Research on heart rate estimation algorithm based on dynamic PPG

Jiawei Guo[1,2†], Shiyuan Chen[3†], Ting Lan[2], Ruochen Li[1,2], Lichao Wang[3], Yunchong Wu[3], Jun Zhong[2]* and Wei Zhu[3]*

[1]School of Biomedical Engineering (Suzhou), Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, China, [2]Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Science, Suzhou, China, [3]PLA Naval Medical Center, Shanghai, China

Heart rate is one of the most vital physiological parameters and is clinically widely used to assess human health status. In recent years, wearable devices based on photoplethysmography (PPG) have been extensively applied in real-time monitoring. However, PPG signals are susceptible to interference from various types of noise during acquisition, particularly motion artifacts (MA), which pose a significant challenge to the accurate extraction of physiological parameters. This study focuses on heart rate extraction from dynamic PPG signals and explores denoising methods combining traditional signal processing and machine learning techniques. The main research contents of this paper are as follows: further improvements are made on the basis of existing algorithms by integrating support vector machines (SVM). A more comprehensive signal quality assessment is performed via SVM, which incorporates the time-domain and frequency? domain statistical characteristics of both PPG signals and triaxial acceleration (ACC) signals. In addition, the short-time Fourier transform (STFT) is integrated to capture time-varying characteristics, thereby mitigating the impact of local signal quality degradation on the analysis of full-window signals. For spectral peak tracking, a Gaussian window is adopted to optimize the spectral search range and a comprehensive analysis is conducted by fusing spectral amplitude information with historical heart rate data. Experimental results demonstrate that the heart rate error of the test set is 1.71 beats per minute (BPM).

Introduction

Alongside the rapid development of society, the prevalence of cardiovascular diseases (CVDs) in China has been on a steady rise. For CVDs, early identification and prevention represent effective strategies for curbing the escalation of morbidity and mortality rates. Wearable devices capable of continuous human activity monitoring have broken through the constraints of traditional physiological data collection methods and are now widely employed in health monitoring for CVD patients (Sun et al., 2024). Moreover, the integration of wearable devices with artificial intelligence (AI) offers distinct advantages and considerable potential in the field of biomedical engineering. For instance, the multidisciplinary approach adopted by Mazumdar et al. (2025) in their design of a soft robotic system for Parkinson's disease highlights the potential of combining soft robotics, functional materials, and machine learning to develop novel healthcare solutions.

Heart rate is one of the critical indicators for evaluating human health status and is therefore a physiological parameter that wearable devices need to monitor continuously.

Compared with electrocardiogram (ECG) signals, devices based on PPG technology are more suitable for daily health monitoring due to advantages such as portability and ease of wear. Despite these notable merits, several non-negligible issues persist during practical application. For example, physical exercise and daily activities can lead to gaps between the sensor and the skin, allowing ambient light to penetrate and consequently generating motion artifacts (Maeda et al., 2011). These motion artifacts can significantly degrade the quality of the collected PPG signals, which in turn impairs the accuracy of measuring physiological parameters such as heart rate.

Heart rate extraction from dynamic PPG signals based on traditional signal processing can be divided into two phases: signal enhancement and heart rate estimation. Signal enhancement primarily leverages signal processing techniques to improve signal quality, thereby reducing motion artifacts and other types of noise in PPG signals. This phase can be further subdivided into two sub-stages: preprocessing and signal denoising. The core objective of preprocessing is to eliminate noise outside the heart rate frequency range to ensure signal purity; signal denoising, by contrast, focuses on removing complex noise such as motion artifacts and enhancing the valid components of the signal, laying a solid foundation for subsequent analyses. The heart rate estimation phase typically comprises two steps: spectral peak tracking and post-processing. In the spectral peak tracking step, the signal is transformed into the frequency domain to identify spectral peaks associated with heart rate. In the post-processing step, methods such as filtering and smoothing are applied to further optimize heart rate estimation results, improving calculation accuracy and stability. To enhance the accuracy of heart rate extraction, research efforts worldwide have primarily focused on optimizing signal denoising technologies and refining spectral peak tracking algorithms.

Related work

Sun and Jia (2020) proposed a PPG signal denoising method based on ensemble empirical mode decomposition (EEMD) and wavelet threshold filtering. The combination of these two techniques can effectively avoid misjudgment caused by noise-dominated conditions. Results demonstrate that this method can maximize the preservation of the nonlinear and non-stationary characteristics of PPG signals. Khan et al. (2015) put forward a two-stage denoising algorithm for PPG signals. The first stage employs the absolute criterion of EEMD to eliminate outlier errors and does not rely on historical heart rate data. The second stage integrates recursive least squares (RLS) filtering and time-domain extraction techniques to enhance the algorithm's robustness. By iteratively adjusting the filter parameters, RLS can adapt to the noise characteristics under varying exercise intensities. Coupled with a forgetting factor, it balances convergence speed and estimation stability during signal processing, thus preventing error drift induced by sudden changes in motion states.

Chung et al. (2018) proposed constructing a finite state machine (FSM) to determine the reliability of the current heart rate value, based on the prominence of the dominant spectral peak within the frequency spectrum and its deviation from the previously estimated heart rate. By discarding unreliable heart rate values, this method enhances the accuracy of heart rate estimation. Meng et al. (2022) extracted step frequency information from acceleration signals and established an adaptive model based on the correlation among step

frequency information, historical heart rate, and current heart rate. This model narrows the potential frequency range of heart rate, thereby reducing estimation errors.

Lan et al. (2024) seamlessly integrated the adaptive noise suppression advantage of RLS filtering for non-stationary PPG signals with a reliability evaluation mechanism for verifying heart rate estimation, thereby constructing a novel framework for heart rate extraction from PPG signals. When judging the intensity of MA, triaxial acceleration signals were used as reference; RLS filtering and empirical wavelet transform (EWT) were added for denoising. Meanwhile, FSM was introduced to evaluate the reliability of historical heart rate data, which optimized the spectral peak selection strategy and improved the stability of the algorithm.

