A Framework for Cognitive Load Recognition Based on Machine Learning and Multimodal Physiological Signals by Wearable Sensors

1st ChenWei Wang
School of Computer and Information
Hefei University of Technology
Hefei, China
ChenWeiWang@mail.hfut.edu.cn

2nd XinYu Li
School of Computer and Information
Hefei University of Technology
Hefei, China
xinyuli@mail.hfut.edu.cn

3rd JinYang Huang
School of Computer and Information
Hefei University of Technology
Hefei, China
hjy@hfut.edu.cn

4th Peng Zhao
School of Computer and Information
Hefei University of Technology
Hefei, China
ZhaoPengWork@mail.hfut.edu.cn

5th Meng Wang
School of Computer and Information
Hefei University of Technology
Hefei, China
mengw512@mail.hfut.edu.cn

6th GuoHang Zhuang
School of Computer and Information
Hefei University of Technology
Hefei, China
guohang_zhuang@mail.hfut.edu.cn

7th Xiao Sun
School of Computer and Information
Hefei University of Technology
Hefei, China
sunx@hfut.edu.cn

***Abstract*—Cognitive load recognition is of great significance to the study of emotional health, as excessive cognitive load not only impairs work and learning efficiency but also leads to mental pressure and fatigue, affecting overall well-being. Advances in wearable technology and understanding the role of physiological signals in assessing human health have enabled the possibility of cognitive load recognition using these signals. This work proposes a framework that combines multiple physiological signals (ECG, EDA, and PPG) collected through wearable sensors for cognitive load recognition. Heart rate variability (HRV) features are extracted from ECG signals, while statistical characteristics are extracted from EDA and PPG signals. The RFECV feature selection algorithm is utilized in the feature selection module to achieve optimal feature subset selection and redundant feature removal. The classification of cognitive load states is implemented based on the ASAPSO-LightGBM(Adaptive Simulated Annealing Particle Swarm Optimization-Light Gradient Boosting Machine) machine learning model. We validated the effectiveness of the framework on two publicly available datasets (MAUS and WESAD). Experimental comparisons of feature combinations and machine learning models demonstrate that the proposed framework achieves superior performance, with recognition accuracy of 98.52% and 88.38% on the two datasets, respectively.**

***Index Terms*—cognitive load recognition, multimodal physiological signals, feature selection, machine learning, parameter optimization**

I. Introduction

Mental health is an essential component of physical health, and mental health disorders can affect our mood, thinking, and behavior. Cognitive load has been an increasingly important theoretical concept in human mental health research in recent years. The cognitive load includes mental load and mental effort. The mental load is imposed based on the amount of information provided, while mental effort is the mental capacity that must be assigned to the information [1]. Humans have a limited cognitive load capacity for work and learning. It has been shown that learning ability is impaired when a task exceeds the cognitive load capacity of the learner [2] [3]. The common cognitive load measures currently available include subjective measures, task performance measures, and physiological signal measures. Subjective measures are susceptible to subjective factors, and task performance may be influenced by multiple factors, such as cognitive ability and intelligence level, that cannot visually reflect changes in cognitive load. Physiological signals have received much attention in recent years because they can visually reflect the real-time state of the human body in response to external stimuli. Physiological signal measures are able to assess a person's cognitive load, and many previous studies have confirmed these indicators as a proxy for cognitive load [4]. Most existing studies have collected evidence for the construct validity of these metrics by manipulating the difficulty and complexity of the task and have shown the high sensitivity of physiological metrics in detecting individual differences in cognitive load [5].

The proliferation of wearable sensors and devices has dramatically facilitated the acquisition and analysis of physiological signals. With tiny, accurate, and low-cost sensors, it can be used to track people's activities and collect biometric signals in real-time, providing sufficient data for task analysis.

In addition, machine learning methods with excellent data processing and analysis capabilities can help analyze cognitive load situations from physiological signals quickly and provide reliable results. All these provide suitable conditions for cognitive load measurement based on physiological signals. However, the critical question is how to combine physiological signal features with machine learning to achieve a good cognitive load recognition effect.

The common physiological signals used to analyze human mental states mainly include electrodermal activity (EDA), electrocardiogram (ECG), photoplethysmograph(PPG), respiration (RSP), electroencephalogram (EEG), and electromyogram (EMG). These signals can provide much information about the individual's mental state, such as anxiety, depression, and fatigue. In [6], the authors used information on heart rate variability (HRV) extracted from ECG signals to analyze the fatigue and non-fatigue states of orchard workers before and after work. They concluded that some HRV parameters differed significantly in fatigue and non-fatigue states. The effect of surgical treatment, a stimulus, on anxiety and depression in cancer patients was confirmed in [7], where HRV parameters played an important role. Reference [8] achieved major depression detection based on EDA and support vector machine-recursive feature elimination (SVM-RFE) method with an accuracy of up to 74%. Reference [9] used multiple deep neural network models based on EDA for emotion recognition to validate the performance of each model on emotion recognition. Among them, the CNN model achieved the best performance, with an average accuracy of 86.73% in valance and 86.92% in arousal. In [10], the authors trained a one-dimensional convolutional neural network using single-pulse features extracted from PPG signals to achieve the detection of short-term emotions, and the method achieved accuracies of 75.3% and 76.2% for valence and arousal states, respectively. In addition, the combination of multiple physiological signals is also a commonly used method. Multimodal physiological signals have more joint features and therefore have the potential to provide better performance than single signals. Reference [11] used a wristband device to record data from blood volume pulses (BVP), EDA, TEMP, and ACC to train a stress detection model. In [12], ECG, RESP, EDA, and EMG data were used to detect the emotional response of the human body to music. In [13], the authors used EDA, PPG, EEG, and pupil size to assess the level of anxiety of the driver's face under different road conditions.

