samples, the expression is:

$$R = \frac{TP}{TP + FN} \quad (13)$$

- (5) F-value: harmonic average of precision and recall, the expression is:

$$F - \text{value} = \frac{2}{\frac{1}{P} + \frac{1}{R}} = \frac{2(P \times R)}{P + R} \quad (14)$$

- (6) ROC curve: Receiver Operator Characteristic curve, which is a tool for classification model selection based on performance factors such as model True Positive Rate (TPR) and False Positive Rate (FPR). Among them, TPR and FPR reflect the ratio of the true class to the positive class and the ratio of the false positive class to the negative class. The equations are defined as follows:

$$TPR = \frac{TP}{FN + TP} \quad (15)$$

$$FPR = \frac{FP}{FP + TN} \quad (16)$$

- (7) AUC value: Area Under the Curve, whose value is within 1. The larger the AUC value, the better the classifier performance.

The evaluation indexes of each classifier can be calculated through the confusion matrix, and the specific evaluation results are shown in Table 8. The ROC curve of each classifier is shown in Fig. 16.

Based on the above evaluation indexes, the classification effect of the established SVM hierarchical combination classification model can be evaluated systematically. From the comprehensive evaluation index F-value and AUC value, the positive index value of each classifier is above 0.8, so it can be considered that the model has a good classification effect on the mission load of the five mission difficulties.

To further verify the validity and credibility of the mission load classification model established in this paper, 150 groups of eye movement and physiological data collected in the experiment are classified again according to the previously determined classification order using the Back Propagation (BP) neural network algorithm.

When the BP neural network is used for mission classification, the maximum number of training iterations is 1000, the learning rate is 0.01, and the training accuracy is 0.01. After the classification is completed, the evaluation index of accuracy is selected to compare the two algorithms, and the comparison result is shown in Fig. 17.

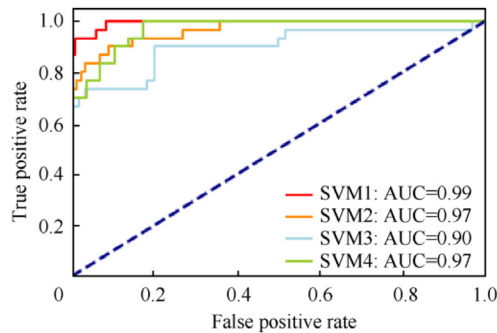
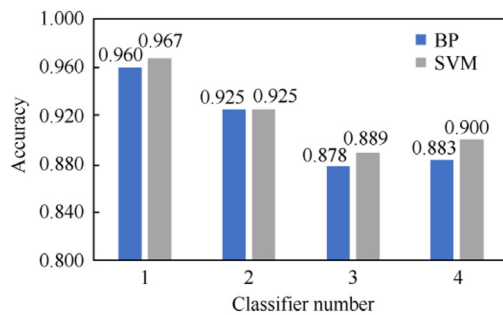
Fig. 17 shows that when the classification of the first, third, and fourth layers is performed, SVM's classification accuracy is higher than that of the BP neural network by 0.007, 0.011, and 0.007, respectively. In the second layer of classification, the accuracy of the two classification algorithms is equal. In general, both algorithms have achieved good results in terms of accuracy.

However, for the high-dimensional small sample data collected in this experiment, the BP neural network is easy to fall into the local extreme value, and the weights converge to the local minimum point so that the algorithm will get different results every time it is trained. The SVM algorithm has strong generalization ability and is a convex function, which is not easy to fall into a local minimum.

In addition, the BP algorithm has a multi-layer network structure, so more model parameters need to be optimized. As a result, it occupies a large amount of memory and is slower (Each layer's average training running time is 0.04 s). On the other hand, the SVM algorithm is simple to operate, requires

Table 8 Evaluation indexes of each classifier effect.

Evaluation indexes	Classifier			
	1	2	3	4
ERR	0.033	0.075	0.110	0.100
ACC	0.967	0.925	0.889	0.900
P	1	0.920	1	0.900
R	0.833	0.767	0.667	0.900
F-value	0.909	0.836	0.800	0.900
AUC	0.99	0.97	0.93	0.97

**Fig. 16** ROC curve of each classifier.**Fig. 17** Comparison of the accuracies of two algorithms.

fewer model parameters to be optimized, and occupies a small amount of memory and is fast (The average training running time of each layer is 0.001 s).