Xiong et al. (2017) regarded spectral peak selection as a classification problem. They extracted the peak coefficient ratio of candidate spectral peaks and the distance from the previous heart rate spectral peak as features, and utilized a SVM to perform binary classification on the spectral peaks, with the classification results being either true spectral peaks or false spectral peaks. The FSM framework proposed by Lan et al. (2024) has certain limitations in the utilization of features in terms of state transition rules. In contrast, machine learning methods can extract features related to the prediction target from signals and perform optimization based on these extracted features. Therefore, this study integrates machine learning into the PPG signal heart rate extraction framework, which can effectively integrate different spectral peak selection rules and further reduce the error of heart rate estimation.

The main contributions of this study are as follows:

1. Based on the algorithm proposed by Lan et al. (2024), SVM is adopted to replace the FSM framework. By constructing an SVM classification model and taking multi-dimensional feature vectors as the decision basis, a more adaptive state transition criterion is established to evaluate the quality of signals processed by RLS filtering, which compensates for the limitations of FSM in feature utilization.
2. According to the signal quality evaluation results, different spectral peak selection rules are employed. Spectral peak tracking is performed on denoised signals in combination with the STFT, followed by the post-processing of estimated heart rate values. This enables more refined and accurate analysis of the local time-frequency domain characteristics, thereby mining more abundant information from the signals.
3. The experiments supplement data from various types of arm movements, and the algorithm is validated using both public datasets and self-collected datasets. Experimental results demonstrate that the proposed algorithm achieves lower estimation error in heart rate prediction and maintains high accuracy across diverse motion states.

Materials and methods

Datasets

Public dataset

The public dataset was sourced from the open data of the 2015 IEEE Signal Processing Cup (Temko, 2017). This dataset is

TABLE 1 Attributes of the test dataset.

No.	Subject ID	Motion type	Age	Weight (kg)	Height (cm)	Gender	Healthy or not
1	S1	T1	29	70	169	Male	Yes
2	S2	T1	21	77	188	Male	Yes
3	S2	T2	21	77	188	Male	Yes
4	S3	T2	19	54	174	Male	Yes
5	S4	T2	20	64	162	Male	Yes
6	S5	T2	20	57	174	Male	Yes
7	S6	T1	19	70	180	Male	Yes
8	S6	T2	19	70	180	Male	Yes
9	S7	T2	21	73	180	Male	Yes
10	S8	T1	58	70	156	Female	No

primarily designed for research on heart rate extraction under motion states and has been widely applied in the field of PPG-based heart rate extraction (Choe et al., 2024; Huang et al., 2023; Ray et al., 2022; Zhang et al., 2022). Although it dates back many years, its rigorous synchronized acquisition method and high data quality ensure that it still retains considerable research value in this domain.

To enhance data diversity, 12 training sets and 10 test sets were selected from the aforementioned dataset. Each training set contains 2-channel PPG signals, 3-axis ACC signals, and 1-channel ECG data, with a sampling frequency of 125 Hz. The PPG signals were collected by pulse oximeters equipped with 515 nm green LEDs. All participants were healthy male subjects aged between 18 and 35 years old. During data acquisition, the subjects performed 5-min exercises on a treadmill with varying speeds following this protocol: 30 s at 1–2 km/h, 1 min at 6–8 km/h, 1 min at 12–15 km/h, 1 min at 6–8 km/h, 1 min at 12–15 km/h, and 30 s at 1–2 km/h. The test set data cover a wider range of hand movements, and the specific dataset attributes are presented in Table 1. In the table, T1 denotes common rehabilitation exercises involving various forearm and upper arm movements (e.g., hand gripping and stretching), while T2 represents more vigorous forearm and upper arm movements such as boxing.

Self-collection dataset

Data acquisition equipment

The self-collected data were acquired using a device developed by the Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences. Each dataset includes 4-channel PPG signals and 3-axis ACC signals, with the sampling frequency of PPG signals set at 250 Hz and that of ACC signals at 25 Hz. Both PPG and ACC signals were collected by a smart wristwatch (Model: MPPB-V1), and the PPG data were captured via a green LED with a wavelength of 540 nm.

Data acquisition protocol

The self-collected data in this study were obtained from 8 healthy male subjects (aged 18–35 years). Prior to the experiment, each subject wore a chest strap on the chest and the smart wristwatch on the left wrist. The experiment was conducted in an indoor environment, and the specific protocol is as follows:

TABLE 2 Attributes of the self-collected dataset.

Dataset	Motion state	MA intensity
1	Walking	Medium
2	Resting	Weak
3	Walking	Medium
4	Running	Strong
5	Resting-walking-running	Strong
6	Resting	Weak
7	Resting	Weak
8	Resting	Weak

1. At the start of the experiment, all subjects maintained a static sitting posture for 5 min. This step allowed the subjects to acclimatize to the device-wearing state, regulate their breathing, and ensure the stable operation of the acquisition equipment.

2. All subjects performed three different types of activity tasks as required by the experiment: static sitting, walking, and running. The specific arrangement was as follows: 4 subjects maintained a static sitting posture throughout the experiment; 2 subjects performed continuous walking; 1 subject engaged in continuous running; and 1 subject completed the task sequence of 1-min static sitting, 2-min walking, and 2-min running. During walking and running, all subjects were required to maintain a natural arm-swinging motion, and the exercise intensity was self-regulated by the subjects. The data acquisition duration was set to 5 min for each subject. This experiment was designed to cover diverse exercise intensities and heart rate variation scenarios, so as to verify the adaptability and robustness of the proposed method across multiple motion states.

3. After the completion of data acquisition, the subjects remained seated with the devices still worn for an additional 1 min.
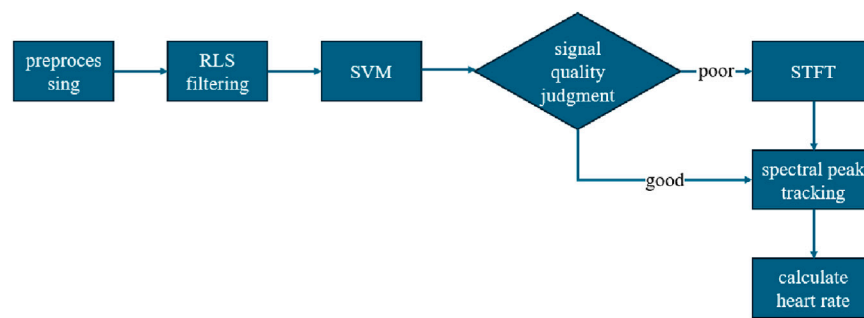
FIGURE 1
An SVM classification model is constructed to evaluate the quality of signals after preprocessing and RLS filtering. Based on the evaluation results, different spectral peak screening rules are adopted. Combined with STFT, spectral peak tracking is performed on the denoised signals, and finally, post-processing is conducted on the estimated heart rate.

