Research on a Personalized Classifier of Health Status Based on Pulse Signal

Xiaoping Jiang, Leli Sun*, Shuyao Feng, Zhuojing Li, Ying Chen, Xingzhuo Chen, Chunman Wang, Aolai He
China University of Mining and Technology (Beijing) School of Mechanical and Electrical engineering, Beijing, CHINA 100083

ABSTRACT

At present, the workload of mental workers in society is getting heavier and heavier, and it is necessary to assess their health status. Compared with other physiological signals, the pulse is easy to obtain and non-invasive. In this paper, through pulse signal detection, pulse data preprocessing and feature extraction, 12 sets of feature values are selected. Then based on these feature data, using support vector machine algorithm modeling, for different testers to build different personalized human physiological state discrimination system. The experimental results show that the classification accuracy rate reaches 91.17%, which proves that the selected feature value has a strong correlation with the physiological state, and the classifier is effective.

Keywords: Pulse wave, Feature extraction, Classifier design, Physiological state evaluation

1. INTRODUCTION

1.1 RESEARCH STATUS

With the rapid development of science and technology, the demand for young mental workers in society is increasing. Typical examples include scientific researchers in aero-space and other industries, program designers who work intensively, and students preparing for various important exams. These mental workers need to continuously main-tain a high degree of concentration at work. After a long time of concentrated work and study, the physiological state of the worker will change greatly, from a state of fullness to a state of mental fatigue which is dominated by mental fatigue and supplemented by physical fatigue. This fatigue state is different from the state of mental fatigue. Physical fatigue caused by physical exercise. Under the combined effect of working pressure and mental fatigue for a long time, people are prone to serious psychological problems such as depression and anxiety, which in turn bring serious hidden dangers to physical health and work safety. Therefore, it is very necessary to evaluate the physiological state of workers, detect whether they are tired and make early warning research, which has strong practical significance.

At present, there are three main research ideas for physiological state assessment [1]: physiological state detection based on subjective evaluation, physiological state detection based on behavior characteristics, and physio-logical state detection based on physiological signals. The accuracy of subjective evaluation is relatively low, and it is often used as an aid to other methods. The behavior characteristic method needs to arrange various sensors according to application scenarios (such as driving scenes, sports scenes), and the process is cumbersome and it is difficult to obtain a widely used unified method. The physiological signal method judges the physiological state by collecting and analyzing the physiological signal of the tester. It is easier to collect and has a wide range of applicable scenarios. Therefore, this method is selected in this article.

In previous studies, the state assessment and detection methods based on physiological signals mainly selected pulse signals, electrocardiogram signals (ECG), brain wave signals (EEG), and myoelectric wave signals (EMG) as the research objects. Wang Fei et al. [2] took the extracted driver's brainwave signals and the corresponding steering wheel information as feature values, and used support vector machines to classify and judge the features to achieve qualitative analysis of fatigue status. Xu Lisheng et al. [3] used a short period of ECG signal extraction features, combined with random forest algorithm to design a classifier to judge the physiological state of the human body. Fu Menglong [4] explored the judgment and classification of muscle fatigue based on the characteristic relationship between surface EMG signal and muscle fatigue, and then developed a real-time online assessment system for muscle state. In these studies, based on physiological electrical signals, most subjects need to wear multiple electrodes on the head, chest, etc., which are invasive and complicated, and are not easy to carry. Compared with the above-mentioned physiological signals, pulse signals have the advantage of being easy to detect. You only need to wear an infrared photoelectric clip or a wristband on the fingertips to complete data collection, which is less invasive to the body and facilitates the production and actual use of the detection device. Patel M et al. [5] confirmed that the pulse signal is closely related to the degree of fatigue and established a mathematical model. According to the generation mechanism of pulse signal, Li Yongping[6] extracted and analyzed the power spectrum peak value, center of gravity and other characteristics, and integrated them, using the Back Pro-pagation (BP) neural network method to identify the state of mental alertness and mental fatigue, and achieved 95% Accuracy. Guo Yanjie [7] used the two-way Long Short-Term Memory (LSTM) neural network algorithm to classify and recognize the pulse data measured before/after lunch break, before/after exercise, and before/after smoking, with an accuracy rate of more than 80%. It can be concluded from previous studies that the pulse signal of the human body does change under different conditions. Some behaviors such as lunch breaks and running will cause differences in the pulse, and this difference can be identified by machine learning and other methods. This is the use of pulse Signals provide a basis and method to evaluate and detect the physiological state of the human body.

1.2 THE NECESSITY AND SIGNIFICANCE OF PERSONALIZED MODELING METHOD

Previous studies have shown the relationship between pulse signal and physiological state and provided a basis for the use of pulse signal to evaluate and detect human physiological state. However, the main problem with the current research methods based on physiological signals is that the physio-logical signals have individual differences. This leads to a situation in which a certain characteristic value collected on different volunteers has the same value, but belongs to different physiological states, which brings trouble to the design of the discrimination algorithm. However, there is no way to completely eliminate individual differences in physiological signals. Previous studies mainly used methods such as limiting the scope of the group and normalization to weaken the influence of this difference. Normalization can only be applied to features related to amplitude, and cannot be used to process features such as heart rate and cycle; narrowing the scope of the study will not completely eliminate individual differences, but will also lead to the loss of universality of research results. Xie Zhi’s research [8] clustered the measured data of each volunteer to obtain multiple cluster centers for analysis, which gave us inspiration. Therefore, on the basis of previous research, this project analyzes the physiological state of samples of different volunteers, and proposes a personalized state detection algorithm that can be used by different people. This avoids the problem of individual differences and makes the research more practical. In other words, the evaluation and detection model for a volunteer is trained by the pulse signal characteristic data collected from the volunteer for a long time, so the model can be guaranteed to be effective for this volunteer. This paper aims to explore such a data training modeling method, to build different assessment and detection models for different people, so as to avoid the problem of individual differences, so that the research has more practical significance.