In this work, we propose a framework for cognitive load recognition based on multimodal physiological signal fusion. High and low cognitive load states are recognized using multimodal physiological signals of EDA, ECG, and PPG collected from wearable sensors. First, heart rate variability (HRV) is estimated from the pre-processed ECG signal, from which time domain and frequency domain features and nonlinear features are extracted, and statistical features are extracted from the EDA and PPG signals. After that, the selection and fusion of multimodal features are implemented based on the RFECV algorithm to remove redundant features from the high-dimensional features. Finally, the LightGBM machine learning model and the adaptive simulated annealing particle swarm optimization algorithm are used to realize the classification and parameter optimization schemes for cognitive load state recognition, respectively. Two publicly available datasets are used for training and testing the framework. The experimental results show that the framework achieves 16.68% higher recognition accuracy than the baseline model in MAUS and 5.4% higher than that in WESAD.

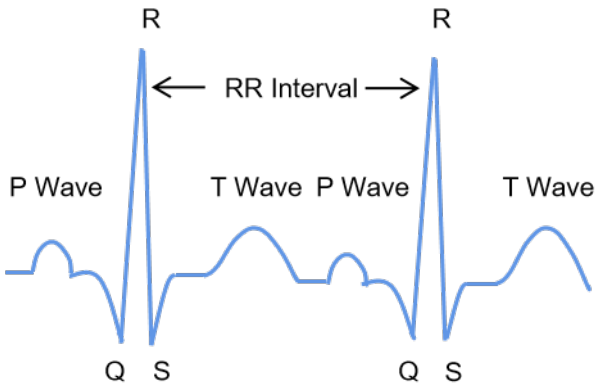


Fig. 1. Description of ECG signal components

The main contributions of this work are summarized as follows: (1) A cognitive load recognization framework based on multimodal physiological signal fusion and machine learning is proposed to achieve an accurate assessment of high and low cognitive load levels. (2) The RFECV algorithm is used to achieve the selection and fusion of high-dimensional features of physiological signals, and the extracted feature subset has the best performance in recognizing cognitive load by comparing the results with those of other feature combinations. (3) A parameter optimization method for the LightGBM model based on an adaptive simulated annealing algorithm (ASAPSO) is proposed,which can effectively improve the performance of the LightGBM model on cognitive load recognition. The experimental results show that the ASAPSO-LightGBM model performs better than the other five machine-learning models.

The rest of the paper is organized as follows: Section II discusses the features of the three signals used in this work and the dataset used for model training. Section III describes our overall approach, Section IV presents the analysis and discussion of the experimental results, and Section V concludes this work.

II. Physiological Signal Data

As described in the previous section, physiological signals have been widely used in the study of human mental state recognition. The basic principle is that fluctuations in the mental state will have an impact on the indicators of human physiological signals. We can study the transformation of the human mental state by analyzing the changes in these indicators. In this study, we focused on three signals: ECG,

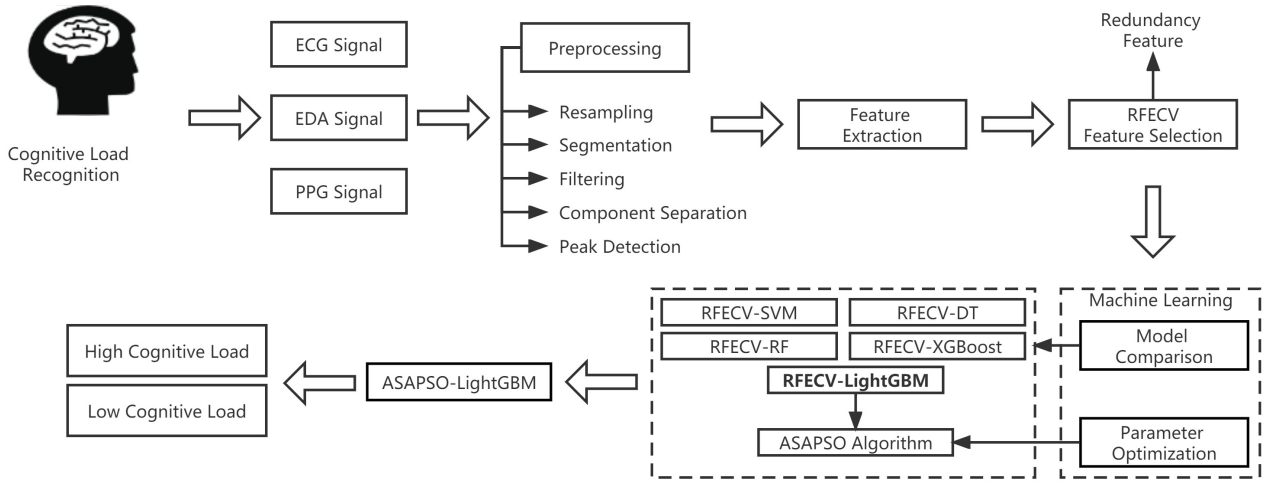


Fig. 2. Overall method flow chart

EDA, and PPG. [18].In addition, heart rate variability (HRV) can be calculated from PPG signals as well, which is why PPG signals are often combined with ECG signals in many affective computing studies. In summary, we selected three physiological signals that are closely related to changes in human mental states for the identification of human cognitive load levels.