Data attributes

The 8 collected datasets were parsed and sorted in chronological order of timestamps. The corresponding PPG signals and ACC signals were extracted and saved as. txt format files, respectively. The specific attributes of the datasets are presented in Table 2. Specifically, Datasets 2, 6–8 correspond to the static sitting state; Datasets 1 and 3 correspond to the walking state; Dataset 4 corresponds to the running state; and Dataset 5 covers the transition state from rest to walking and then to running.

Dataset splitting

To ensure data independence and prevent data leakage, this study split the 22 datasets into training and test sets at a ratio of 8:2, with the division performed on a per-subject basis. Specifically, the first 17 datasets were selected as the training set, and the remaining 5 datasets were used as the test set. This ratio is a classic split in the field: 80% of the subjects allocated to the training set provide sufficient sample diversity and data volume, ensuring that the model can learn the features of physiological signals; the remaining 20% assigned to the test set have an adequate sample size to support statistical analysis, thus guaranteeing the stability of evaluation results.

Splitting the datasets by subjects enables accurate assessment of the model's true generalization ability, directly reflecting the model's adaptability to new subjects and avoiding the overfitting problem caused by random splitting.

PPG signal heart rate extraction framework

Building upon the heart rate estimation framework proposed by Lan et al. (Lan et al., 2024), this study replaces the FSM framework with the SVM and integrates it with different spectral peak selection rules, thereby effectively reducing the error of heart rate estimation. The specific algorithm flow is illustrated in Figure 1.

PPG signal preprocessing

Consistent with the signal preprocessing method proposed by Lan et al. (2024), each dataset was segmented using a window length and sliding window consistent with the reference heart rate settings, which were 8 s and 2 s, respectively. Band-pass filtering was applied

to the PPG signals of each channel and the triaxial ACC signals. The band-pass range was set to 0.4–3.5 Hz, corresponding to a heart rate range of 24–210 BPM.

To reduce computational complexity and signal dimensionality while preserving the main features of the signals, averaging processing was performed on both the PPG and ACC signals. The band-pass filtered signals were normalized via L2-norm normalization, and then the signals of each channel were subjected to weighted averaging to obtain the averaged PPG signal (PPGcom) and averaged ACC signal (ACCcom).

The specific computing method are presented in Equations 1, 2.

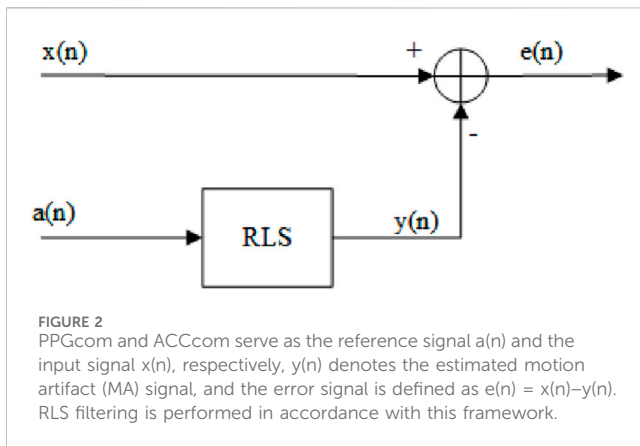$$PPG_{com} = \frac{1}{2}\left(\frac{PPG_1}{\|PPG_1\|_2} + \frac{PPG_2}{\|PPG_2\|_2} \right) \quad (1)$$

$$Acc_{com} = \frac{1}{3}\left(\frac{Acc_x}{\|Acc_x\|_2} + \frac{Acc_y}{\|Acc_y\|_2} + \frac{Acc_z}{\|Acc_z\|_2} \right) \quad (2)$$

Where PPG1 and PPG2 denote the two-channel PPG signals after band-pass filtering, respectively; ACCx, ACCy, and ACCz denote the triaxial acceleration signals after band-pass filtering, respectively.

RLS filtering

Adaptive filtering minimizes the discrepancy between the reference signal and the input signal by continuously iteratively adjusting the filter parameters. When the reference signal can well reflect MA information, adaptive filtering can effectively remove MA. Basic adaptive filtering methods include least mean square (LMS) filtering and RLS filtering. Compared with LMS filtering, RLS filtering offers the advantages of faster convergence speed and better adaptability to non-stationary signals (Geng and Zhang, 2008), making it more suitable for processing non-stationary signals such as PPG signals. Therefore, RLS filtering was selected for this study.

PPGcom and ACCcom described in Section 3.2.1 were used as the reference signal a(n) and the input signal x(n), respectively. The estimated MA signal is denoted as y(n), and the error signal is defined as e(n) = x(n) − y(n). The framework is illustrated in Figure 2. The parameters of RLS were set as follows: order N = 55, forgetting factor λ = 0.999, and initial covariance estimate Θ = 0.1.

FIGURE 2
PPGcom and ACCcom serve as the reference signal a(n) and the input signal x(n), respectively, y(n) denotes the estimated motion artifact (MA) signal, and the error signal is defined as e(n) = x(n)–y(n). RLS filtering is performed in accordance with this framework.

SVM-based imbalanced classification
Data balancing based on SMOTE

Taking the difference between the frequency domain information of PPG signals processed by RLS filtering and the reference heart rate as the evaluation metric, signal quality labeling was performed on each sample. The results indicated that the number of samples with good signal quality was far greater than that of samples with poor signal quality, resulting in a data imbalance phenomenon. Direct classification on imbalanced data will cause the classifier to be biased toward the majority class in prediction results. Therefore, data balancing is required prior to classification.

The Synthetic Minority Over-Sampling Technique (SMOTE) was selected herein to achieve data balancing. SMOTE was proposed by Chawla et al. (Arunkumar and Bhaskar, 2020) in 2002 to address the problem of large class imbalance ratios in classification tasks. Its core idea is to balance the data by generating minority class samples through interpolation between neighboring samples. Specifically, it calculates the Euclidean distance (i.e., geometric distance) between each minority class sample and other minority class samples to measure the dissimilarity between two samples. Based on the calculation results, the K nearest samples are selected from all minority class samples, which is the K-nearest neighbor (KNN) method (Holmes and Adams, 2002). Then, one sample is randomly selected from the chosen ones, linear interpolation is performed, and new samples are generated. By augmenting the minority class samples in the training data using SMOTE, the total number of training samples is increased from 2,427 to 4,822, and the class ratio is adjusted from the original 1:150 to 1:1.