Therefore, in general, compared to the BP neural network algorithm, the SVM algorithm is more suitable for the online mission load evaluation research based on eye movement and physiological data required in this paper.

6. Conclusions

Based on the sensitivity analysis of physiological factors, the real-time evaluation method of mission load is carried out. The main contributions of this paper include:

- (1) Flight mission experiments, including aircraft control, human-computer interaction and mental arithmetic tests, are designed to simulate five mission loads with

different flight difficulties. Compared with other studies, the flight experiment mission designed in this paper is more comprehensive and the cockpit equipment used is closer to the real flight scene. However, there are still some gaps between the simulated flight scenes in the laboratory and the real flight scenes, for example, the reliability of mental arithmetic simulation analysis. In the future, we will continue to improve the experimental design to increase the reality of the experiment and the reliability of pilot data.

- (2) A sensitivity analysis method based on a comprehensive test is proposed. The physiological factors that are sensitive to the difficulty of the flight mission are selected, including fixation duration, number of fixations, number of saccades, pupil area, EMG, heart rate, respiration rate (abdomen) and blood oxygen saturation. Compared with other studies that have not been tested or used a single test method, modeling using the indicators selected by the comprehensive test method proposed in this article can significantly improve classification accuracy.
- (3) The Support Vector Machine (SVM) two classifiers are combined hierarchically, and a five-class real-time evaluation model of flight mission load based on high-sensitivity physiological factors is established. The experimental results show that the model's classification accuracy is high. It meets the requirements of actual engineering application, which can provide methodological references for mission load evaluation research in other fields. In future work, we will further improve the model to improve the accuracy of the classification and applicability.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study was co-supported by the Aeronautical Science Foundation of China (No. 2020Z023053002) and the National Natural Science Foundation of China (No. 61305133).

References

1. Wiegmann DA, Shappell SA. Human error analysis of commercial aviation accidents: Application of the Human Factors Analysis and Classification system (HFACS). *Aviat Space Environ Med* 2001;**72**(11):1006–16.
2. Zhao Z, Niu YF, Shen LC. Adaptive level of autonomy for human-UAVs collaborative surveillance using situated fuzzy cognitive maps. *Chin J Aeronaut* 2020;**33**(11):2835–50.
3. Chen H, Pang LP, Wanyan XR, et al. Effects of air route alternation and display design on an operator's situation awareness, task performance and mental workload in simulated flight tasks. *Applied Sciences* 2021;**11**(12):5745.
4. Dismukes RK, Kochan JA, Goldsmith TE. Flight crew errors in challenging and stressful situations. *Aviat Psychol Appl Hum Factors* 2018;**8**(1):35–46.