A. Electrocardiogram, Electrodermal activity, Photoplethysmograph

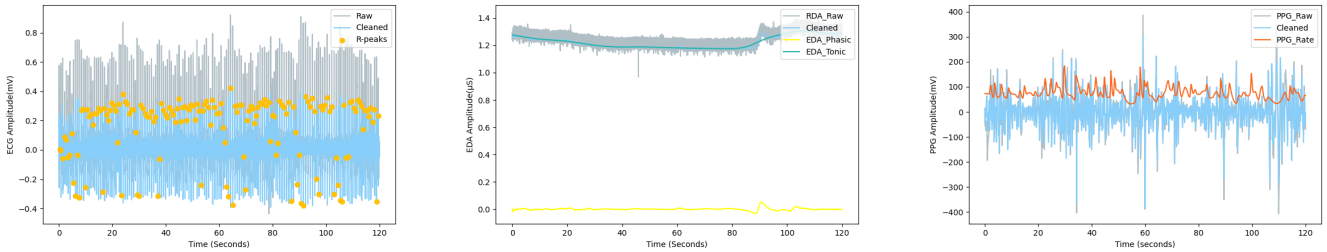
The ECG signal captures the change in cardiac potential over time, showing a quasi-periodic behavior consisting of a series of heartbeats [14]. Each beat consists of three waves, a P-wave, a QRS complex, and a T-wave, as shown in Figure 1. QRS detection can be used as a starting point for ECG feature extraction by detecting successive R-peaks (positive peaks of the QRS complex), which can extract the RR interval. Heart rate variability (HRV) reflects the variability of ECG adjacent R-wave intervals, and the RR interval itself can be used to analyze heart rate variability, which provides important information about cardiovascular behavior, which may be influenced by factors such as the health and emotional/mental state of the subject [14]. Numerous studies have shown that heart rate (HR) and heart rate variability (HRV) can distinguish different mental states. The time-frequency analysis of HRV reveals that HRV can differentiate whether a subject is in a normal or mentally fatigued state [15].

When our organism is subjected to sensory stimuli or emotional changes, it secretes large amounts of sweat, which alters skin conductance, a property known as EDA or GSR. Dermal electrical activity can respond very quickly and sensitively to the degree of impact of a stimulus event on an individual. Therefore, EDA can be used as a physiological indicator to detect mood changes [16]. The EDA signal is divided into two main components, including the Tonic Data and the Phasic Data composition. Skin conductance level (SCL) is the most commonly used method to measure the Tonic component. Related studies have shown that changes in SCL are related to an individual's level of autonomic arousal, such as the general level of emotion and stress. Because the SCL changes slowly, the measurement time needs to be long enough. Phasic Data refers to the rapidly changing component caused by sympathetic activity and is also considered as the skin conductance response (SCR). The SCR is a state of activation of physiological and psychological caused by stimuli that can be visualized as rapid fluctuations and large wave amplitudes in the electrodermal signal.

Finally, the PPG signal is a noninvasive method to detect changes in blood volume caused by cardiac activity [17].PPG has a periodic pulse consisting of systolic and diastolic phases. Some studies have shown that certain properties of the PPG signal can provide information about the level of mental stress

B. Datasets

Two publicly available datasets, MAUS [19] and WESAD [20], were used in this study. Physiological signal data from both datasets were collected using wearable sensors. In particular, the MAUS dataset recorded ECG, GSR, and PPG signals at a sampling rate of 256 Hz, and data were recorded from a total of 22 participants. The MAUS dataset used a series of N-Back tasks (0-Back, 2-Back, and 3-Back) to elicit different cognitive load levels and statistically validated the difference between 0-Back and 2-Back and 3-Back induced significant differences in cognitive load levels. Therefore, we assigned a low cognitive load label to the data during the 0-Back task and a high cognitive load label to the data during the 2-Back and 3-Back tasks. The WESAD dataset was designed for the detection of cognitive stress and emotion, with multiple physiological signal information collected via the wrist-worn device and the chest-worn device. Seventeen participants were



(a) ECG signals before and after pre-processing (b) EDA signals before and after pre-processing (c) PPG signals before and after pre-processing

Fig. 3. The physiological signals before and after pre-processing of the WESAD dataset

asked to read magazines, give public speeches, and count numbers to induce different cognitive stressors. In this study, we obtained data labels by treating the baseline phase of reading magazines as a low cognitive load phase and the phase of making speeches and counting numbers as a high cognitive load phase.