Classifier selection

To compensate for the limitations of the FSM framework in feature utilization, machine learning is adopted in this study to perform signal quality classification on the balanced data, classifying the signals into two categories: good quality and poor quality. This approach also enables the integration of different spectral peak selection rules for subsequent processing. Three classic machine learning classifiers were employed for performance comparison, namely SVM, random forest (Breiman, 2001), and K-nearest neighbor (Cover and Hart, 1967). These methods are classic, representative, and practically validated algorithms in the field of machine learning, covering three core technical routes: discriminative learning, ensemble learning, and instance-based

learning. Verified by long-term academic research and engineering practice, they share common advantages such as stable generalization, clear parameter tuning logic, and broad scenario adaptability, which can fully support the experimental requirements for model comparison and the goal of scenario adaptability.

First, parameter sensitivity analysis is carried out through preliminary experiments to identify the key parameters that exert a significant impact on classification performance for each classifier. For these key parameters, multiple groups of gradient parameter combinations are designed for comparison experiments. The model performance metrics under each parameter combination are evaluated to complete the parameter optimization process. After the key parameters of each model are adjusted to their optimal values, a horizontal comparison of the performance of the SVM, random forest, and K-nearest neighbor classifiers is conducted based on a unified experimental environment. Ultimately, the optimal classifier suitable for the target task is selected.

Feature selection

During the feature selection process for each classifier, a total of 11 features were selected as classification features, which were derived from the time-domain and frequency-domain features of the original PPG signal (PPGpre), the PPG signal processed by RLS filtering (PPGRLS), and the triaxial acceleration signals. The detailed descriptions are as follows:
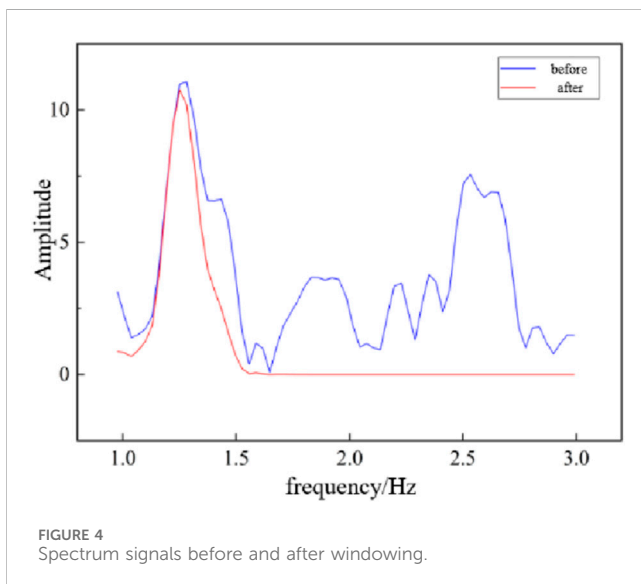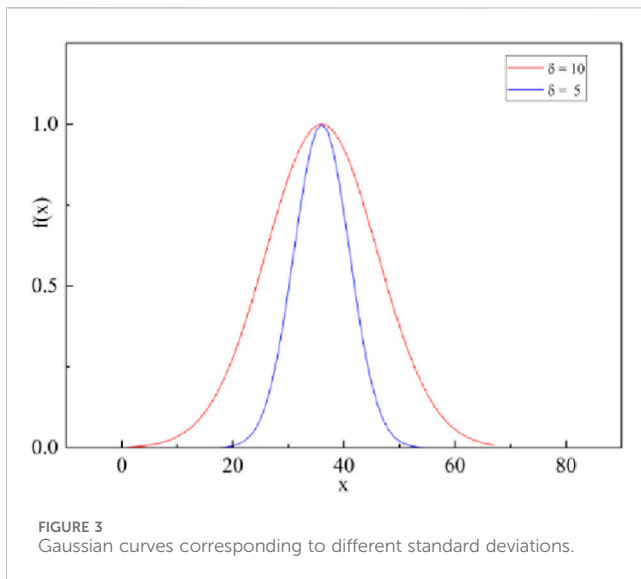
Ratio.   Ratio is defined as the ratio of the total amplitude of the region of interest (ROI) to the total amplitude of the entire frequency spectrum. Since the heart rate of the previous window is close to that of the current window, the historical heart rate has significant reference value for heart rate estimation of the current window. Therefore, based on the heart rate position of the previous window, a Gaussian window is superimposed on the frequency spectrum of the PPG signal in the current window instead of the rectangular window commonly used in traditional algorithms, so as to optimize heart rate estimation. The parameters of the Gaussian function include amplitude A, mean μ and standard deviation δ.

$$f(x) = A \cdot e^{-\frac{(x-\mu)^2}{2\delta^2}} \tag{3}$$

Here, the mean is determined by the heart rate position of the previous window, and the amplitude is set to 1. The value of the standard deviation determines the coverage range of the Gaussian function. As illustrated in Figure 3, it shows the Gaussian curves corresponding to the standard deviations of 10 and 5, respectively. To reduce the impact of motion artifacts on heart rate estimation, the standard deviation is set to 5.

For the frequency spectrum of each window, it is multiplied point-wise by the Gaussian window function described above. The comparison of spectral signals before and after Gaussian windowing is illustrated in Figure 4. The spectral signal after Gaussian windowing is accumulated to obtain the total amplitude denoted as Sum1, while the total amplitude of the spectral signal before Gaussian windowing is recorded as Sum2. The corresponding calculation formula for ratio is thus given by ratio = Sum1/Sum2. This ratio feature can reflect the intensity of motion artifacts to a certain extent: when motion artifacts are significant, the total

FIGURE 3
Gaussian curves corresponding to different standard deviations.

FIGURE 4
Spectrum signals before and after windowing.

amplitude of non-interest regions increases, thus resulting in a smaller ratio value.

Mean and variance of the frequency domain signal after fourier transform of PPGRLS. The mean helps identify the main frequency components of the PPG signal, while the variance reveals the distribution in the frequency domain. A larger variance typically corresponds to a broader frequency distribution.