1.3 RESARCH PROCESS

In this article, we first collected the pulse signal data of 8 volunteers in different mental work situations; then pre-processed it to remove the high-level interference and baseline drift, and extract the pure pulse signal; The third step is to extract the characteristics of the pulse signal, and analyze and screen the validity of the characteristic value to provide a basis for the subsequent classification research; In the end, according to different volunteers pulse sample build eight individual physiological condition assessment system and demonstrates the rationality of the system and practical significance.

Figure 1. steps of the research

2. DATA COLLECTION AND PREPROCESSING

2.1 DATA COLLECTION

In order to fit the mental work physiological state change scenario studied in this article, the objects selected for data collection are college students and graduate students who are mainly engaged in mental work at ordinary times, and they are 20-28 years old. And the pulse is stable, there is no arrhythmia, no cardiovascular and cerebrovascular diseases, and no strenuous physical activity was performed on the day of the measurement. A total of 8 test subjects meeting this standard were selected for this study, including 6 males and 2 females. Multiple people measure at the same time and collect data for 20-30 days. During the measurement, the volunteer sits quietly on a chair, clamps the tip of the index finger or middle finger with the clip of the instrument, and measures the real-time pulse for one minute.

The experimental equipment we used is an infrared pulse sensor produced by Hefei Huake Information Technology Co. Ltd, named HKG-07C. The instrument can convert the pulse signal into an electrical signal and store the amplitude in hexadecimal form for MATLAB to call.

After the collection is over, the volunteers then self-test through the subjective state judgment form, obtain the subjective judgment result of the state corresponding to the pulse data at that moment according to the score, and put it together with the corresponding pulse data to form a data set. This table uses 10 groups of adjectives to describe the state of the body [8], which are "pleasant-painful", "sleepy-awake", "wandering-energy concentration" and so on. And these adjective pairs are divided into 7 grades, defined as "very-comparative-a little bit-no influence-a little bit-comparative-very", and give the corresponding score "+3,+2,+1,0,−1 ,−2,−3" The higher the score, the more fatigued.

Count the total scores of volunteers at the measurement moment and divide the data into four categories based on the scores: when the score is less than −20 points, the label is set as “very easy”; when the score is greater than −20 points and less than 0, the label is set as “easy”; when the score is greater than 0 and less than 20, the label is regarded as “fatigue”; when the score is greater than 20, the label is regarded as deep fatigue. And record the very obvious subjective state (such as: very sleepy, loss of appetite, unable to concentrate, full of energy, etc.). In the end, 940 sets of pulse data from 8 volunteers were obtained. See appendix table 1 for the specific subjective judgment form.

All pulse data in this study were measured with the consent of volunteers. And this research aims to propose a personalized research method, the experiment is only for 8 individual volunteers, not involving the general public.

Figure 2. Data sampling process

2.2 B. DATA PREPROCESSING

In the process of collecting pulse data, because the electronic components of the sensor are susceptible to the surrounding environment and magnetic field during the measurement process, as well as the possible competition-risk phenomenon in AD conversion, the collected signal is usually doped with some high frequencies. Spike noise, also known as "burr". In addition, the existence of distributed capacitance of human organs and tissues and the AC power supply itself can also cause interference. The waveform appears as a 50Hz power frequency sine wave noise superimposed on the pulse signal. The unfiltered pulse signal is shown in Figure 3.

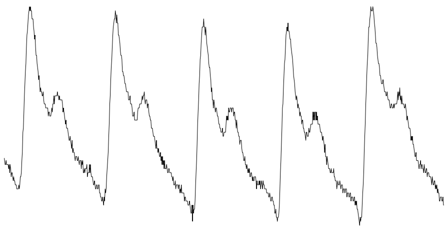


Figure 3. Pulse data graph before smoothing and denoising

This article calls the smoothdata function in the MATLAB library function to smooth and filter the pulse data from the image angle, and the use form is as follows:

$$smoothdata\left(d,'gaussian',10\right)$$

Where *d* is the data, and 'gaussian' is the Gaussian filter. ‘10’ means that the data window is 10. The pulse data image after smoothing and denoising is shown in Figure 4.

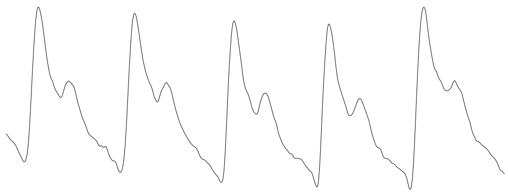


Figure 4. Pulse data graph after smoothing and denoising

In the process of pulse data collection, the subject’s breathing, small muscle movements, or changes in the light intensity in the surrounding environment will affect the infrared photoelectric sensor, resulting in a significant rise in the waveform of the collected pulse data and an abnormal phenomenon. Regular up and down disturbances, this kind of disturbance is called baseline drift [9]. The baseline drift cannot be eliminated or avoided during data collection. Therefore, it is necessary to separate the pure pulse amplitude data from the baseline drift amount through a software algorithm. The acquired pulse signal with baseline drift is shown in Figure 5.