III. Methodology

The proposed framework includes the processing steps detailed in the following sections. The overall method flowchart is shown in Figure 2. Experimental data are obtained from publicly available multimodal physiological signal datasets, from which we receive raw ECG, EDA, and PPG signal recordings. A pre-processing operation is performed on the raw data to obtain physiological signal time segments that can be used for feature extraction. The feature extraction method will then be used to extract these physiological signals to get signal features initially. The feature selection module further eliminates redundant features. Finally, the selected results will be fed into a machine learning model for training. We compared different machine learning algorithms (SVM, DT, RF, XGBoost, LightGBM), from which the best performing model was selected. At the same time, parameter optimization was performed to obtain the final solution to achieve an accurate classification of cognitive load levels.

A. Pre-processing

a) Upsampling: The ppg signal in WESAD comes from a wrist-worn device with a sampling frequency of 64hz, while a chest-worn device sampling collects the ECG and EDA signals at a frequency of 700hz. This results in the ppg signal having a shorter time series length than the ECG and EDA signals. Therefore, before performing all the pre-processing operations, we implemented the upsampling process of the ppg signal using the scipy library of Python.

b) Segmentation: Segmentation is an essential step in data processing. Raw physiological signal data were collected during prolonged activity. After the signal is segmented, the processing efficiency will be higher. At the same time, dividing the signal into shorter segments can increase the data volume to achieve better results in training machine learning models. In our study, we applied a two-minute sliding window to split all signals into segments of two minutes in length.

c) Filtering and processing: The original ECG signal was filtered with a Butterworth bandpass filter to remove interference such as muscle noise, power line noise, and baseline drift. The passband frequency was set to 5-15 Hz. After that, the R-peaks were obtained using the Pan-Tompkins (PT) peak detection algorithm [21] [22], which will be further used for the analysis of HRV. The EDA signal was filtered using a Butterworth low-pass filter with a cutoff frequency of 3 HZ, which can remove most motion artifact noise. In order to extract richer EDA features, the EDA signal is next component separated using the cvxEDA model [23] to distinguish the SCL and SCR components. As for the processing of the PPG signal, we used a bandpass filter of 0.5-8hz to process it, while the Elgendi algorithm [24] was used afterward to detect the peak systolic pressure to get the PPG rate. The statistical features will be calculated from the obtained PPG rate. Taking the WESAD dataset as an example, Figures 3 show the physiological signals before and after pre-processing.

B. Feature extraction

With the preprocessing step, we have the conditions to derive the HRV from the ECG signal. Next, we performed a time-frequency domain analysis and a nonlinear analysis of the HRV. Seven time-domain features were extracted. Three frequency-domain features (LF, HF, TF) and three relative frequency-domain features (LFn, HFn, LF/HF) were extracted using band power of different frequency ranges. In addition, three nonlinear features were also incorporated into our study. The time and frequency domain analysis metrics of HRV can quantitatively reflect the heart rate variability at different time scales, while the nonlinear analysis metrics of HRV can quantitatively reflect the structure and complexity of the R-R intervals (RRI) [25]. The HRV feature parameters and definitions are shown in Tables I-III. As for the EDA signal, we extracted statistical features (maximum, minimum, mean, standard deviation) and features related to SCL/SCR (number of SCR occurrences, mean amplitude of SCR peak occurrences, SCR-mean, SCR-standard deviation, SCL-mean, SCL-standard deviation) according to the method given in [20]. Since the PPG signal is highly correlated with the ECG signal, we have extracted the HRV features of the ECG signal earlier, so only the statistical features (maximum, minimum,

TABLE I
THE DEFINITION OF PARAMETER INDICES FOR TIME-DOMAIN ANALYSIS OF HRV

Parameter Indices	Definition
AVNN	The mean of the R-R intervals.
SDNN	The standard deviation of the RR intervals.
RMSSD	The square root of the mean of the squared successive differences between adjacent RR intervals.
pNN50	The proportion of RR intervals greater than 50ms, out of the total number of RR intervals.
SDSD	The standard deviation of the successive differences between RR intervals.
TINN	The baseline width of the RR intervals distribution obtained by triangular interpolation, where the error of least squares determines the triangle.
HTI	The HRV triangular index, measuring the total number of RR intervals divided by the height of the RR intervals histogram.

TABLE II
THE DEFINITION OF PARAMETER INDICES FOR FREQUENCY-DOMAIN ANALYSIS OF HRV

Parameter Indices	Definition	Frequency band
LF	The spectral power of low frequencies.	0.04-0.15Hz
HF	The spectral power of high frequencies.	0.15-0.4Hz
TF	The total power.	0-0.4Hz
LF/HF	The ratio obtained by dividing the low-frequency power by the high-frequency power.	/
LFn	The normalized low frequency, obtained by dividing the low-frequency power by the total power.	/
HFn	The normalized high frequency, obtained by dividing the low-frequency power by the total power.	/

mean, standard deviation, median) of the PPG signal were extracted here.

C. Feature selection

In the feature extraction module, we extracted 31-dimensional physiological signal features. High-dimensional attributes not only increase the overhead of model training but also some redundant features can negatively affect the classification accuracy. Therefore, we utilize the Recursive Feature Elimination and Cross Validation (RFECV) algorithm to perform feature filtering to obtain the best feature subset.