Mean and variance of PPGpre. The mean reflects issues such as signal drift in the PPG signal, while the variance reflects the signal's fluctuation characteristics.

ACC. The original triaxial acceleration signals before band-pass filtering were processed to calculate their scalar sum, denoted as ACC. This scalar sum can effectively reflect the intensity of motion: when the ACC value is relatively large, it indicates a relatively high

motion intensity. The specific computing method is presented in Equation 4.

$$Acc = \sum_{i=1}^{N} \sqrt{Acc_{x(i)}{}^2 + Acc_{y(i)}{}^2 + Acc_{z(i)}{}^2} \quad (4)$$

Where N denotes the total number of sampling points in each window, and ACCx(i), ACCy(i), and ACCz(i) represent the acceleration signals in the x, y, and z directions, respectively.

Absolute value of the difference in Acc calculated from adjacent windows. This reflects the change in motion state between two consecutive windows. A smaller difference indicates consistent motion states between the two adjacent windows.

Mean and variance of the magnitude of the resultant acceleration (Ampacc) of the triaxial acceleration. The mean is used to distinguish motions of different intensities—high-intensity motions typically correspond to a higher mean. The variance characterizes the regularity of the current motion; abrupt changes in motion state will lead to an increase in variance. The specific computing method is presented in Equation 5.

$$Amp_{acc}(i) = \sqrt{Acc_{x(i)}{}^2 + Acc_{y(i)}{}^2 + Acc_{z(i)}{}^2} \quad 1 \le i \le N \quad (5)$$

Crest factor (CF) of PPGRLS and ACCcom after fourier transform. CF is defined as the ratio of the peak value to the root mean square (RMS) of the entire signal. A larger crest factor indicates that the dominant frequency peak is relatively more prominent compared to other frequency peaks. For PPG signals, a larger crest factor implies better current signal quality. For acceleration signals, a larger crest factor suggests that the current motion has stronger stability and regularity. The specific computing method is presented in Equation 6.

$$CF = \frac{x_{peak}}{x_{rms}} \quad (6)$$

Where xpeak denotes the peak value of the dominant frequency in the frequency spectrum of each window, and xrms denotes the root mean square of the spectral amplitude in each window.

Classification model performance evaluation metrics

This paper will demonstrate the classification accuracy through both numerical and graphical methods. Numerically, Accuracy, Precision, Recall, F1-score, and Macro-F1 are selected. A graphical confusion matrix is used to more intuitively reflect the classification accuracy. The specific descriptions are as follows:

Accuracy. The proportion of correctly predicted samples to the total number of samples. The specific computing method is presented in Equation 7.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (7)$$

TP (True Positive): The model predicts a sample as belonging to a certain class, and the actual label of the sample also belongs to that class.

TN (True Negative): The model predicts a sample as not belonging to a certain class, and the actual label of the sample also does not belong to that class.

FP (False Positive): The model predicts a sample as belonging to a certain class, but the actual label of the sample does not belong to that class.

FN (False Negative): The model predicts a sample as not belonging to a certain class, but the actual label of the sample belongs to that class.

Precision.   The proportion of correctly predicted positive samples among all samples predicted as positive. The specific computing method is presented in Equation 8.

$$Precision = \frac{TP}{TP + FP} \qquad (8)$$

Recall.   The proportion of correctly predicted positive samples among all actually positive samples. The specific computing method is presented in Equation 9.

$$Recall = \frac{TP}{TP + FN} \qquad (9)$$

F1-score.   The harmonic mean of precision and recall.

This metric comprehensively considers both precision and recall, and thus better reflects the overall performance of the model. In some cases, improving precision may lead to a decrease in recall, and this metric balances the two. The specific computing method is presented in Equation 10.

$$F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \qquad (10)$$

Macro-F1.   The arithmetic mean of F1-scores calculated for each class.

This metric emphasizes that all classes are equally important and is not affected by differences in the number of samples across classes, thereby reflecting the comprehensive performance of the model. Therefore, it can provide a balanced evaluation in cases of class imbalance or when each class is of equal importance. The specific computing method is presented in Equation 11.

$$Macro\_F_1 = \frac{1}{C}\sum_{i=1}^{C} F_i \qquad (11)$$

C represents the number of classes.

Confusion matrix.   The structure diagram of the confusion matrix is shown in Figure 5A; when the value of each cell is represented by the depth of color, a corresponding heatmap can be generated, as shown in Figure 5B.

Spectral peak tracking
STFT

Based on the classification results in Section 3.2.3.2, signals judged to have poor quality require further analysis. Since the traditional Fourier transform cannot meet the needs of spectral analysis for such non-stationary signals whose spectral structure

varies with time, a joint time-frequency analysis method is adopted to adapt to these non-stationary signals (Xiao and Feng, 2010). Therefore, STFT applies a window function with finite length to perform sliding window processing on the signal. It is assumed that the signal within each window is stationary, and then Fourier analysis is conducted on the signal of each window. Finally, the spectral information of all windows is combined to obtain time-frequency spectrum information.

STFT involves several key parameters: window length, sliding step, and window function. The sampling frequency of the PPG signal is 125 Hz, and the total window length is set to 8 s. Considering that the dataset was collected under motion conditions and to ensure that each sub-window contains at least one complete PPG cycle, the corresponding window length is set to 100 sampling points, with a sliding step of 100 sampling points, resulting in no overlap between adjacent sub-windows. The Hanning window is selected as the window function to reduce spectral leakage.

STFT analysis is performed on the PPG signal to identify the position of the dominant frequency peak in the frequency spectrum of each sub-window. If the corresponding peak positions of adjacent sub-windows remain consistent, the frequency spectrum of the corresponding time-domain signals is considered continuous, and the time-domain signals in these sub-windows can be spliced together. The spliced signal can be regarded as the low-artifact segment within the window. As illustrated in Figure 6, although the spectrogram of the PPG signal processed by STFT retains the characteristics of the real PPG signal, it is subject to the time-frequency resolution limitation of STFT. Signal segmentation reduces the signal length, which in turn leads to a corresponding decrease in spectral resolution. Consequently, the exact position of the real heart rate cannot be directly located from the spectrum, and only its approximate range can be determined.
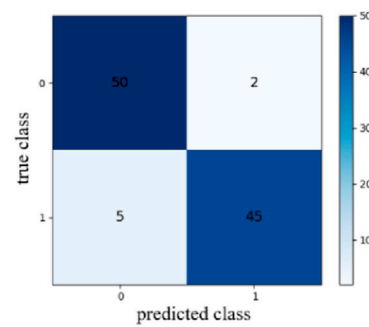
Spectral peak selection

During the spectral peak selection process, different spectral peak selection rules are adopted based on the signal quality assessment results, and the specific workflow is illustrated in Figure 7.