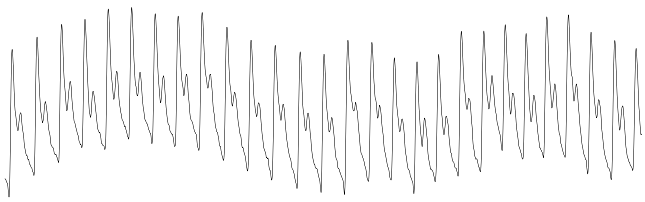


Figure 5. Pulse signal with baseline drift

This paper uses a 7-layer wavelet decomposition method to remove baseline drift. First import the pulse signal. Then, combined with the researched literature, Sym8 wavelet [22] is selected as the wavelet basis function for 7-layer wavelet decomposition. Since the acquisition frequency of the sensor is 200Hz, according to the Nyquist sampling theorem, the frequency band of each layer should be 0.78125Hz. The second step is to set the approximate coefficient A1 after expansion of the seventh layer to zero, that is, to remove the 0-0.78125Hz baseline signal. The separated baseline signal is shown in Figure 6. The third step is to set the detail coefficients of the first and second layers to zero (D1, D2), and remove the signals in the two frequency bands above 25Hz-50Hz and 50Hz to achieve a more thorough high-frequency filtering effect. Finally, perform signal re-construction to obtain pulse data that removes baseline drift and strengthens high-frequency filtering. The effect before and after baseline separation is shown in Figure 7.

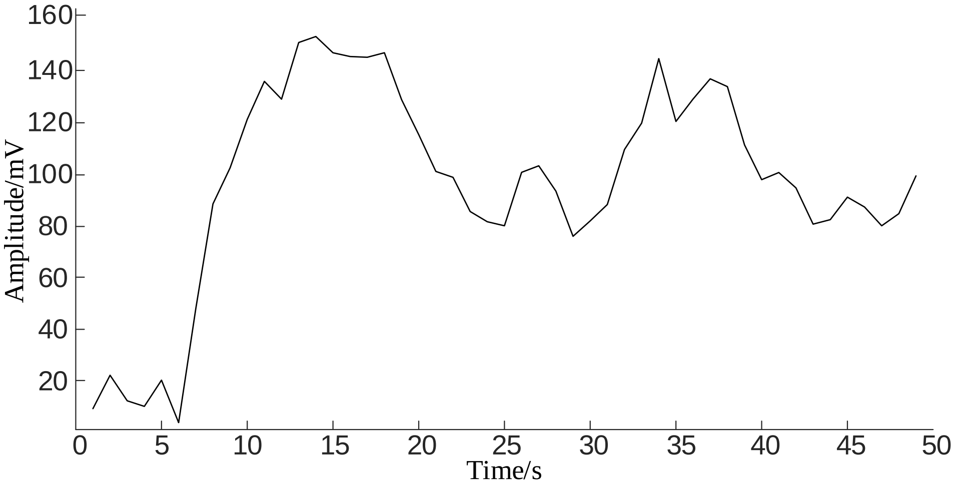


Figure 6. Schematic diagram of 7-layer wavelet decomposition

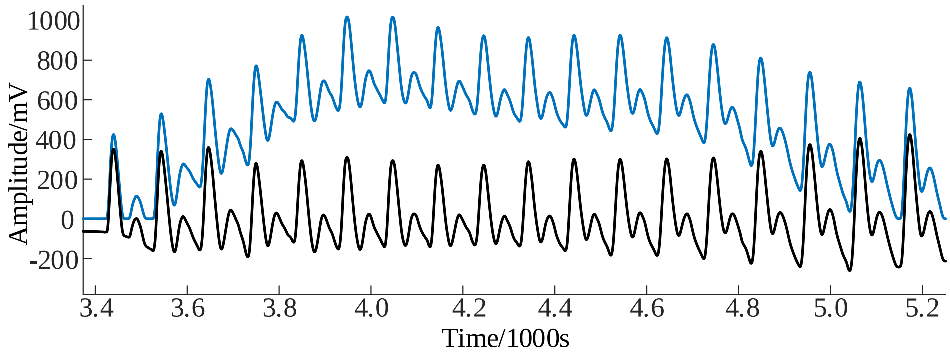


Figure 7. Effect picture before and after baseline separation

3. FEATURE EXTRACTIO OF PULSE DATA

3.1 TIME DOMAIN CHARACTERISTICS

Based on the summary of previous studies on the time domain characteristics of pulse images, and comparing the waveforms actually measured during fatigue and relaxation, it was decided to calculate and extract the heart rate (pulse period), main wave height, main wave period, and weight from the data. Data such as the height of the stroke wave, the period of the dicrotic wave, and the corresponding ratio are used as characteristic values. The relationship between these values and the physiological processes and states of the human body and the basis for their extraction are described as follows:

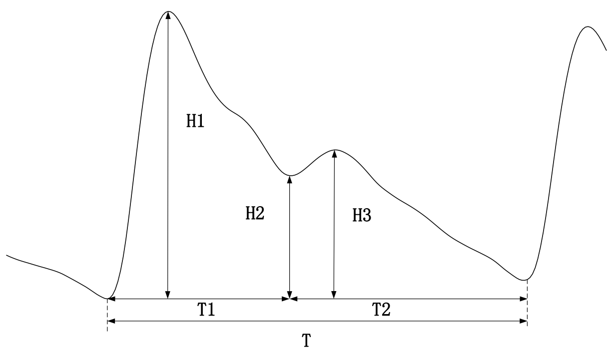


Figure 8. Pulse image of a single cycle

H1 is the main wave peak value, which reflects the expansion of the arterial wall caused by the pressure of the blood flow at the moment when the heart pumps blood, and reflects the heart state during the measurement period. H3 is the peak value of the dicrotic wave, which is caused by the backflow of arterial blood hitting the tube wall. It can reflect the state and elasticity of the artery and the size of the peripheral resistance of the artery. T1 represents the systolic period, T2 represents the diastolic period, T represents the entire pulse cycle, and its reciprocal is the heart rate value. Relevant medical experiments show that when the human body is in different fatigue states, the pulse waveform and related values will change. Therefore, these time-domain values of pulse diagram are theoretically related to the physiological state of human body and can be extracted as eigenvalues. In the actual measurement, the used sensor always modifies its magnification according to the change of the pulse amplitude, which is conducive to the observation of the image, but is not conducive to the extraction of the peak amplitude. Therefore, instead of extracting the specific size of H1 and H3, the size of H3/H1 is extracted instead, which is defined as the *dicrotic coefficient C*. And there is a mathematical correlation between the values of T1 and T2 (T1=T−T2). So only extract T2 as the feature value.

The kurtosis factor and margin factor are two dimensionless features that describe the overall trend of pulse in time domain, which can reflect the trend of pulse pattern, that is, whether the pulse pattern tends to be flat or gradually steep within a period. The experimental results show that the kurtosis factor and margin factor of the image have obvious changes in the data of different states. For example, in the data collected under some fatigue states, the amplitude of the heavy beat wave is small and the period is narrowed, which leads to the flattening of the whole image and changes in kurtosis and margin. Therefore, the correlation between the characteristic values and the state of human body is analyzed.

In summary, the extracted time-domain feature values are: heart rate (the inverse of pulse period T), dicrotic wave period T2, period ratio T2/T, dicrotic coefficient C, kurtosis factor, and margin factor.

When extracting, first use the *findpeaks* function to find the peak and valley points in the pulse data image, divide the pulse image into the main band and the dicrotic band, and then calculate the indicators as needed according to the formula. The *findpeaks* function is a function for finding extreme points in the MATLAB library function. It can collect the peak and valley point coordinates of the function image, and save the peak points in the matrix and output them for recall. Use this function to segment the time domain image, and then obtain the four characteristic values of heart rate, dicrotic wave period T2, period ratio T2/T, and dicrotic coefficient C. Its use form is as follows:

$$\left[maxl, minl\right] = findpeaks\left(y, 'minpeakdistance', d\right)$$

Where *y* is the desired pulse data image, *d* is the minimum distance between the two peaks, and *[maxv, maxl]* is the matrix that saves the extracted peak point coordinates.

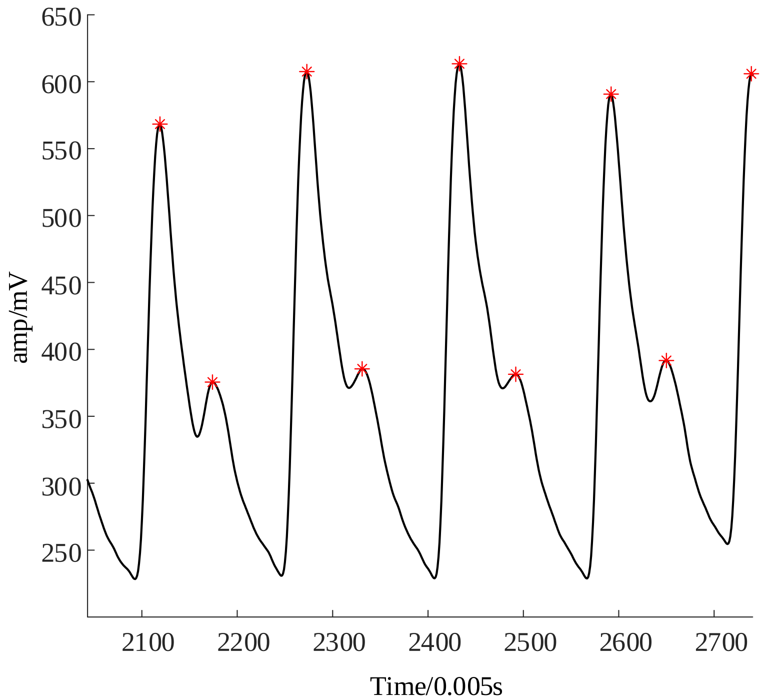


Figure 9. Peak and valley values extracted using the findpeaks function

The calculation methods of the other two eigenvalue margin factors and kurtosis factors are as follows:

$$K=\frac{\mu}{\sigma}=\frac{\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^4}{\sigma}-3$$

$$C_e=\frac{x_{peak}}{x_r}=\frac{x_{peak}}{\left(\frac{1}{n}\sum_{i=1}^{n}\sqrt{|x_i|}\right)^2}$$

In the formula, K is the sheath factor , and Ce is the margin factor

3.2 FREQUENCY DOMAIN CHARACTERISTICS

The pulse signal is a periodic signal, and the frequency domain map can be obtained by Fourier transform as shown in Figure 10. Since each harmonic of the pulse is closely related to the physiological state of multiple organs of the human body [10], the corresponding characteristic value can be extracted and applied to the evaluation and detection of the physiological state of the human body. In this paper, the first three resonant peak amplitudes, which are more obvious, are extracted as characteristic values.