RFECV is a heuristic feature selection method that adds a cross-validation step to the RFE algorithm. The first step is recursive feature elimination. All features in the complete feature set are fed into the model for training, the importance of each feature is obtained and ranked, and the feature set is updated by removing the least important features from the current set of features. Then, a new round of training learning and rating calculation is performed on the feature set until it recursively completes rating the importance of all features. Then, different numbers of features are selected from the complete feature set in turn to construct feature subsets

TABLE III
THE DEFINITION OF PARAMETER INDICES FOR NONLINEAR ANALYSIS OF HRV

Parameter Indices	Definition
SD1	Index of short-term HRV changes.
SD2	Index of long-term HRV changes.
SD1/SD2	The ratio of SD1 to SD2. Describes the ratio of short-term to long term variations in HRV.

TABLE IV
OPTIMIZED PARAMETERS AND THEIR VALUE RANGES

Parameter	Parameter Meaning	Value Ranges
learning_rate	Learning rate for model training.	[0.001,0.2]
n_estimators	Number of iterations for model training.	[100,1200]
max_depth	The maximum depth of the tree model.	[3,12]
num_leaves	The number of leaf nodes in a tree.	[1,2^{max_depth})
feature_fraction	The feature sampling ratio.	(0.5,1]
bagging_fraction	The data sampling ratio.	(0.5,1]
lambda_L1	L1 regularization parameter.	[1e-5,3)
lambda_L2	L2 regularization parameter.	[1e-5,3)

according to the feature importance ratings eliminated by the recursive features.Cross-validation is performed on each feature subset separately to obtain the average performance score of each feature subset. Finally, the best feature subset and the number of features are determined based on the feature subset with the highest cross-validation accuracy.

D. Machine learning and parameter optimization

The filtered optimal feature subset is fed into the machine learning module for training. The proposed ASAPSO-LightGBM model was compared with five other machine learning models (LightGBM, XGBoost, RF, DT, and SVM) To validate the classification effect. We trained and tested the models using five-fold cross-validation, and the final evaluation results were averaged.

a) LightGBM: GBDT (Gradient Boosting Decision Tree) is an iterative decision tree algorithm whose main idea is to iteratively train weak classifiers (decision trees) to obtain the optimal model, which has the advantages of excellent training effect and less overfitting. LightGBM (Light Gradient Boosting Machine) is a framework for implementing the GBDT algorithm, which supports efficient parallel training. Compared with GDBT, which requires multiple iterations of the entire dataset during training and puts the whole dataset into memory, the LightGBM model supports parallel training and distributed processing of large amounts of data. The LightGBM saves the discretized values of features so as to occupy less memory, which solves the problem of GBDT being time-consuming when processing large amounts of data. Several decision tree models are integrated into LightGBM, and the final model is approximated by iterative training with the formula calculated

as:

$$f(x) = \sum_{t=1}^{N} T\left(x; \theta_t\right) \tag{1}$$

where $T\left(x; \theta_t\right)$ is a single decision tree, N denotes the total number of iteration rounds (number of decision trees), and θ_t is a decision tree parameter.

b) ASAPSO algorithm: After training the machine learning models, the ASAPSO (Adaptive Simulated Annealing Particle Swarm Optimization) algorithm [26] is used to optimize the performance of the LightGBM model to maximize the model's classification performance. Since the LightGBM model has a huge number of parameters, an inappropriate combination of parameters can cause the model to fall into a local optimum, which affects the model's classification accuracy and may also cause the model to be overfitted. ASAPSO can solve this problem very well. The algorithm is based on the improvement of the particle swarm optimization algorithm by introducing a simulated annealing operation in the particle search process, setting a temperature T_0 according to the initial state of the population, and decaying with a specific cooling factor μ after each iteration as follows:

$$T(k) = \begin{cases} E\left(G_{best}\right) / \log(0.2), & k = 1 \\ T(k-1)\mu, & k > 1 \end{cases} \tag{2}$$

where k is the number of iterations, G_{best} represents the current global optimal position of the particle seeking, and $E(G_{best})$ represents the optimal fitness value of the particle (the fitness value corresponds to the classification accuracy of the model in our study). In addition to maintaining the global optimal position G_{best}, each particle i has to hold its own individual optimal position $P_{best,i}$. The change of temperature guides the population to accept the current solution with a certain probability, which ensures the ability of the particle to jump out of the local optimal solution.The acceptance probability is calculated by:

$$p_i = \begin{cases} 1, & E_i(k) \geq E(G_{best}) \\ exp(-\frac{E_i(k) - E(G_{best})}{T_i}), & E_i(k) < E(G_{best}) \end{cases} \tag{3}$$

$E_i(k)$ represents the fitness value of the ith particle after the kth iteration. At the same time, according to the number of iterations, the algorithm can adaptively adjust the inertia weights w as:

$$\begin{aligned} w =& (w_{max} + w_{min})/2 + \tanh(-4 + 8\times \\ & (k_{max} - k)/k_{max})(w_{max} - w_{min})/2 \end{aligned} \tag{4}$$

and adjust learning factors $c1$ and $c2$ as:

$$c_1 = c_{1max} - k(c_{1max} - c_{1min})k_{max} \tag{5}$$

$$c_2 = c_{2min} - k(c_{2min} - c_{2max})k_{max} \tag{6}$$

the parameters with subscripts max and min represent the maximum and minimum values selected for this parameter, and the parameters are assigned according to the method given in [21]. Next, the particle will update its evolutionary speed $v_i(k+1)$ and position $x_i(k+1)$ according to the parameters given above:

$$\begin{aligned} v_i(k+1) = & wv_i(k) + c_1 r_1(P_{best,i}(k) - x_i(k)) \\ & + c_2 r_2(G_{best} - x_i(k)) \end{aligned} \tag{7}$$

$$x_i(k+1) = x_i(k) + v_i(k+1) \tag{8}$$

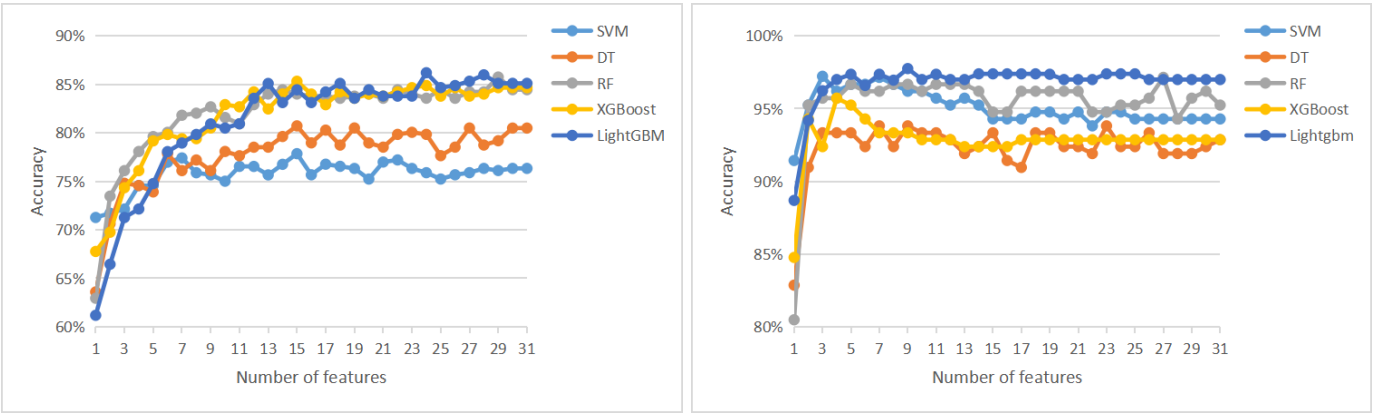
After the velocity and position of individual particles are updated, the global optimal position of the corresponding particle population may also be changed according to (3). The optimal position of the particle population will be obtained after all the iterations are completed. In this study, we use the ASAPSO algorithm to find the optimal position for the eight main parameters of LightGBM. The cross-validation accuracy of the model is used as the particle fitness value, and the final obtained optimal positions of the particles are assigned to the model parameters. The optimized parameters and their value ranges are shown in Table IV.

IV. Results and Discussion

The experimental results will be presented and discussed in this part, and the two evaluation indexes of accuracy and F1-score will be selected to evaluate the results. To verify the effect of the proposed framework, we compared the performance of the feature combination based on the three physiological signals with the optimal feature subset based on the RFECV algorithm on the cognitive classification. Meanwhile, the performance of six machine learning models was compared. Figures 4 shows the cross-validation accuracy of each model as a function of the number of features during feature selection on the MAUS and WESAD datasets. The LightGBM model with feature selection performed best on the two datasets, and the optimal number of feature subsets was 19 and 9, respectively. Tables V and VI show the experimental results of the proposed method on the two datasets. As can be seen from the tables, our method achieves 98.52% recognition accuracy on the WESAD dataset, which is the best accuracy for this study. This result is 5.4% higher than the best result achieved in the baseline. Our method also performs better than the baseline of MAUS, with 16.68% higher accuracy.

The experimental results show that the model based on full features performs better than that on the unimodal features in general but not necessarily better than the bimodal model. The complete feature set has a limited advantage over the bimodal features since it contains the largest number of redundant features. The feature subset after feature selection by the RFECV algorithm has the highest recognition accuracy, which confirms that the RFECV feature selection method can effectively reduce the feature dimensionality as well as improve the accuracy of cognitive load recognition. The LightGBM model shows the best performance among all machine learning models, and the ASAPSP algorithm further enhances this advantage, which illustrates the effectiveness of the method.

Furthermore, the table shows that the recognition accuracy of the proposed framework on WESAD is significantly better than that on MAUS. This is because the devices used to collect



(a) The RFECV algorithm was used for feature selection on the MAUS dataset(b) The RFECV algorithm was used for feature selection on the WESAD dataset