For PPG signals with good quality, the PPGRLS is first subjected to Fourier transform. Then, according to the position of the previous window (locpre), a Gaussian window as described in Section 3.2.3.3 is superimposed on the frequency spectrum. The heart rate spectral peak of such signals is sharp, single, and stable, which does not require coverage by a wide window. A narrow window can improve spectral peak resolution and avoid introducing noise spectral peaks. Moreover, Chung et al. (2019) reported that 99% of the absolute heart rate differences within 2 s between consecutive windows in the ISPC database are approximately 5 BPM. Therefore, the standard deviation is set to δ = 5 and the mean is set to μ = locpre.

For signals with poor quality, the dominant frequency position (locSTFT) is first determined by STFT. However, it should be noted that the interval corresponding to locSTFT does not necessarily reflect the true heart rate. For instance, if motion artifacts affect the entire window, the extracted signal may correspond to the frequency range of motion artifacts instead. Therefore, it is necessary to compare the relative distance between locSTFT and locpre to determine whether their distance falls within the preset threshold range THloc (THloc = 15). If the distance between the two is less than THloc, the interval is

FIGURE 5
Confusion matrix diagram. **(A)** Confusion matrix structure diagram. **(B)** Confusion matrix heatmap.
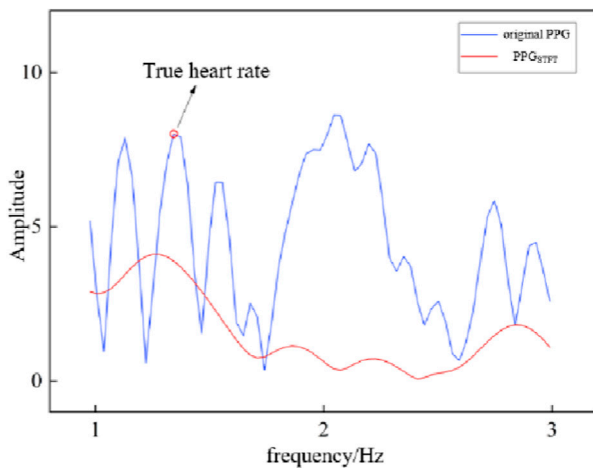

FIGURE 6
The original PPG signal spectrum contains multiple spectral peaks with large amplitudes. After STFT analysis, the spectral energy of the PPG signal is concentrated around the true heart rate, which facilitates subsequent spectral peak screening. In addition, it can be clearly observed from the figure that STFT analysis retains the true PPG signal while removing most of the motion artifact (MA) information.

regarded as the heart rate interval. Due to the low resolution of STFT, it is necessary to conduct analysis in combination with the original PPG signal. Specifically, the spectral peak (loc1) closest to locSTFT is identified in the frequency spectrum of the PPG signal. Motion artifacts may cause the heart rate spectral peak to broaden, shift, or even split. A wide window can cover the potential range of the true spectral peak and avoid missed detection caused by spectral peak shift. Furthermore, Zhang et al. (2015) mentioned that the BPM variation between two consecutive windows rarely exceeds 10 bpm. Thus, a Gaussian window with the mean of loc1 and standard deviation of 10 is superimposed on the original PPG spectral peak. If the distance does not meet the threshold requirement, the spectral filtering method is similar to that for signals with good quality, except that the standard deviation is set to 10.

In addition, multiple groups of gradient combinations of standard deviations should be designed around the two key values of 5 and

10 for comparative experiments. The model performance under different parameters is evaluated to select the optimal standard deviation of the Gaussian window suitable for the target task.

After the above windowing process, the abscissa corresponding to the maximum point of the frequency spectrum is taken as the estimated heart rate position (Loccur), and the heart rate value corresponding to each window is calculated. The specific computing method is presented in Equation 12.

$$BPM_{cur} = \frac{Loc_{cur} - 1}{N_{FFT}} \times 60 \times F_S \tag{12}$$

NFFT represents the number of points in the Fourier transform, which is 4,096, and Fs corresponds to the sampling frequency.

Post-processing

Post-processing is mainly aimed at the removal of outliers generated during the heart rate estimation process and the smoothing of the heart rate curve, so as to improve the accuracy and stability of heart rate data. For outlier removal, this study adopts the backtracking verification method, which conducts detection when the heart rate values are reliable and corrects the heart rate values based on the information of abnormal windows. The specific flowchart is illustrated in Figure 8.

After completing backtracking verification, the heart rate curve after backtracking is smoothed using the Savitzky-Golay filter. The Savitzky-Golay filter can better preserve the local characteristics of the signal while achieving smoothing. The performance of this filter is mainly determined by two parameters: window size and polynomial order. Through experimental verification and parameter tuning, the window size is set to 11 and the polynomial order is set to 3. Filtering can remove the outliers that were not identified during backtracking verification, making the heart rate curve smoother and closer to the true heart rate.