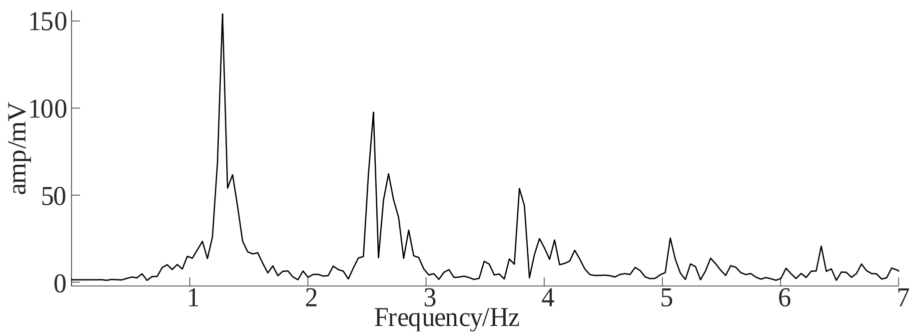


Figure 10. Pulse signal frequency domain diagram

In addition, some scholars have studied the frequency domain energy distribution of pulse. The experiment of Qiu Jian et al. [11] pointed out that "whether the energy of the pulse signal spectrum above 5Hz accounts for more than 1%" can be used as a basis for fatigue determination. Liu Fenghua's research [12] showed that the energy ratios of 7.8Hz-23.4Hz and 23.4Hz-31.25Hz have significant changes in sleepiness and awake states. Based on the above research, this paper also selects the proportion of the spectral energy in these three ranges as the eigenvalues, and uses the wavelet packet decomposition method to extract.

First, the pulse data is decomposed by 7-layer wavelet packet to obtain 128 frequency bands, and the wavelet packet diagram is drawn to check the correctness of the program. Then determine the frequency band to be extracted. Since the sampling frequency of the instrument used is 200Hz, according to the Nyquist sampling theorem, the frequency of the pulse signal in the center of the frequency domain is 0-100Hz. It can be calculated that each frequency band contains 100/128=0.78125Hz. According to the serial number assigned by MATLAB to each frequency band decomposed, it can be known that the corresponding calculation method of each eigenvalue is shown in Table 1. Finally, calculate the proportion of energy to obtain the characteristic value.

Table 1. Corresponding calculation table of spectral characteristic value

Spectral eigenvalue	Energy ratio calculation
0-4.6875Hz	$\sum_{0}^{6} E(X)$
7.8Hz-23.4Hz	$\sum_{10}^{30} E(X)$
23.4Hz-31.25Hz	$\sum_{30}^{40} E(X)$

In summary, the feature values extracted in this article are:

Table 2. Characteristic value summary table

	The serial number	Name of eigenvalue
Time Domain	1	The pulse cycle
	2	Heart rate
	3	Dicrotic wave period
	4	Two-period ratio
	5	C (Dicrotic coefficient)
	6	Kurtosis factor
	7	Margin factor
Frequency domain	8	The first peak in the frequency domain
	9	The second peak of the frequency domain

10	The third peak in the frequency domain
11	0-4.6875Hz energy ratio
12	7.8Hz-23.4Hz energy ratio
13	23.4Hz-31.25Hz energy ratio

3.3 SEPARABILITY ANALYSIS OF EIGENVALUES BASED ON STUDENT'S T TEST

After the feature values are extracted, their separability must be analyzed. Only when the feature values under different labels are significantly different from each other, can the extracted features be proved to be effective, and different subjective labels can be distinguished, otherwise it means that they cannot be used to distinguish the physiological state and the features are invalid. This article uses the Student's T-Test to analyze whether the eigenvalues under different subjective labels are different.

The student T-Test method, referred to as the T-Test method, was first proposed by William Gossell in 1908. It is a statistical method that can determine whether there is a significant difference between the data of two samples. Because the T-Test method has better accuracy when the sample size is not large, and is easy to implement, this method is selected for research in this article. The MATLAB implementation method is as follows:

$$\left[h, p, ci\right] = ttest\left(data1, data2, \alpha\right)$$

Among them, *data1* and *data2* are the two sets of input data that need to be compared, and the number of data contained must be the same; the value of α is set to control the confidence; P is the calculated value. In this paper, α is set to 0.05. At this time, if the calculated result is $P<0.05$, it proves that there is a 95% certainty that the two sets of data are different; if $P<0.01$, it proves that there is a 95% certainty that the two sets of data are significantly different.

The separability analysis is mainly divided into two steps: (1) Group "fatigue" and "deep fatigue" in the subjective label into one group; group "easy" and "very easy" into one group, and carry out T-Test analysis to explore whether the feature value can distinguish the two types of labels. (2) Carry out a detailed study, that is, whether there is a significant difference between the data labeled "fatigue" and the data labeled "deep fatigue"; whether there is a significant difference between the "relaxed" data and the "very easy" data. In summary, the three T-Test analysis is performed on each group of features, and the process is shown in the following figure:

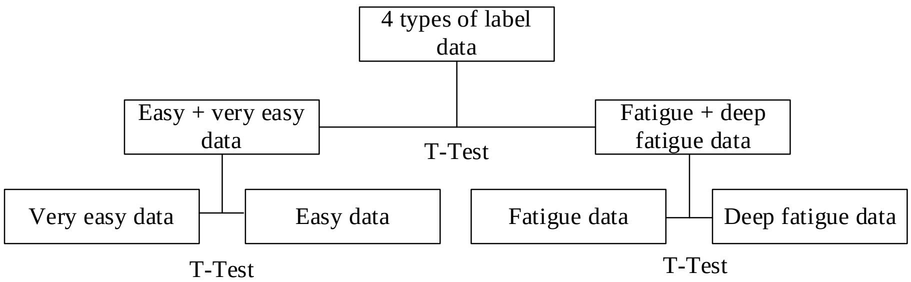


Figure 11. T-Test process of the reserach

The calculation results of P for the first fatigue-easy category of each time domain feature are as follows:

Table 3. Fatigue-easy category t-test results of each time domain eigenvalue

Time domain feature name	P
Heart rate	$4.86*10^{-7}$

Dicrotic wave period	0.004
Ratio of two periods	0.0042
Ratio of two crests	0.9953
Margin factor	$5.71*10^{-7}$
Kurtosis factor	0.0434

It can be seen from the table that, with the exception of "the ratio of the two crests", they are 95% confident that the data of each feature under different categories of labels are different, and the difference is significant. Therefore, the "two crest ratio" is eliminated, and other features are retained for further judgment.

