[1]Jinglong Fang

Emotion Recognition using PPG Based on Multi-Time Window PRV Feature Selection

Journal of Electrical Systems

Abstract: - Emotion is a series of feelings triggered by specific objects or activities, influenced by various factors such as personality, mood, and motivation, and manifested in multiple forms. As a crucial part of emotion research, emotion recognition has found extensive use in various fields such as mental health monitoring, and intelligent system design. In the study of emotion recognition based on PPG, PPG have received considerable attention because of their objective nature and the ease with which they can be collected., and have been widely used in various emotion recognition systems. However, utilizing PPG signals for emotion recognition still presents several challenges. Traditional feature extraction from the entire signal can cause local features to be overshadowed by overall statistics, failing to capture critical local emotion-related features. To address these issues, this paper focuses on PPG, aiming to improve the efficiency of feature extraction and utilization in PPG emotion recognition tasks. This, in turn, broadens the potential applications of wearable devices for emotion monitoring. Experimental results demonstrate that compared to existing PPG emotion recognition methods, this method shows advantages on the DEAP public dataset.

Keywords: PPG; Emotion Recognition; Multi-Time Window; Feature Selection

1 INTRODUCTION

Emotion is a complex state produced by the human brain under external or internal stimulation. It is related to a person's cognition, personality, physiological state and other factors, and guides people's behavior and response to surrounding events[1]. Currently, Researchers use many physiological signals when developing emotion recognition systems, such as electroencephalogram (EEG)[2], electrocardiogram (ECG)[3], and Photoplethysmography (PPG)[4], Galvanic Skin Response (GSR)[5] and Electrooculogram (EOG)[6], etc.

Among the various physiological signals, EEG signals have been demonstrated to be the most precise and reliable for emotion identification[7]. However, using EEG signals for emotion recognition also has its limitations. Usually, the collection of EEG signals requires the wearing of specific equipment. Meanwhile, another signal closely related to human emotions is the PPG, which monitors variations in blood volume during human heart activity. Emotional changes can cause fluctuations in the autonomic nervous system, thereby affecting blood flow to the skin and changing the amplitude and shape of PPG[8]. Therefore, it is feasible to use PPG for emotion recognition.

In recent years, Researchers primarily explore PPG in conjunction with other physiological signals. Among the extracted features, Pulse Rate Variability (PRV) related to emotions has garnered significant attention. For example, Singh R R[9] collected PPG and skin conductance responses from subjects while driving to detect their emotional states. They computed statistical features and rhythm variations of both signals and analyzed them using a recurrent neural network. This method demonstrated a high accuracy rate of 89.23% in tests conducted on 19 drivers. Similarly, Udovičić G[10] aimed to develop a practical wearable device for emotion recognition. They collected and analyzed PPG and skin conductance responses, using standard statistical parameters and power spectral density as features. SVM and K- KNN were utilized as classifiers. The results indicated that the single-subject model

*Corresponding author: Jinglong Fang, jinglongfang01@gmail.com

[1] Department of Computer Science, College of Information Science and Technology, Jinan University, Guangzhou, 510632, China

achieved better average accuracy compared to the multi-subject model, suggesting that single-subject emotion detection models are more feasible for real-life applications. Raheel[11] derived frequency domain features from EEG signals, time domain features from skin conductance responses, and heart rate and time-frequency features from PPG. They then used a KNN classifier to recognize emotions based on each modality and their fused features. The final experimental results indicated that among the various modalities, the features based on PPG achieved the highest accuracy of 78.57%. SEONGSIL HEO[12]. integrated various lightweight denoising and peak detection algorithms, proposing a Coordinated Multiple Denoising and Peak Detection Algorithm (OMDP). They applied this algorithm to address the issue of inaccurate feature extraction due to noisy PPG with low signal quality. The final test accuracy on the WESAD dataset was 96.50%, with an F1 score of 93.36%. This result suggests an increased feasibility of real-time emotion monitoring using PPG.

In summary, these studies demonstrate that emotion recognition using PPG is feasible and holds potential for real-time emotion state monitoring.

I MATERIALS AND METHODS

A Emotion Model

Accurately quantifying and analyzing emotions necessitates the modeling of emotions. However, modeling human emotions is challenging due to their complex nature, which results from the interplay of various elements. To address this challenge, two main types of emotion models are commonly used: discrete emotion models and dimensional emotion models[1].

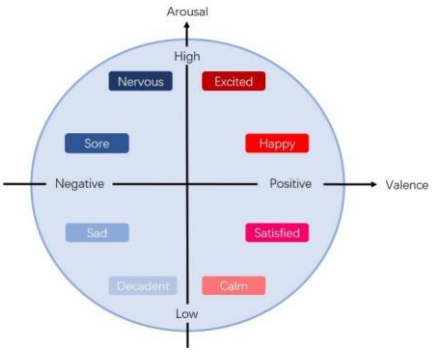

Fig. 1 Arousal valence model.

The fundamental idea of the discrete emotion model is to simplify emotions into a limited number of basic categories. For instance, six basic emotions (Anger, Happiness, Fear, Surprise, Disgust, Sadness) are widely recognized and utilized by many researchers[13]. The advantage of the discrete emotion model lies in its conceptual simplicity. However, it has notable drawbacks. It cannot represent mixed emotions or the degree of emotional intensity, such as how happy or how sad a person is. While the discrete model is clear and straightforward, its ability to describe emotional states is relatively limited. In contrast, the dimensional emotion model provides a more comprehensive and intuitive way to capture the complexity of emotions[14]. This model emphasizes the expression of emotions across different dimensions, with each dimension representing an emotional metric, creating multiple continuous scales instead of a single scale. To classify emotions, this study adopts the two-dimensional arousal-valence model based on the circumplex model. As shown in Fig. 1, the horizontal axis represents valence, indicating the pleasantness of the emotion, ranging from negative to positive. The vertical axis denotes arousal, reflecting the activation level of emotions, which ranges from calm to intense[15].