Fig. 4. Results of feature selection using the RFECV algorithm

TABLE V
COMPARISON OF EXPERIMENTAL RESULTS OF MAUS DATASET

Model		ALL[1]	ECG	EDA	PPG	ECG+EDA	ECG+PPG	EDA+PPG	SF(n_feature)[2]
SVM	Accuracy	0.7633	0.7698	0.6644	0.7763	0.7545	0.7720	0.7391	**0.7785(15)**
	F1-score	0.7629	0.7688	0.6638	0.7760	0.7544	0.7717	0.7385	**0.7776**
DT	Accuracy	0.7852	0.7633	0.7017	0.7610	0.7567	0.7566	0.7894	**0.8071(15)**
	F1-score	0.7748	0.7581	0.6960	0.7577	0.7544	0.7511	0.7682	**0.7996**
RF	Accuracy	0.8444	0.8247	0.7960	0.8225	0.8401	0.8378	0.8487	**0.8575(29)**
	F1-score	0.8434	0.8237	0.7955	0.8218	0.8395	0.8375	0.8478	**0.8567**
XGBoost	Accuracy	0.8465	0.8006	0.7698	0.8115	0.8203	0.8115	0.8487	**0.8532(15)**
	F1-score	0.8456	0.7998	0.7670	0.8108	0.8197	0.8111	0.8478	**0.8526**
LightGBM	Accuracy	0.8510	0.7962	0.7808	0.8004	0.8576	0.8203	0.8618	**0.8684(24)**
	F1-score	0.8506	0.7950	0.7799	0.8014	0.8570	0.8198	0.8615	**0.8680**
ASAPSO-LightGBM	Accuracy	0.8633	0.8043	0.8020	0.8105	0.8587	0.8238	0.8832	**0.8838(24)**
	F1-score	0.8615	0.8026	0.8009	0.8096	0.8569	0.8220	0.8813	**0.8818**
Baseline	Accuracy	**0.7170**							
	F1-score	**0.6900**							

[1] ALL: The combination of all features of ECG, EDA, PPG signals.
[2] SF: The combination of features selected by RFECV. (n_feature: the number of selected features)

TABLE VI
COMPARISON OF EXPERIMENTAL RESULTS OF WESAD DATASET

Model		ALL[1]	ECG	EDA	PPG	ECG+EDA	ECG+PPG	EDA+PPG	SF(n_feature)[2]
SVM	Accuracy	0.9429	0.9286	0.9095	0.8476	0.9524	0.9095	0.9429	**0.9722(3)**
	F1-score	0.9377	0.9224	0.9011	0.8334	0.948	0.9004	0.9373	**0.969**
DT	Accuracy	0.9238	0.8952	0.9048	0.8286	0.9286	0.9095	0.8857	**0.9381(7)**
	F1-score	0.9116	0.8969	0.8994	0.8018	0.9274	0.8997	0.8752	**0.9306**
RF	Accuracy	0.9524	0.919	0.919	0.8714	0.9524	0.9238	0.9429	**0.9714(27)**
	F1-score	0.9472	0.9123	0.9098	0.8583	0.9472	0.915	0.9363	**0.9651**
XGBoost	Accuracy	0.9386	0.9285	0.919	0.8667	0.9333	0.9333	0.9381	**0.9571(4)**
	F1-score	0.932	0.9211	0.9114	0.8521	0.9272	0.9259	0.9326	**0.9502**
LightGBM	Accuracy	0.9704	0.9556	0.963	0.8926	0.963	0.9556	0.963	**0.9775(9)**
	F1-score	0.9703	0.9536	0.9629	0.8923	0.9629	0.9552	0.9629	**0.9771**
ASAPSO-LightGBM	Accuracy	0.978	0.9632	0.9711	0.8997	0.9706	0.9633	0.9712	**0.9852(9)**
	F1-score	0.9776	0.9629	0.9708	0.8995	0.9703	0.9631	0.971	**0.9849**
Baseline	Accuracy	**0.9312**							
	F1-score	**0.9147**							

[1] ALL: The combination of all features of ECG, EDA, PPG signals.
[2] SF: The combination of features selected by RFECV. (n_feature: the number of selected features)

the WESAD signals contain wearable sensor devices, while the devices used for MAUS are wristband or fingertip devices. Since participants need to maintain interaction with the computer during the experiment, devices worn on the hand are more likely to generate artificial noise due to hand movement, even if it is a slight movement, which will affect the quality of signal acquisition and eventually lead to unsatisfactory results in the classification model. Even so, we still achieved a recognition accuracy close to 90%. In summary, the proposed framework of cognitive load recognition based on multimodal physiological signals can achieve accurate recognition of high and low cognitive loads.

V. Conclusion

In this paper, a framework for cognitive load recognition based on multimodal physiological signals is proposed. The framework consists of four modules: physiological signal preprocessing, feature extraction, feature selection based on the RFECV algorithm, and cognitive load classification based on the ASAPSO-LightGBM model. We fuse three physiological signals related to human mental states and select key features from them to form the optimal feature subset and optimize the best-performing LightGBM model using the ASAPSO algorithm. Two publicly available affective datasets, MUAS and WESAD, were used to validate the effect of the framework. The framework achieves an accuracy of up to 98.52% for the recognition of cognitive load level, outperforming the results of the two baseline works referenced in this paper.

References

[1] Paas, F. G. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: a cognitive-load approach. Journal of educational psychology, 84(4), 429.

[2] Brünken, R., Seufert, T., & Paas, F. (2010). Measuring cognitive load. Cognitive load theory, 181-202.

[3] Teigen, K. H. (1994). Yerkes-Dodson: A law for all seasons. Theory & Psychology, 4(4), 525-547.

[4] Ayres, P., Lee, J.Y., Paas, F., & van Merriënboer, J.J. (2021). The Validity of Physiological Measures to Identify Differences in Intrinsic Cognitive Load. Frontiers in Psychology, 12.

[5] Luque-Casado, A., Perales, J.C., Cárdenas, D., & Sanabria, D. (2016). Heart rate variability and cognitive processing: The autonomic response to task demands. Biological Psychology, 113, 83-90.