After that, the T-Test of the refined labels was performed on the time domain features, and the T-Test results of the relaxed-very-easy label and the fatigue-deep fatigue label were obtained respectively, as shown in the following two tables. It can be seen that the "kurtosis factor" is not good for the separability of the two types of more refined labels, and the P value exceeds 0.05. Therefore, in the subsequent classification algorithm design, it should be decided whether to exclude it according to the actual situation. The calculation results of other characteristic P values are all less than 0.05, which proves that they are separable for the four types of labels and the characteristics are effective.

Table 4. Ease of each time domain eigenvalue-very easy label t-test results

Time domain feature name	P
Heart rate	$5.4574*10^{-4}$
Dicrotic wave period	0.0113
Ratio of two periods	0.0020
Ratio of two crests	0.0269
Margin factor	0.5508

Like the time domain feature analysis, the T-Test method is used to judge the correlation between each frequency domain feature and the physiological state. The results of the first T-Test for major categories of tags are as shown in Table 5.

It can be seen from the table that the P values of all features are less than 0.05, which proves that there is 95% confidence that they have differences in the data under different categories of labels, and the P values of four features are less than 0.01, the difference is very significant.

Table 5. Fatigue-depth fatigue label t-test method results of each time domain eigenvalue

Time domain feature name	P
Heart rate	$1.4*10^{-4}$
Dicrotic wave period	0.0013
Ratio of two periods	0.0317

Ratio of two crests	0.0301
Margin factor	0.8944

After that, the T-Test of the refined label was also performed on the frequency domain features, and the T-Test results of the easy-very easy label and the fatigue-deep fatigue label were obtained respectively, as shown in Table 6 and Table 7. It can be seen that the "third wave peak in the frequency domain" is not good for the separability of the two types of refined tags for easy and very easy, and the P value exceeds 0.05, which is 0.0666; and "23.4Hz-31.25Hz energy ratio "The separability of the two types of refined labels for fatigue and deep fatigue is not good, and the P value reaches 0.3709. Therefore, in the design of subsequent classification algorithms, it is necessary to decide whether to exclude these two types of labels according to the actual situation. The calculation results of other characteristic P values are all less than 0.05, which proves that it is separable for the four types of labels and the characteristics are effective.

Table 6. Fatigue of each frequency domain eigenvalue-easy label's t-test method

Frequency domain feature name	P
The first peak in the frequency domain	$3.11*10^{-8}$
The second peak of the frequency domain	$2.97*10^{-6}$
The third peak in the frequency domain	0.0419
0-4.6875Hz energy ratio	$1.39*10^{-7}$
7.8Hz-23.4Hz energy ratio	$2.97*10^{-6}$
23.4Hz-31.25Hz energy ratio	0.0488

Table 7. Ease of each frequency domain feature value-very easy label t-test results

Frequency domain feature name	P
The first peak in the frequency domain	0.0164
The second peak	0.0037
The third peak	0.0666
0-4.6875Hz energy ratio	$1.3492*10^{-5}$
7.8Hz-23.4Hz energy ratio	0.0201
23.4Hz-31.25Hz energy ratio	0.04688

4. DESIGN OF CLASSIFIER BASED ON SUPPORT VECTOR MACHINE ALGORITHM

The classification algorithm adopted in this paper is Support Vector Machine (SVM), which is a commonly used classification method. Its central idea is to form a multidimensional space with each eigenvalue as the axis, and then establish a classification hyperplane as the decision surface, so that the isolation edge between the positive example and the negative example can be maximized. Finally, the purpose of distinguishing different kinds of data based on feature decision surface is achieved [13-14]. LIBSVM (A Library for Support Vector Machines) is A set of Support Vector machine libraries developed by Professor Lin Chih-jen in 2001 [15].In this paper, functions in the library are used to construct the SVM classifier on MATLAB. The four groups of eigenvalues corresponding to the input data and the labels obtained from clustering are input for training, and the classification accuracy is judged by using ten-fold cross validation.

Table 8. Fatigue-depth fatigue label t-test method results of each frequency domain eigenvalue

Frequency domain feature	P
The first peak in the frequency domain	$8.155*10^{-4}$
The second peak of the frequency domain	$2.9152*10^{-5}$
The third peak in the frequency domain	0.0027
0-4.6875Hz energy ratio	$2.4816*10^{-5}$
7.8Hz-23.4Hz energy ratio	0.0313
23.4Hz-31.25Hz energy ratio	0.3709

4.1 Principle of support vector machines

Support Vector Machine [14](SVM for short) is a classification method based on statistical learning theory proposed by the research group of AT&Tbell Laboratory led by Vapnik in 1995.SVM using structural risk minimization principle of statistical learning theory and the theory of VC dimension and space by kernel function input samples from the original nonlinear mapped to high-dimensional feature space, and in this structure, the optimal separating hyperplane in the high dimensional feature space, makes the SVM has high fitting precision, strong learning ability, short training time, select few parameters, good generalization ability, marketing ability and whole The local best advantage provides an effective tool for solving small sample, high dimension and nonlinearity problems, and can be generalized to other machine learning problems such as function fitting.