B Proposed Methodology

The main steps of a PPG emotion recognition method based on multi-window PRV feature selection proposed in this article are shown in Fig. 2. First, the entire preprocessed signal is divided according to the preset time window

size and overlap degree to form multiple non-overlapping time windows. Each time window is a fixed-length signal segment, which can be continuous or spaced. Then the PRV feature extraction operation is performed within each time window to capture the local key information in the PPG. After completing the feature extraction of all time windows, the features extracted from each time window are fused. Finally, a feature selection algorithm is employed to identify the most valuable features from the fused feature set. These selected features are then used for training and establishing the emotion classification model.

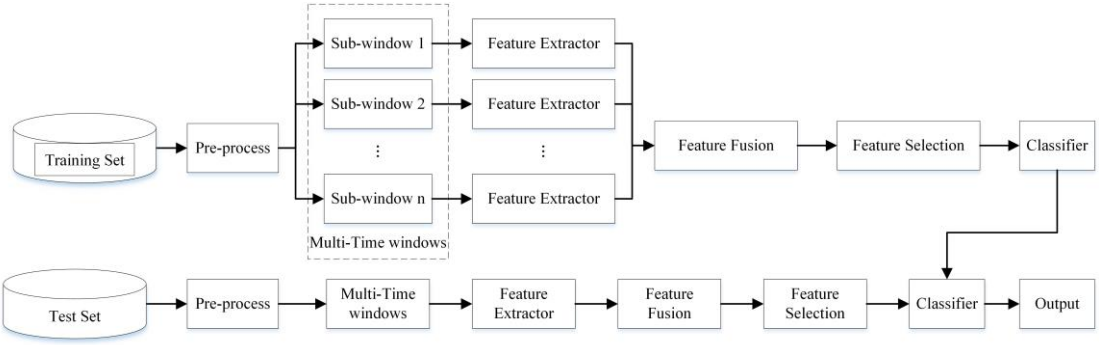

Fig. 2 Overall flow chart of the proposed method

C Multi-Time Window

Since PPG are non-stationary, meaning their statistical characteristics change over different periods, extracting features directly from the entire length of the signal is a common method when the signal length is relatively short[16]. This approach is simple, straightforward, and computationally efficient, but it also has some drawbacks. Firstly, extracting features from the entire signal length can lead to partial information loss. When the signal length is relatively long, some crucial local features may be obscured by the overall statistical measures of the entire signal. Additionally, critical information in the signal may only exist in certain local regions. For example, in experiments where physiological signals related to emotions are collected, subjects are asked to watch emotion-related videos or listen to music while their physiological signals are recorded. In such cases, due to the extended duration of data collection, the time points at which emotions are triggered often vary among subjects. Therefore, extracting features directly from the entire signal fails to capture these important local features. Finally, extracting features from the entire signal cannot capture the dynamic changes in the signal, thereby neglecting the signal's temporal variation characteristics.

To address these issues, we propose to use multi-time windows to extract PRV features from PPG. As shown in Fig. 3, the entire signal is first segmented into multiple overlapping time windows based on the chosen window size and overlap degree. Each time window typically consists of fixed-length signal segments, which can be either continuous or spaced apart. Next, feature extraction is performed within each time window. The specific 30 emotion-related PRV features from PPG are detailed in next section. After extracting features from every window, the features from each window are integrated. Finally, a feature selection algorithm is used to identify the most valuable features for building and training the emotion classification model.

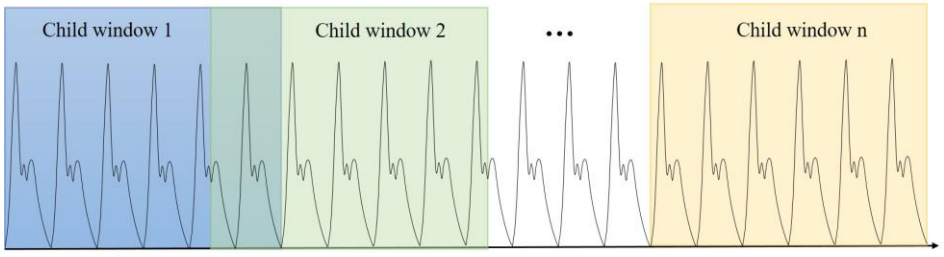

Fig. 3 Multi-Time windows

D PRV Features

Pulse Rate Variability (PRV) refers to the variation patterns in the intervals between heartbeats, with these intervals measured by the PP intervals (the time interval between consecutive heartbeats) in continuous PPG (as shown in Fig. 4) . PRV is used to describe the fluctuations in heart rate over time. These fluctuations are influenced by the body's autonomic nervous system and result from the autonomic regulation of the heart. PRV features are various parameters or characteristics extracted from PRV signals. These features describe the heart rate's fluctuations and variation patterns over time, indirectly reflecting an individual's emotional state.