5. Department of Defense. Unmanned systems integrated roadmap. Washington, D.C.: Defense Technical Information Center; 2009. Report No.: FY2009-2034.
6. Zhan H, Chen Y. *Research on flight fatigue*. Beijing: National Defense Industry Press; 2011 [Chinese].
7. Chen J, Zhang QL, Cheng L, et al. A cognitive load assessment method considering individual differences in eye movement data. *2019 IEEE 15th international conference on control and automation (ICCA)*; 2019 Jul 16-19; Edinburgh, UK. Piscataway: IEEE Press; 2019.p. 295–300.
8. Wei ZM, Zhuang DM, Wanyan XR, et al. A model for discrimination and prediction of mental workload of aircraft cockpit display interface. *Chin J Aeronaut* 2014;**27** (5):1070–7.
9. Lehrer P, Karavidas M, Lu SE, et al. Cardiac data increase association between self-report and both expert ratings of task load and task performance in flight simulator tasks: An exploratory study. *Int J Psychophysiol* 2010;**76**(2):80–7.
10. Sugiono S, Widhayaturiyawan D, Andriani DP. Investigating the impact of road condition complexity on driving workload based on subjective measurement using NASA TLX. *MATEC Web Conf* 2017;**136**:02007.
11. Wei ZM, Hao HX, Xu QZ, et al. Review of research on pilots' mental workload evaluation indices and measurement methods. *Sci Technol Eng* 2019;**19**(24):1–8 [Chinese].
12. Luo J, Cheng P, Lu BC, et al. Analysis of driver workload on optical tunnel based on driver HRV. *J Chongqing Univ Technol Nat Sci* 2019;**33**(8):44–50, 64 [Chinese].
13. Putze F, Jarvis JP, Schultz T. Multimodal recognition of cognitive workload for multitasking in the car. *2010 20th International Conference on Pattern Recognition*; 2010 Aug 23-26; Istanbul, Turkey. Piscataway: IEEE Press; 2010.p.3748–51.
14. Yang YQ, Chen YB, Wu CX, et al. Effect of highway directional signs on driver mental workload and behavior using eye movement and brain wave. *Accid Anal Prev* 2020;**146**:105705.
15. Chihara T, Kobayashi F, Sakamoto J. Evaluation of mental workload during automobile driving using one-class support vector machine with eye movement data. *Appl Ergon* 2020;**89**:103201.
16. Orlandi L, Brooks B. Measuring mental workload and physiological reactions in marine pilots: Building bridges towards redlines of performance. *Appl Ergon* 2018;**69**:74–92.
17. Othman N, Romli FI. Mental workload evaluation of pilots using pupil dilation. *Int Rev Aerosp Eng IREASE* 2016;**9**(3):80.
18. Qi WW, Shen B, Wang LH. Model of driver's eye movement and ECG index under tunnel environment based on spatiotemporal data. *J Adv Transp* 2020;**2020**:1–11.
19. Li WC, Chiu FC, Wu KJ. The evaluation of pilots' performance and mental workload by eye movement. *30th conference of european association for aviation Psychology*; 2012 Sep 24-28; Sardinia, Italy; 2012.
20. Xin T, Luo Z, Jiang L, et al. Analysis of pilot workload during the landing phase of the aircraft. *Chin J Ergonom* 2019;**25**(2):62–5 [Chinese].
21. Feng C, Wanyan X, Yang K, et al. A comprehensive prediction and evaluation method of pilot workload. *Technol Heal Care* 2018;**26**:65–78.
22. Guo SN, Wanyan XR, Liu S, et al. Influences of intelligent design and information processing modality complexity on occupant mental workload. *Acta Armamentarii* 2021;**2**:234–41 [Chinese].
23. Peng GC, Nourani M, Harvey J, et al. Feature selection using F-statistic values for EEG signal analysis. *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*; 2020 Jul 20-24; Montreal, Canada. Piscataway: IEEE Press; 2020.p.5963–6.
24. Sun Z, Qiao YC, Lelieveldt BPF, et al. Integrating spatial-anatomical regularization and structure sparsity into SVM: Improving interpretation of Alzheimer's disease classification. *NeuroImage* 2018;**178**:445–60.
25. Kaur B, Singh D, Roy PP. EEG based emotion classification mechanism in BCI. *Procedia Comput Sci* 2018;**132**: 752–8.
26. Wilson GF, Russell CA. Operator functional state classification using multiple psychophysiological features in an air traffic control task. *Hum Factors* 2003;**45**(3):381–9.
27. Zhang J, Pang LP, Wanyan XR, et al. Method for operator mental workload assessment based on power spectral density of EEG. *Acta Aeronaut et Astronauti Sin* 2020, **41** (10): 123618 [Chinese].
28. Cordeiro H, Fonseca J, Guimarães I, et al. Hierarchical classification and system combination for automatically identifying physiological and neuromuscular laryngeal pathologies. *J Voice* 2017;**31**(3), p. 384.e9-e14.
29. Chen J, Xue L, Liu ZC. A pilot workload evaluation method based on EEG data and physiological data. *2020 IEEE international conference on signal processing, communications and computing (ICSPCC)21*. China. Piscataway: 24; MacaoIEEE Press; 2020 Aug. p. 2020.p.1–6.
30. Chen J, Liu ZC, Fu XW. *A sensitivity analysis method of physiological signals for assessing pilot workload*, 9. 11; Singapore. Piscataway: IEEE Press; 2020 Oct. p. 2020.p.737–42.
31. Zhou ZH. *Machine learning*. Beijing: Tsinghua University Press; 2016, p. 121-8 [Chinese].