[6] Gao, R., Yan, H., Duan, J., Gao, Y., Cao, C., Li, L., & Guo, L. (2022). Study on the nonfatigue and fatigue states of orchard workers based on electrocardiogram signal analysis. Scientific Reports, 12(1), 4858.

[7] Kim, E. H., Park, J. H., Lee, S. M., Gwak, M. S., Kim, G. S., & Kim, M. H. (2016). Preoperative depressed mood and perioperative heart rate variability in patients with hepatic cancer. Journal of clinical anesthesia, 35, 332-338.

[8] Kim, A. Y., Jang, E. H., Kim, S., Choi, K. W., Jeon, H. J., Yu, H. Y., & Byun, S. (2018). Automatic detection of major depressive disorder using electrodermal activity. Scientific reports, 8(1), 17030.

[9] Yu D, Sun S. A Systematic Exploration of Deep Neural Networks for EDA-Based Emotion Recognition. Information. 2020; 11(4):212. https://doi.org/10.3390/info11040212

[10] Lee, M. S., Lee, Y. K., Pae, D. S., Lim, M. T., Kim, D. W., & Kang, T. K. (2019). Fast emotion recognition based on single pulse PPG signal with convolutional neural network. Applied Sciences, 9(16), 3355.

[11] Gjoreski, M., Gjoreski, H., Luštrek, M., & Gams, M. (2016). Continuous stress detection using a wrist device: in laboratory and real life. In proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing: Adjunct (pp. 1185-1193).

[12] J. Kim and E. André. 2008. Emotion recognition based on physiological changes in music listening. IEEE Transactions on Pattern Analysis and Machine Intelligence 30, 12 (2008), 2067–2083.

[13] Lee, S., Lee, T., Yang, T., Yoon, C., & Kim, S. P. (2020). Detection of drivers' anxiety invoked by driving situations using multimodal biosignals. Processes, 8(2), 155.

[14] Sarkar, P., Ross, K., Ruberto, A. J., Rodenbura, D., Hungler, P., & Etemad, A. (2019). Classification of cognitive load and expertise for adaptive simulation using deep multitask learning. In 2019 8th international conference on affective computing and intelligent interaction (ACII) (pp. 1-7). IEEE.

[15] Taelman, J., Vandeput, S., Vlemincx, E., Spaepen, A., & Van Huffel, S. (2011). Instantaneous changes in heart rate regulation due to mental load in simulated office work. European journal of applied physiology, 111, 1497-1505.

[16] Setz, C., Arnrich, B., Schumm, J., La Marca, R., Tröster, G., & Ehlert, U. (2009). Discriminating stress from cognitive load using a wearable EDA device. IEEE Transactions on information technology in biomedicine, 14(2), 410-417.

[17] Zhu, L., Spachos, P., & Gregori, S. (2022). Multimodal physiological signals and machine learning for stress detection by wearable devices. In 2022 IEEE International Symposium on Medical Measurements and Applications (MeMeA) (pp. 1-6). IEEE.

[18] Banerjee, S., Bailón, R., Lazáro, J., Marozas, V., Laguna, P., & Gil, E. (2017). A two step Gaussian modelling to assess PPG morphological variability induced by psychological stress. In 2017 Computing in Cardiology (CinC) (pp. 1-4). IEEE.

[19] Beh, W. K., & Wu, Y. H. (2021). MAUS: A dataset for mental workload assessmenton N-back task using wearable sensor. arXiv preprint arXiv:2111.02561.

[20] Schmidt, P., Reiss, A., Duerichen, R., Marberger, C., & Van Laerhoven, K. (2018, October). Introducing wesad, a multimodal dataset for wearable stress and affect detection. In Proceedings of the 20th ACM international conference on multimodal interaction (pp. 400-408).

[21] Pan, J., & Tompkins, W. J. (1985). A real-time QRS detection algorithm. IEEE transactions on biomedical engineering, (3), 230-236.

[22] Hussein, A. F., Burbano-Fernandez, M., Ramírez-Gonzalez, G., Abdulhay, E., & De Albuquerque, V. H. C. (2018). An automated remote cloud-based heart rate variability monitoring system. IEEE access, 6, 77055-77064.

[23] Greco, A., Valenza, G., Lanata, A., Scilingo, E. P., & Citi, L. (2015). cvxEDA: A convex optimization approach to electrodermal activity processing. IEEE Transactions on Biomedical Engineering, 63(4), 797-804.

[24] Elgendi, M., Norton, I., Brearley, M., Abbott, D., & Schuurmans, D. (2013). Systolic peak detection in acceleration photoplethysmograms measured from emergency responders in tropical conditions. PloS one, 8(10), e76585.

[25] Wang, C., & Guo, J. (2019). A data-driven framework for learners' cognitive load detection using ECG-PPG physiological feature fusion and XGBoost classification. Procedia computer science, 147, 338-348.

[26] Yan Qunmin, Ma Ruiqing, Ma Yongxiang. An adaptive simulated annealing particle swarm optimization algorithm [J]. Journal of xi 'an university of electronic science and technology, 2021 (04) : 13. 120-127 DOI: 10.19665 / j.i ssn1001-2400.2021.04.016.