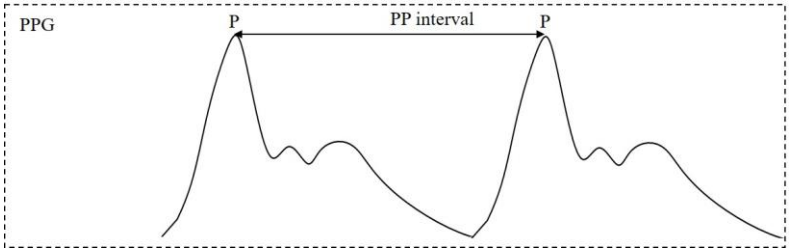

Fig. 4 PP interval

Referring to previous studies[17][18], it is known that PRV features mainly include linear and nonlinear characteristics. Therefore, this section will introduce the extracted PRV features.

In the time domain analysis, six parameters related to PRV were calculated: the mean of PP intervals ($MEAN_{PP}$), the standard deviation of PP intervals ($SDNN_{PP}$), the root mean square of successive differences of adjacent PP intervals ($RMSSD_{PP}$), the standard deviation of differences between adjacent PP intervals ($SDSD_{PP}$), the number of successive PP intervals differing by more than 50 ms ($NN50$), and the percentage of successive PP intervals differing by more than 50 ms ($pNN50$). The calculation formulas are shown as follows.

$$MEAN_{PP} = \frac{1}{n} \times \sum_{i=1}^{n} PP_i \qquad (1)$$

$$SDNN_{PP} = \sqrt{\frac{\sum_{i=1}^{n} (PP_i - MEAN_{PP})^2}{n}} \qquad (2)$$

$$RMSSD_{PP} = \sqrt{\frac{\sum_{i=2}^{n} (PP_i - PP_{i-1})^2}{n-1}} \qquad (3)$$

$$SDSD_{PP} = \sqrt{\frac{\sum_{i=2}^{n} (PP_i - PP_{i-1} - RMSSD_{PP}{}^2)^2}{n-1}} \qquad (4)$$

$$NN50 = \sum_{i=1}^{N-1} I(\left|PP_{i+1} - PP_i\right| > 50) \qquad (5)$$

$$pNN50 = \frac{\sum_{i=1}^{N-1} I(\left|PP_{i+1} - PP_i\right| > 50)}{N-1} \times 100\% \qquad (6)$$

In frequency domain analysis, 10 PRV-related parameters were computed. First, the PP interval sequence signal was resampled and interpolated, followed by using Fast Fourier Transform (FFT) to calculate the power spectral density (PSD) of the resampled PP interval sequence signal. The specific power in certain frequency ranges and the peak frequencies of three distinct frequency bands were calculated at last.. Very low frequency range (VLF) (0.003–0.04 Hz), low frequency range (LF) (0.04–0.15 Hz), and high frequency range (HF) (0.15–0.4 Hz).The calculation formulas are shown as follows. Then we calculated the following features:Total power (TP), Energy ratio between low frequency and high frequency (Ratio_LH), Normalized LF power relative to LF and HF power (LFnorm), Normalized HF power relative to LF and HF power (HFnorm), Percentage of VLF power in total power (pVLF),

Percentage of LF power in total power (pLF), Percentage of HF power in total power (pHF). The calculation formulas are shown as follows.

$$TP = VLF + HF + HF. \quad (7)$$

$$Ratio\_LH = LF/HF. \quad (8)$$

$$LF_{norm} = LF/(LF + HF). \quad (9)$$

$$HF_{norm} = HF/(LF + HF). \quad (10)$$

$$pVLF = (VLF / TP) \times 100\%. \quad (11)$$

$$pLF = (LF / TP) \times 100\%. \quad (12)$$

$$pHF = (HF / TP) \times 100\%. \quad (13)$$

In nonlinear analysis, a total of 12 Poincaré plot-related features and 2 nonlinear dynamical parameters were computed.

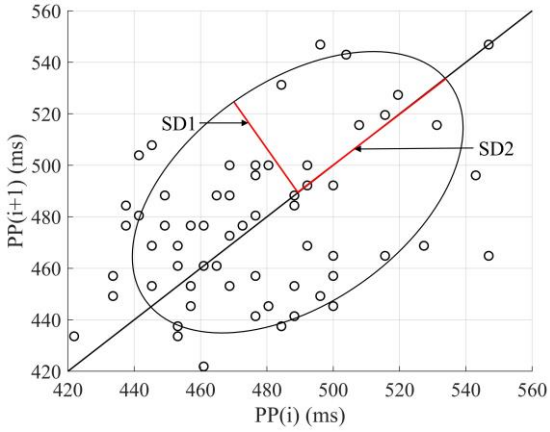

Fig. 5 Poincaré's Scatter plot

Poincaré Plot Analysis[19]: A nonlinear method used to analyze pulse rate variability, it's a scatter plot that illustrates the relationship between successive heart beat intervals. When plotting a Poincaré scatter plot, for a given PP time series $(pp_1, pp_2 \cdots, pp_n)$, it's generated by analyzing the intervals delayed by a certain number of heartbeats. Here, we set the delay to 1, thus plotting the scatter plot $pp_1, pp_2 \cdots, pp_{n-1}$ with as the horizontal axis, and $pp_2, pp_3 \cdots, pp_n$ as the vertical axis. The scatter plot typically exhibits a "comet-like" shape and can be fitted with an ellipse. From this, we extract three nonlinear features: SD1, SD2, and SD12, as the Fig. 5 shows. SD1 represents the standard deviation of the short axis of the ellipse, reflecting the instantaneous beat-to-beat variability of the PP interval. SD2 represents the standard deviation of the long axis of the ellipse, reflecting the long-term successive beat-to-beat variability of the PP interval. SD12 represents the ratio of SD1 to SD2, which signifies the activity level of the sympathetic nervous system relative to the parasympathetic nervous system.