Methods and metrics for performance evaluation of heart rate extraction model

In the algorithm performance verification of this study, three comparative algorithms are selected: WFPV (Temko, 2017) with Wiener filtering denoising, SPECMAR (Islam et al., 2019) with
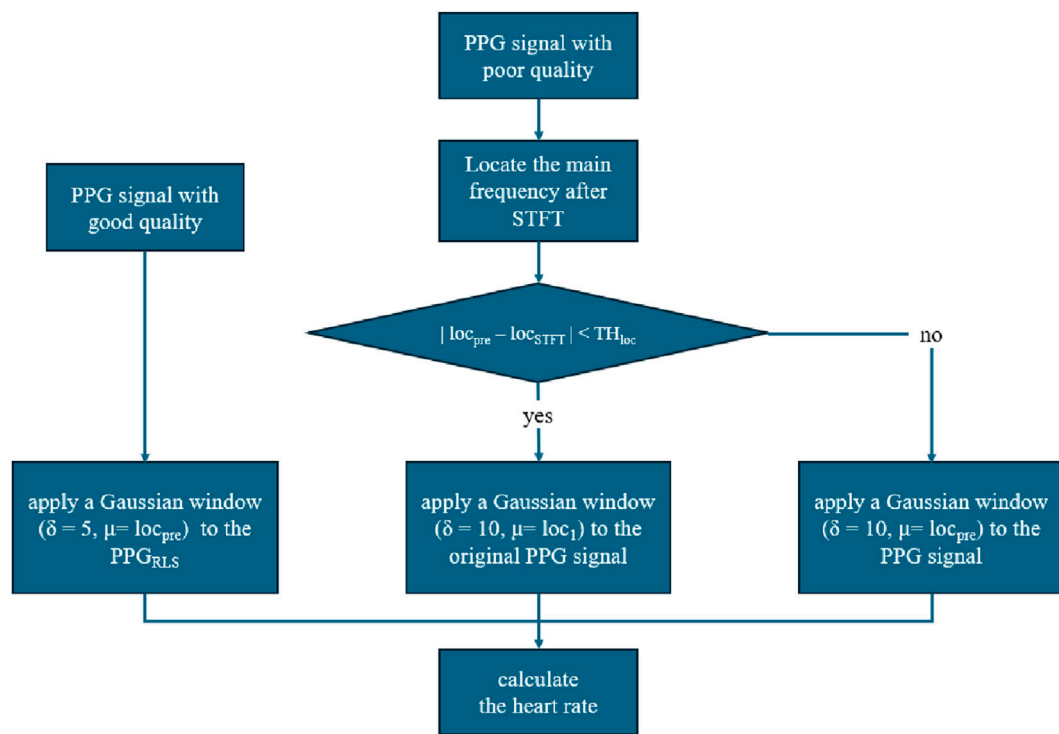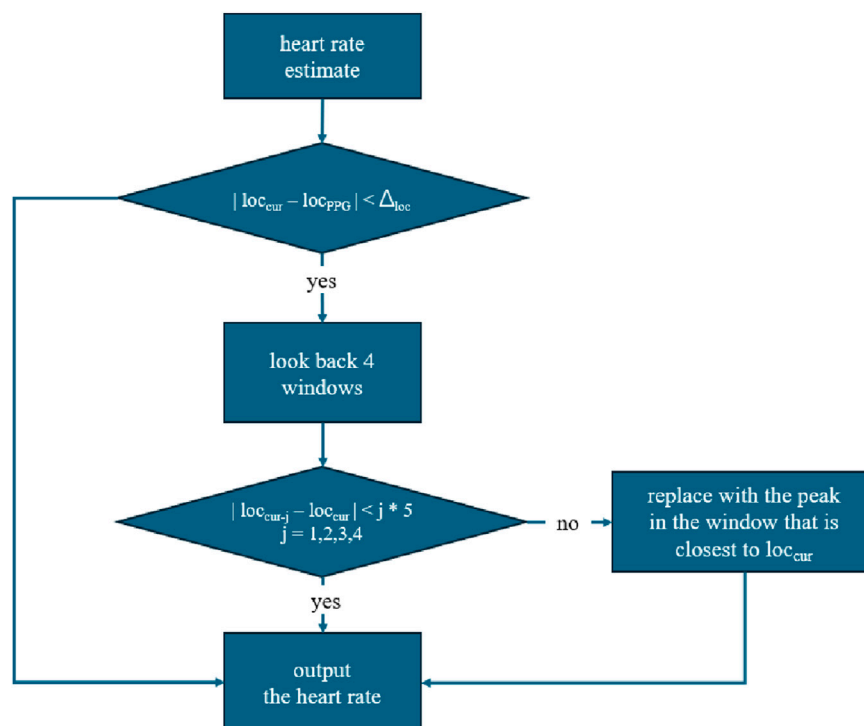
FIGURE 7
Spectral peak screening flow chart.

FIGURE 8
Backtracking verification flow chart.

spectral subtraction denoising, and SSR_Kalma (Zhang et al., 2022) based on Sparse Signal Reconstruction (SSR). To verify the generalization ability of the proposed algorithm, its performance is analyzed based on the results on the test dataset. Meanwhile, to further compare the performance of the proposed algorithm with that of the algorithm proposed by Lan et al. (2024), additional tests are conducted on the self-collected dataset.

To better evaluate the algorithm performance, three evaluation metrics are adopted in this study, namely Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Pearson correlation coefficient plot.

MAE

It refers to the average of the absolute values of the differences between the estimated heart rate and the true heart rate, which can comprehensively reflect the accuracy of the calculation results. The specific computing method is presented in Equation 13.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|BPM_{est}(i) - BPM_{true}(i)| \qquad (13)$$

Where N denotes the total number of windows in each group of data, $BPM_{est}(i)$ denotes the estimated heart rate value of the i-th window, and $BPM_{true}(i)$ denotes the reference heart rate value of the i-th window.

MAPE

It is the average of the ratios of absolute errors to the true heart rate, and it is a relative indicator. When comparing errors across different datasets, it can eliminate the impact caused by individual differences. The specific computing method is presented in Equation 14.

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|BPM_{est}(i) - BPM_{true}(i)|}{BPM_{true}(i)} \times 100\% \qquad (14)$$

Pearson correlation coefficient plot

From a visualization perspective, the Pearson correlation coefficient plot is used to reflect the relationship between the estimated heart rate and the true heart rate. The Pearson correlation coefficient is a parameter that measures correlation, and its value is positively correlated with the degree of correlation: the larger the value, the stronger the correlation. The specific computing method is presented in Equation 15.

$$\rho(X, Y) = \frac{cov(X, Y)}{\delta_X \delta_Y} \qquad (15)$$

Where cov (X,Y) denotes the covariance between X and Y, and δX, δY denote the standard deviations of X and Y, respectively.

Bland-Altman plot

The Bland-Altman plot visually demonstrates the difference between the true heart rate and the predicted heart rate in a graphical manner. This plot takes the mean of the two heart rates as the abscissa and their difference as the ordinate to draw a scatter plot. By calculating the mean of the differences and the standard deviation (sd) of the differences, if the scatter points are concentrated within the 95% confidence interval of the

differences (i.e., mean ± 1.96sd), the predicted heart rate and the true heart rate can be considered to have a high degree of consistency.