LibSVM is a SVM pattern recognition and regression software package developed and designed by Professor Lin Chih-jen of Taiwan University. It is simple, easy to use and fast and effective. It not only provides the compiled executable files which can be used in the Windows series system, but also provides the source code. Easy to improve, modify and apply on other operating systems; The software adjusts the parameters of SVM relatively little, provides a lot of default parameters, using these default parameters can solve many problems; Cross Validation is also provided. The software can solve the problems of C-SVM, v-SVM, ε-SVR and v-SVR, including multi-class pattern recognition based on one-to-one algorithm.

4.2 Design of classifier based on support vector machine

Based on the above package, we built a classifier model, which can realize multiple classifications by dividing hyperplanes. The flowchart of the classifier model constructed using this algorithm is shown in the figure below.

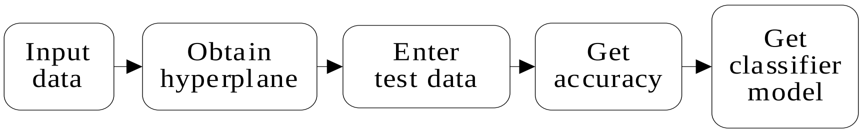


Figure 12. Flow chart of support vector machine algorithm

Then we built a basic classifier according to the LibSVM software package. First, we need to set the training data set and test data set. Then, we call the svmtrain function in the program package for model training, and then use the svm-predict function to classify the test set to get the accuracy.

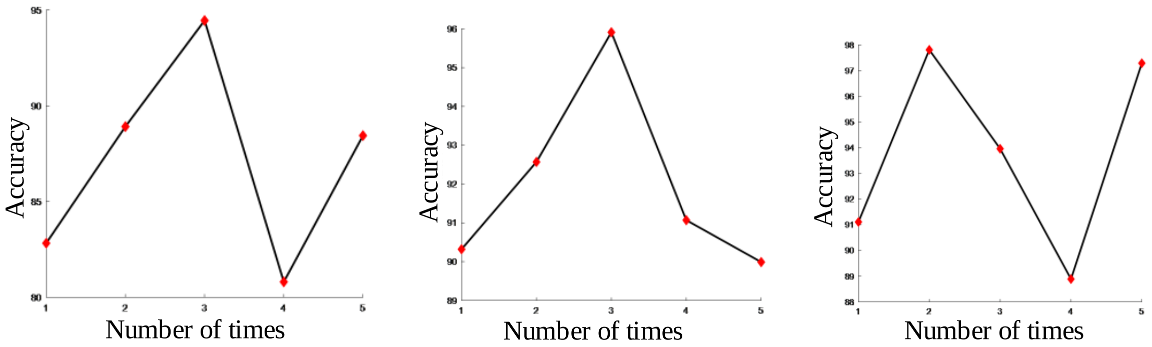
4.3 Support vector machine classifier test

In this paper, three sets of UCI data sets were used to verify the classifier constructed above based on support vector machine. They were Iris, Wine, and Breast-cancer-Wisconsin data sets, and the degree of conformity between the clustering tag and the actual tag of the clustering algorithm was taken as the evaluation index. The parameters of these three datasets are shown in Table 9.

Table 9. UCI data set parameters

data sets	Number of samples	Number of attributes	Number of categories
Iris	150	4	3
Wine	178	13	3
Breast-cancer	6	9	2

We performed 5-fold cross-validation on the above UCI dataset, and the validation results of the above dataset are shown in Figure (a), (b) and (c) below



(a) Iris (b) Wine (c) Breast-cancer-wisconsin

Figure 13. Accuracy of three UCI datasets

According to the 50%-fold cross validation accuracy of the three UCI data sets, we can get Table 10.

Table 10. UCI data set validation results

Data set	Iris	Wine	Breast-cancer-wisconsin
Accuracy	87.09%	91.97%	93.81%

The above results show that the classification apparatus based on support vector machine constructed in this study has high accuracy and can obtain better classification results.

5. CLASSIFICATION RESULTS AND ANALYSIS OF PULSE SAMPLES

5.1 Pulse sample classification results

In this paper, the 12 groups of eigenvalues previously extracted were normalized, and then the pulse samples of each volunteer were classified. The classification accuracy of four volunteers with better experimental results was shown in the following table. Only 50%-fold cross validation is performed in this paper.

Table 11. The classification accuracy of four volunteers based on the selected eigenvalue samples

Number	First	Second	Third	Forth	Fifth	Average
1	87.08%	91.54%	95.58%	86.32%	93.39%	90.78%
2	90.37%	87.69%	93.61%	92.95%	92.66%	91.46%
3	93.23%	94.01%	94.27%	93.67%	86.78%	92.39%
4	85.90%	93.40%	92.34%	89.03%	89.75%	90.08%

5.2 Analysis of the results of pulse classification

It can be seen from Table 11 that the accuracy of the physiological status assessment and detection device we used in this paper is mostly above 90%, with an average accuracy of 91.18%. Combined with the UCI data set in the previous chapter, the verification results are good, and the physiological status assessment system constructed in this paper has achieved good results.