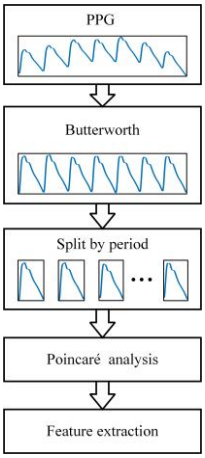

Fig. 6 Poincaré diagram feature extraction process of PPG

In addition to analyzing PP interval sequences using Poincaré plots, we use PPG as the subject of Poincaré plot analysis. The analysis process is as follows:

(1)Apply a Butterworth filter (band-pass filter: 0.6–30 Hz) for filtering. Then normalize using Eq.(14).

$$X_{normalized} = 2\frac{X - \min(X)}{\max(X) - \min(X)} - 1. \quad (14)$$

(2) Segment the original PPG into individual cycles, obtaining segmented PPG, where each segment represents a single-cycle PPG.

(3) Perform Poincaré plot analysis on each segment, with a delay of 4. For a given segment of a single-cycle PPG $(x_1, x_2 \cdots, x_n)$, plot the scatter plot with $x_1, x_2 \cdots, x_{n-4}$ as the horizontal axis and $x_5, x_6 \cdots, x_n$ as the vertical axis, as shown in Fig. 7

(4) For a single-cycle PPG cycle $(x_1, x_2 \cdots, x_n)$, 9 relevant features are extracted from the Poincaré diagram as follows: the standard deviation of the horizontal coordinates of its Poincaré scatter plot, the minimum value of the areas of all cycle scatter plot trajectories, the maximum value of the areas of all cycle scatter plot trajectories, the mean of the areas of all cycle scatter plot trajectories, the mean of the standard deviations of the horizontal coordinates of all cycle scatter plot trajectories, the mean of the standard deviations of the vertical coordinates of all cycle scatter plot trajectories, the mean of the third moments of the horizontal coordinates of all cycle scatter plot trajectories, the mean of the third moments of the vertical coordinates of all cycle scatter plot trajectories, the mean of the fourth moments of the horizontal coordinates of all cycle scatter plot trajectories, and the mean of the fourth moments of the vertical coordinates of all cycle scatter plot trajectories

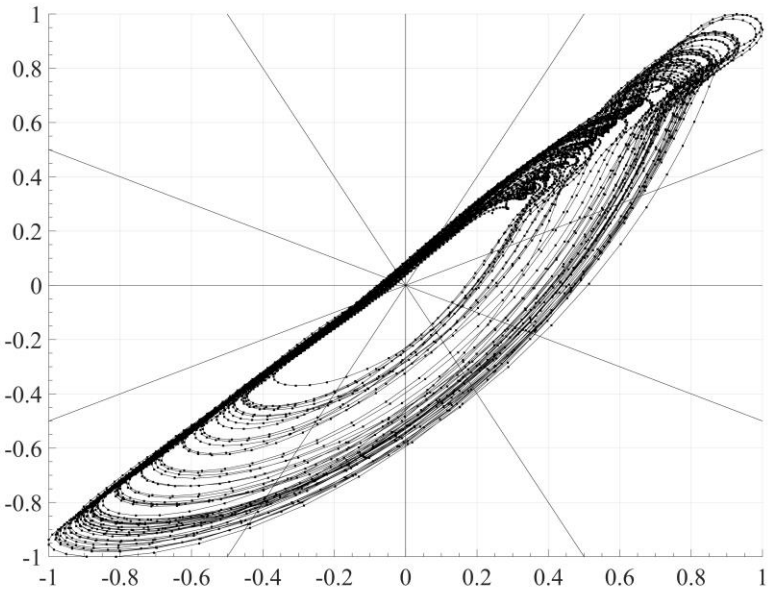

Fig. 7 Poincaré analysis of single-cyclePPG signal

In nonlinear dynamical analysis, a total of 2 features are computed, namely, Approximate Entropy and Sample Entropy.

E Feature Selection Algorithm

We employ the proposed feature selection algorithm to perform feature selection on the PRV features obtained in the previous step, aiming to obtain the optimal feature subset. First, we introduce some relevant descriptive statistics, as detailed in TABLE 1.

TABLE 1 Symbol Descriptions

Symbols	Descriptions
n	Number of samples
d	Dimensionality of features
c	Number of labels
$X \in \mathbb{R}^{n\times d}$	Training set
$Y \in \{0,1\}^{n\times c}$	Label matrix
$W \in \mathbb{R}^{d\times c}$	Weight matrix
μ	Balancing parameter
$tr(\cdot)$	Trace of a square matrix

Below is a detailed introduction to the proposed algorithm and its solving process. Given a training set $X = \{x_1, \cdots x_n\} \in \mathbb{R}^{n\times d}$,where $x_i \in \mathbb{R}^d$ represents the i-th sample, with d denoting the dimensionality of features, and n being the total number of training samples. Define the label matrix $Y = \{y_1, \cdots, y_n\}^T \in \{0,1\}^{n\times c}$ where c is the number of labels, and $y_i \in \mathbb{R}^c$ is the label vector for the i-th data.

Our goal is to learn a feature weight matrix $W \in \mathbb{R}^{d\times c}$ such that W can effectively map the data matrix X to the original label matrix Y.

To obtain such a feature weight matrix, we formulate the problem as the objective function(15):

$$\min_W L(W) = V(X, Y, W) + \mu\Omega(W), \quad (15)$$

Where V represents a loss function, Ω is employed to regulate the model's complexity, and μ is the parameter controlling Ω.

Initially, we introduce the first term V. The decision to employ the least squares loss function is based on its demonstrated simplicity and efficiency in numerous studies[20]. The basic idea is to describe the relationship between the independent and dependent variables by fitting a linear model. Then, the least squares method uses an objective function, namely the sum of squared residuals, it indicates the discrepancy between the projected values of the model and the real observed values.. By minimizing this objective function, optimal model parameters can be obtained, allowing the model to better predict the outcomes of unknown data. Specifically, this is shown as Eq.(16):

$$V(X, Y, W) = \sum_{i=1}^{n} s_i \| x_i W - y_i \|_2^2, \quad (16)$$

Where s_i represents the similarity between sample x_i and all samples, $W \in \mathbb{R}^{d\times c}$ is the regression coefficient vector of the training set X ,which represents the weights corresponding to each feature, i.e., this refers to the degree of fluctuation in the dependent variable when the independent variable changes by one unit. By fitting the training dataset, the optimal regression coefficient vector can be obtained, thus establishing a model for predicting unknown data.

For the second term, we utilize regularization based on the $L_{2,1}$ norm because of its robustness to outliers, better handling of noise and outliers, as well as its effectiveness in revealing discriminative features among multiple labels

and facilitating efficient feature selection[21]. Specifically, the $L_{2,1}$ norm penalizes coefficients for each feature group rather than individual features. This means that when certain features within a feature group are important for the classification task, the coefficients of that feature group will be preserved or amplified, while coefficients of other feature groups may be compressed or even reduced to zero. This mechanism encourages the model to select feature groups with significant discriminative power, thereby enhancing the classification performance and interpretability of the model. Specifically, this is shown as

$$\Omega(W)=\| W \|_{2,1}. \quad (17)$$

In summary, our feature selection algorithm can be summarized by the following objective function:

$$\min_W L(W) =\sum_{i=1}^{n} s_i \| x_i W - y_i \|_2^2 +\mu \| W \|_{2,1}. \quad (18)$$

Due to the non-convexity and difficulty in solving the objective function, we solve W iteratively as follows:

First, we define the matrix S, where s_{ij} represents the similarity between x_i and x_j. Then, the Eq.(18) can be written in the following form:

$$\min_W L(W) = Tr((XW - Y)^T S(X^T W - Y)) + \mu \| W \|_{2,1}. \quad (19)$$

Next, we setting $\frac{\partial L(W)}{W} = 0$, then we obtain:

$$\begin{aligned} \frac{\partial L(W)}{W} &= X^T S(XW - Y) + \mu DW = 0 \\ & X^T SXW - X^T SY + \mu DW = 0 \\ & X^T SXW + \mu DW = X^T SY \\ & (X^T SX + \mu D)W = X^T SY, \end{aligned} \quad (20)$$

where D is a diagonal matrix, defined as Eq.(21).

$$D = \begin{bmatrix} \frac{1}{2\| w^1 \|_2} & & \\ & \ddots & \\ & & \frac{1}{2\| w^d \|_2} \end{bmatrix}, \quad (21)$$

where w^i is the i-th column of the feature weight matrix W.

Due to the dependence between D and W, solving this problem is challenging. We employ an EM-style iterative method to solve for W. First, initialize W randomly to obtain D, then we obtain:

$$W = (X^T SX + \mu D)^{-1} X^T SY. \quad (22)$$

Then, update W according to Eq.(22), iterating repeatedly until convergence.

Based on the above introduction, Algorithm 1 depicts the feature selection process for solving W in the multi-time-window PRV:

Algorithm 1: Multi-Time-Window PRV Feature Selection Algorithm

Input: Training set $X_i \|_{i=1}^n \in \mathbb{R}^{d \times n}$, labels of the training set $Y_l \|_{l=1}^n \in \mathbb{R}^{n \times c}$, regularization parameter μ

Output: Feature weight matrix W

1. Compute the weight matrix of the training set S
2. Set $t=0$, initialize W
3. **while** not converged **do**
4. Compute the diagonal matrix D according to Eq.(21)
5. Compute W according to Eq.(22)
6. $t=t+1$
7. **end while**
8. return W

II RESULTS

A Dataset

In this paper, we selected the publicly available DEAP dataset to evaluate model performance[22]. The DEAP dataset comprises multimodal data, encompassing electroencephalogram (EEG) signals, PPG, and various other physiological signals collected from 30 participants as they viewed 40 one-minute music videos.. PPG were collected from the right thumb with a sampling rate of 512 Hz for a duration of 30 seconds. These data were subsequently downsampled to 256 Hz, and each signal was processed with a moving average filter to achieve signal smoothing. After watching the videos, each participant self-assessed on four dimensions: valence, arousal, dominance, and liking, with scores ranging from 1 to 9. This study focuses on the PPG data from this dataset and performs binary classification on the arousal and valence scores.

B Implementation Details

Initially, the dataset undergoes partitioning into a training set (70%) and a test set (30%) via random selection. We then conducted 10 experiments and averaged the results. For the multi-window parameters, we set the time window size to 10 seconds, with no overlap between adjacent windows. During the optimization process, the regularization parameter was established as 0.1, and the maximum iteration limit was defined as 5. The kernel function for the SVM classifier was set to the Gaussian kernel. Finally, we use accuracy and F1 value as evaluation metrics.