This paper further explores the fluctuation of the characteristics of each dimension with the change of physiological state level. By analyzing the classification results of individual eigenvalue samples, it can be seen that the fluctuation of the eigenvalue of "7.8-23.4Hz energy proportion" is in high agreement with the results of literature [12] and literature [16], that is, with the increase of fatigue degree, the value of the eigenvalue tends to decay. The eigenvalue of "0-4.6875Hz energy ratio" is contrary to the experimental conclusion in literature [11]. The more drowsiness and fatigue the test subjects feel, the higher the energy ratio of this frequency band will be. After analysis in this paper, it is found that the energy value of the tester in this frequency band decreases during fatigue, but the energy decrease in other frequency bands is more significant, so the energy proportion in this frequency band increases instead. Therefore, the conclusion of the energy value decrease is consistent with previous studies.

Several groups of characteristic value also have related trends: one cardiac cycle, first peak value, pulse frequency domain image kurtosis, 23.4 Hz to 31.25 Hz energy accounted for the four eigenvalues with the degree of fatigue is reduced, the pulse image margin, heart rate, pulse cycle and frequency domain the third peak value four eigenvalues appeared opposite trend. However, the three characteristic values of the ratio of two peaks, the ratio of two periods and the second peak value in frequency domain have not been found to have clear classification boundaries under different physiological states, which may be because the data is not extensive or the correlation weight between the characteristic value and the physiological state is relatively low. The overall trend is shown in the table below.

Table 12. Summary of the trend of eigenvalues

Decrease as fatigue deepens	Increases with the deepening of fatigue	Almost unchanged with changes in fatigue
Dicrotic wave period	Pulse image margin	Two-period ratio
The first peak in the frequency domain	Pulse cycle	The second peak of the frequency domain
Pulse image kurtosis	Heart rate	
23.4Hz-31.25Hz energy ratio	The third peak in the frequency domain	

7.8Hz-23.4Hz energy ratio	0-4.6875Hz energy ratio

6. CONCLUSION

With the increasing demand for mental workers in the society, it is very important to optimize the model of state assessment and detection device and establish a reasonable state classification model. In this paper, the pulse signals of a large number of volunteers mainly engaged in mental work were collected and recorded with the method of short period and high frequency in combination with the evaluation index of human physiological state, and the classification method was used to complete the classification of human physiological state and the fluctuation analysis of related characteristics. The purpose of this paper is to conduct personalized physiological state grading research, in order to reduce the impact of individual differences on the state assessment and detection device, and optimize the physiological state assessment system based on human physiological signals.

In this paper, 12 groups of eigenvalues were obtained through extraction and screening, and a physiological state evaluation system based on support vector machine was proposed. The classification of physiological state based on human physiological signals was refined through three steps: feature extraction, correlation test calculation and classifier design. For different individual pulse signal analysis aims to study the law of based on the physiological state of the pulse signal level, to explore the pulse signal with the physiological state level change between different individuals, so as to further optimize the human physiological state criterion system, based on the physiological state of the pulse signal analysis evaluation system to provide a new research idea.

REFERENCES

[1] Liang Zhaode. Research on Driver Fatigue Detection Algorithm Based on Convolutional Neural Network [D]. South China University of Technology, 2018.

[2] Xu Lisheng, Zhang Wenxu, Pang Yuxuan, etc. Fatigue driving detection algorithm based on short-term ECG signal [J]. Journal of Northeastern University (Natural Science Edition), 2019, 40(7): 937-941.

[3] Fu Menglong. Research on real-time online muscle fatigue estimation method based on surface EMG signal[D]. University of Chinese Academy of Sciences (Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences), 2020.

[4] Patel M, Lala SKL, Kavanagha D, et al. Applying Neural Network Analysis on Heart Rate Variability Data to Assess Driver Fatigue[J]. Expert Systems with Applications,2011,38(6):7235-7242.

[5] Li Yongping. Recognition of mental fatigue state by pulse imaging detection method [D]. Lanzhou University of Technology, 2008.

[6] Guo Yanjie. Research on Pulse Signal Analysis Based on Recurrent Neural Network [D]. Beijing University of Posts and Telecommunications, 2018.

[7] Xie Zhi. Research on driving fatigue model based on physiological signals [D]. Soochow University, 2017.

[8] Zhang Zuhuai. Research methods and applications of driving fatigue based on human physiological signals [D]. Harbin Institute of Technology, 2006.

[9] Li Sinan. Research on Human Physiological State Discrimination Technology Based on Multi-sensor [D]. Northeastern University, 2014.

[10] He Shasha. Research on Feature Extraction and Selection of Sub-health Recognition Based on Pulse Wave [D]. Xi'an University of Science and Technology, 2020.

[11] Qiu Jian, Liang Shuai, Han Peng, et al. Design of wearable bracelet for monitoring fatigue driving based on pulse spectrum analysis [A]. China Association for Science and Technology, Guangdong Provincial People’s Government. The 17th China Association for Science and Technology Annual Conference ——Sub 3 Proceedings of the Academic Conference on Advanced Measurement and Control Technology for Intelligent Manufacturing [C]. Chinese Association for Science and Technology, Guangdong Provincial People's Government: Academic Department of the Chinese Association for Science and Technology, 2015: 6.

[12] Liu Fenghua. Research on drowsiness detection and recognition method based on ECG pulse signal [D]. Lanzhou University of Technology, 2013.

[13] Xu Xiaoming. SVM Parameter Optimization and Its Application in Classification [D]. Dalian Maritime University,2014.

[14] Zhou Zhihua. Machine Learning [M]. Beijing: Tsinghua University Press :2015-06.

[15] Huang He. Research on Real-time Sports Fatigue Monitoring System [D]. University of Electronic Science and Technology of China,2018.

[16] Liu Qian. Research on drowsiness detection method based on ECG pulse signal [D]. Lanzhou University of Technology, 2012.