C Emotion Recognition Result

We will compare the proposed method with algorithms commonly used for PPG emotion recognition classification and some time-series classification models. The details are as follows:

(1) CNN[23]: This approach implements a 3-layer convolutional neural network to extract profound features from PPG and PP intervals. It then integrates these two deep features to classify valence and arousal.

(2) ALSTM-FCN[24]: A model for time-series classification that combines Attention LSTM and Fully Convolutional Networks (FCN). By introducing an attention layer in ALSTM-FCN's LSTM, the model retains the memory capabilities of LSTM while reducing computational complexity and better capturing key information in time-series data. Moreover, the FCN architecture is employed for feature extraction from time-series data, thereby enhancing the model's robustness.

(3) InceptionTime[25]: A model for time-series classification that combines Inception modules and one-dimensional convolution. The Inception module comprises numerous convolutional kernels with varying sizes, facilitating the concurrent capture of both local and global features, thereby augmenting model efficacy. It also uses one-dimensional convolution to process sequence data, reducing the number of parameters and speeding up computation.

(4) LITE[26]: A model for time-series classification that integrates DepthWise Separable Convolutions (DWSC) into the InceptionTime model. It employs techniques like multiplexing, custom filters, and dilated convolutions to

boost performance, significantly reducing the number of trainable parameters while maintaining performance.

(5) DT[27]: Decision Trees (DT) solve classification problems by iteratively splitting the data according to given parameters. The tree is comprised of decision nodes responsible for data partitioning and leaves dedicated to decision-making.

(6) GNB[27] : Gaussian Naive Bayes (GNB) is a classification algorithm that relies on Bayes' theorem, which calculates the probability of one event happening based on the likelihood of another event occurring.

(7) KNN[27]: The K-Nearest Neighbors (KNN) algorithm represents a straightforward machine learning technique wherein data points are classified according to their proximity to the nearest neighbors (K value).

For CNN, ALSTM-FCN, InceptionTime, and LITE, we use the original length of the PPG as the network input for classification. For SVM, DT, GNB, and KNN, we use 30 PRV features extracted from the entire length of the PPG as the input for classification.

For CNN, ALSTM-FCN, InceptionTime, and LITE, we use the original length of the pulse signals as the network input for classification. For SVM, DT, GNB, and KNN, we use 30 PRV features extracted from the entire length of the pulse signals as the input for classification.The experimental results on the DEAP dataset for each method are shown in TABLE 2, with bold numbers indicating the optimal values. Our method is denoted as Ours. The results clearly show that the proposed method outperforms current leading methods in pulse signal emotion recognition for both the Arousal and Valence dimensions. Specifically, the classification accuracy for the Arousal dimension improved by 1.23%, and for the Valence dimension, it improved by 2.41%. The F1 scores also showed varying degrees of improvement.

Additionally, it is evident that deep learning methods frequently outperform traditional machine learning classification techniques in terms of achieving higher classification accuracy. but the variance in their results is generally larger than that of machine learning methods. Among the four machine learning classifiers, the SVM classifier achieved significantly better results than the other methods. For our method, it not only surpasses the well-performing deep learning methods regarding classification accuracy and F1 score but also has a variance comparable to traditional machine learning methods. This indicates that the proposed method's approach of multi-window feature extraction followed by feature selection effectively yields a feature subset suitable for emotion recognition.

TABLE 2 The Results of the Prosed Model on the DEAP Datasets Compared With Prior Works

Methods	DEAP Arousal		DEAP Valence	
	ACC	F1	ACC	F1
CNN	62.08±1.52	76.60±1.34	56.59±1.67	70.61±1.53
RNN-FCN	62.70±1.67	75.82±1.59	58.10±1.20	71.75±1.30
ALSTM-FCN	63.45±1.76	75.89±1.60	57.78±1.56	69.79±1.28
InceptionTime	65.02±1.64	74.47±1.57	61.34±1.68	71.74±1.78
LITE	66.19±1.56	75.74±1.39	62.89±1.32	72.64±1.20
SVM	63.12±0.67	74.67±0.60	59.12±0.40	70.74±0.54
DT	53.75±0.87	61.05±0.76	50.62±0.79	59.06±0.88
GNB	52.62±0.98	58.20±0.95	52.5±1.10	64.48±1.02
KNN	56.25±0.67	62.76±0.64	53.75±0.52	58.42±0.56
Ours	**66.25±0.40**	**76.92±0.47**	**63.75±0.66**	**73.14±0.48**

D Comparison with Feature Selection Methods

To assess the efficacy of each module within this approach, we conducted experiments employing various combinations of time window quantities and feature selection algorithms. This allowed us to determine the optimal

number of windows and demonstrate the effectiveness of the proposed feature selection algorithm. For the selection of the number of time windows, since the sample length of the dataset used in this paper is 30 seconds, we ensured that each time window has a consistent data length with no overlap between windows. We experimented with four different time window lengths.

As shown in Fig. 8, Fig. 9 and Fig. 10, for example :the time window size is the entire length of the signal, including only window T1 from Fig. 8.

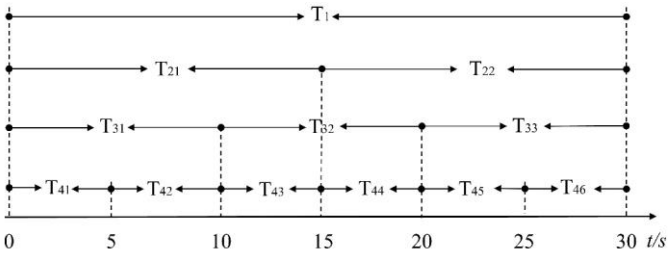

Fig. 8 Division of Multi-Time Windows

We simultaneously selected the following four classic feature selection algorithms: CMIM[28], Chi-square[29], Relief[30], mRMR[31] ,along with no feature selection algorithm (denoted as NO) and the feature selection method proposed in this research (denoted as Ours), to combine with different numbers of time windows to calculate the classification accuracy of the model.

As indicated in the results, employing time window T3, the feature selection approach introduced in this study attains the highest classification accuracy for both the Arousal and Valence dimensions. As the number of time windows increases, the model's classification accuracy generally shows an initial increase followed by a decrease. This suggests that with an increase in the number of time windows, the count of valuable features likewise rises. The model achieves its peak classification accuracy at time window T3, exhibiting an accuracy of 66.25% for the Arousal dimension and 63.75% for the Valence dimension.

However, as the number of time windows continues to increase, the model's accuracy starts to decline. This is because too many time windows, while capturing features well, also increase feature redundancy. Therefore, after using feature selection algorithms to filter features, the model's performance is further improved. Additionally, compared to other feature selection algorithms, the feature selection method proposed in this research achieves the highest classification accuracy in most time windows.

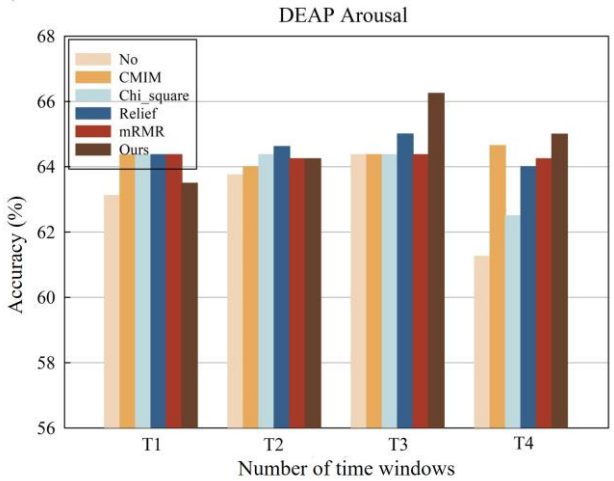

Fig. 9 Accuracy of different time windows and feature selection algorithms on Arousal

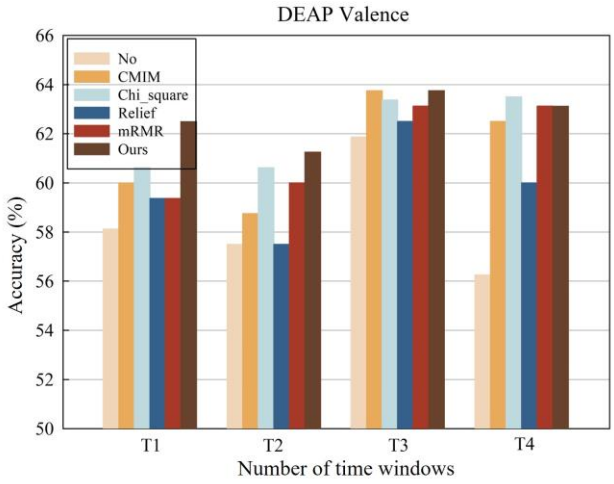

Fig. 10 Accuracy of different time windows and feature selection algorithms on Valence

III Discussion

In the research on using PPG to recognize emotions, due to the traditional method of feature extraction from the entire signal, local features are masked by the overall statistics, and local emotion-related features cannot be fully captured. We proposes a PPG emotion recognition method based on multi-time window PRV feature selection. This method uses Multi-Time windows to extract 30 PRV-related features from the PPG to better capture the local dynamic information of the signal, thereby improving the richness of feature information and being able to extract different features according to different stages of signal change. Features enable the recognition model to more accurately capture the emotion-related information in the PPG, and then fuse the features. Simultaneously, to address the issue of escalating feature redundancy resulting from feature extraction across multiple time windows. Utilizing the least squares loss function and norm, we formulated the objective optimization function and employed an EM-style iterative approach to resolve the feature weight matrix. Finally, the feature weight matrix is used to select the optimal feature subset. The experimental outcomes on the DEAP dataset demonstrate that our approach surpasses the sophisticated PPG-based emotion recognition method in both accuracy and F1 score. Furthermore, the feature selection approach introduced in this study outperforms various other feature selection algorithms in the identification of emotion-related features. In summary, this approach contributes to the enhancement of accuracy in PPG-based emotion recognition research.

Data sharing agreement

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

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, author-ship, and publication of this article.

Funding

The author received no financial support for the research.

References

[1] Wang Y, Song W, Tao W, et al. A systematic review on affective computing: Emotion models, databases, and recent advances[J]. Information Fusion, 2022, 83: 19-52.

[2] Li D, Xie L, Wang Z, et al. Brain emotion perception inspired EEG emotion recognition with deep reinforcement learning[J]. IEEE Transactions on Neural Networks and Learning Systems, 2023.

[3] Sarkar P, Etemad A. Self-supervised ECG representation learning for emotion recognition[J]. IEEE Transactions on Affective Computing, 2020, 13(3): 1541-1554.

[4] Goshvarpour A, Goshvarpour A. Evaluation of novel entropy-based complex wavelet sub-bands measures of PPG in an emotion recognition system[J]. Journal of Medical and Biological Engineering, 2020, 40(3): 451-461.

[5] Paul T, Bhattacharyya C, Sen P, et al. Human emotion recognition using GSR and EEG[J]. International Journal of Scientific and Research Publication, 2020, 10(5): 394-400.

[6] Kose M R, Ahirwal M K, Kumar A. A new approach for emotions recognition through EOG and EMG signals[J]. Signal, Image and Video Processing, 2021, 15(8): 1863-1871.

[7] Li X, Zhang Y, Tiwari P, et al. EEG based emotion recognition: A tutorial and review[J]. ACM Computing Surveys, 2022, 55(4): 1-57.

[8] Ismail S N M S, Aziz N A A, Ibrahim S Z. A comparison of emotion recognition system using electrocardiogram (ECG) and photoplethysmogram (PPG)[J]. Journal of King Saud University-Computer and Information Sciences, 2022, 34(6): 3539-3558.

[9] Singh R R, Conjeti S, Banerjee R. A comparative evaluation of neural network classifiers for stress level analysis of automotive drivers using physiological signals[J]. Biomedical Signal Processing and Control, 2013, 8(6): 740-754.

[10] Udovičić G, Đerek J, Russo M, et al. Wearable emotion recognition system based on GSR and PPG signals[C]//Proceedings of the 2nd International Workshop on Multimedia for Personal Health and Health Care. 2017: 53-59.

[11] Raheel A, Majid M, Alnowami M, et al. Physiological sensors based emotion recognition while experiencing tactile enhanced multimedia[J]. Sensors, 2020, 20(14): 4037.

[12] Heo S, Kwon S, Lee J. Stress detection with single PPG sensor by orchestrating multiple denoising and peak-detecting methods[J]. IEEE Access, 2021, 9: 47777-47785.

[13] Ortony A. Are all “basic emotions” emotions? A problem for the (basic) emotions construct[J]. Perspectives on Psychological Science, 2022, 17(1): 41-61.

[14] Pérez-Toro P A, Vásquez-Correa J C, Bocklet T, et al. User state modeling based on the arousal-valence plane: Applications in customer satisfaction and health-care[J]. IEEE Transactions on Affective Computing, 2021, 14(2): 1533-1546.

[15] Zitouni M S, Park C Y, Lee U, et al. Arousal-valence classification from peripheral physiological signals using long short-term memory networks[C]//2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, 2021: 686-689.

[16] Nitzan M, Ovadia-Blechman Z. Physical and physiological interpretations of the PPG signal[M]//Photoplethysmography. Academic Press, 2022: 319-340.

[17] Prada E J A, Higinio A P. A low-complexity PPG pulse detection method for accurate estimation of the pulse rate variability (PRV) during sudden decreases in the signal amplitude[J]. Physiological measurement, 2020, 41(3): 035001.

[18] Zhou K, Schinle M, Stork W. Dimensional emotion recognition from camera-based PRV features[J]. Methods, 2023, 218: 224-232.

[19] Wang Z H, Wu Y C. A Novel Rapid Assessment of Mental Stress by Using PPG Signals Based on Deep Learning[J]. IEEE Sensors Journal, 2022, 22(21): 21232-21239.

[20] Tarekegn A N, Giacobini M, Michalak K. A review of methods for imbalanced multi-label classification[J]. Pattern Recognition, 2021, 118: 107965.

[21] Qian W, Huang J, Xu F, et al. A survey on multi-label feature selection from perspectives of label fusion[J]. Information Fusion, 2023, 100: 101948.

[22] Koelstra S, Muhl C, Soleymani M, et al. Deap: A database for emotion analysis; using physiological signals[J]. IEEE transactions on affective computing, 2011, 3(1): 18-31.

[23] Lee M S, Lee Y K, Lim M T, et al. Emotion recognition using convolutional neural network with selected statistical photoplethysmogram features[J]. Applied Sciences, 2020, 10(10): 3501.

[24] Karim F, Majumdar S, Darabi H, et al. Multivariate LSTM-FCNs for time series classification[J]. Neural networks, 2019, 116: 237-245.

[25] Ismail Fawaz H, Lucas B, Forestier G, et al. Inceptiontime: Finding alexnet for time series classification[J]. Data Mining and Knowledge Discovery, 2020, 34(6): 1936-1962.

[26] Ismail-Fawaz A, Devanne M, Berretti S, et al. Lite: Light inception with boosting techniques for time series classification[C]//2023 IEEE 10th International Conference on Data Science and Advanced Analytics (DSAA). IEEE, 2023: 1-10.

[27] Ismail S N M S, Aziz N A A, Ibrahim S Z. A comparison of emotion recognition system using electrocardiogram (ECG) and photoplethysmogram (PPG)[J]. Journal of King Saud University-Computer and Information Sciences, 2022, 34(6): 3539-3558.

[28] Thakkar A, Lohiya R. A survey on intrusion detection system: feature selection, model, performance measures, application perspective, challenges, and future research directions[J]. Artificial Intelligence Review, 2022, 55(1): 453-563.

[29] Li J, Cheng K, Wang S, et al. Feature selection: A data perspective[J]. ACM computing surveys (CSUR), 2017, 50(6): 1-45.

[30] Dokeroglu T, Deniz A, Kiziloz H E. A comprehensive survey on recent metaheuristics for feature selection[J]. Neurocomputing, 2022, 494: 269-296.

[31] Jia W, Sun M, Lian J, et al. Feature dimensionality reduction: a review[J]. Complex & Intelligent Systems, 2022, 8(3): 2663-2693.