Results

Comparison of PPG signals before and after RLS filtering

Figures 9A,B respectively show the spectrograms of the PPG signal before and after RLS filtering within the same time window. Among them, the red dots mark the spectral peaks corresponding to the true heart rate. A clear comparison reveals that before RLS filtering, there are two motion artifact (MA) spectral peaks with large amplitudes, and one of these spurious peaks is close to the spectral peak of the true heart rate. If spectral peak tracking is directly performed on this spectrum, the spectral peak of the true heart rate may be masked, impairing the accuracy of heart rate estimation.

After RLS filtering, the 2 MA spectral peaks are significantly suppressed, which highlights the spectral peak corresponding to the true heart rate and makes it the dominant peak within the frequency band. RLS filtering effectively eliminates the interference of MA, thus significantly improving the signal quality.

Classification training results

Performance comparison of key parameter combinations for each classifier

Pre-experimental results indicate that the key parameters affecting the classification performance of the SVM are the kernel function (kernel) and the regularization parameter (C), while the number of neighbors (n_neighbors) is the core parameter determining the classification performance of the K-Nearest Neighbor classifier. For this purpose, multiple groups of gradient parameter combinations are designed for the two classifiers respectively based on the above key parameters, and comparative experiments are conducted. By evaluating the classification performance of the models under different parameter combinations, the parameter configurations suitable for the target task are screened out. Tables 3, 4 respectively list the classification performance comparison results of the SVM and KNN classifiers under different parameter combinations, based on which the values of the key parameters of the two classifiers can be determined.

Subsequently, when evaluating the classification performance on the test set, the parameters of the SVM are set as follows: kernel = 'linear', C = 1.0, class_weight = {0:10,1:100}; the parameters of the K-Nearest Neighbor classifier are set as: n_neighbors = 2. In addition, it was also found through pre-experiments that the performance differences of the Random Forest classifier under different parameter combinations are negligible on the test set, so there is no need to conduct additional parameter comparison experiments. Its final parameter configuration is set as: number of decision trees (n_estimators) = 100, maximum tree depth (max_depth) = None, class weight = {0:10,1:100}.

FIGURE 9
Comparison of PPG signal spectra before and after RLS filtering. **(A)** Before RLS filtering. **(B)** After RLS filtering.

TABLE 3 Classification performance comparison of SVM under different parameter combinations.

Performance metrics	Key parameters	kernel	Linear			poly			rbf		
		C	0.1	0.5	1.0	0.1	0.5	1.0	0.1	0.5	1.0
Accuracy			0.97	0.98	0.98	0.33	0.53	0.67	0.93	0.97	0.98
Macro_F1			0.69	0.73	0.74	0.26	0.37	0.43	0.54	0.57	0.55
Precision			1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.99
Recall			0.97	0.97	0.98	0.32	0.53	0.66	0.93	0.97	0.98
F1			0.98	0.99	0.99	0.49	0.69	0.80	0.96	0.98	0.99

TABLE 4 Classification performance comparison of K-nearest neighbor under different parameter combinations.

Performance metrics	Key parameters	n_neighbors	1	2	3	5	7
Accuracy			0.84	0.86	0.77	0.74	0.73
Macro_F1			0.46	0.46	0.44	0.43	0.43
Precision			0.99	0.99	0.99	0.99	0.99
Recall			0.85	0.87	0.78	0.75	0.74
F1			0.91	0.92	0.87	0.85	0.84

Performance comparison and analysis of classifiers under optimal parameters

Based on the parameters set in Section 4.2.1, the performance of different classifiers was evaluated using the classification results on the test set, as detailed in Table 5.
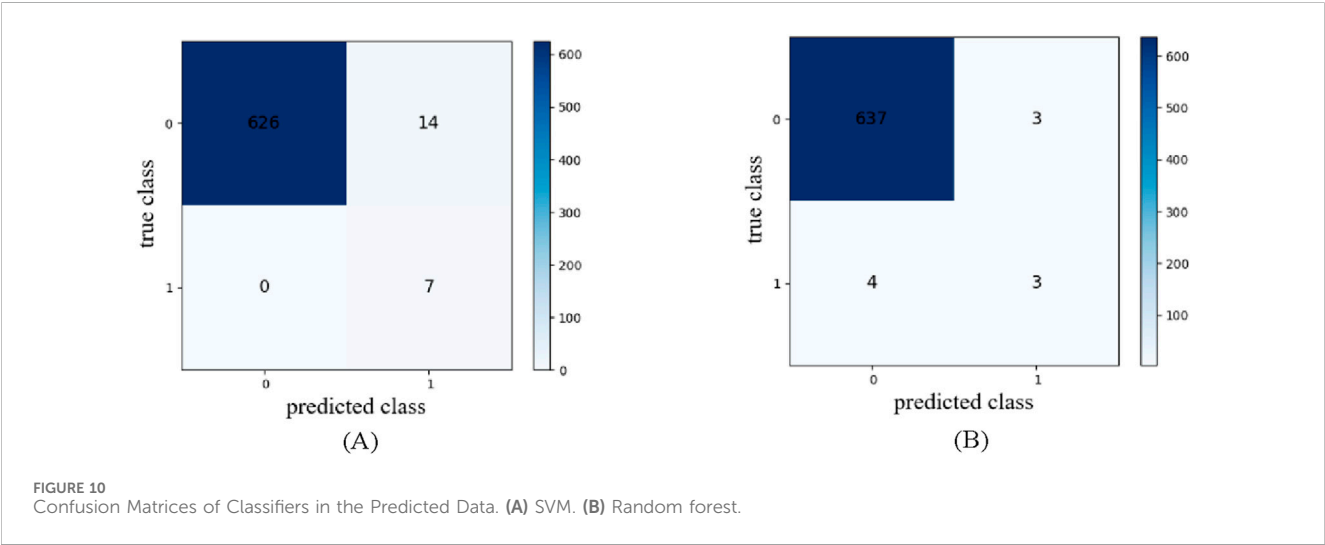
In the performance comparison of classifiers, the SVM and Random Forest significantly outperform the K-Nearest Neighbor. The Accuracy and Macro-F1 values of the K-Nearest Neighbor are much lower than those of the other two classifiers. In terms of numerical values, the gap between Random Forest and SVM is small. In this study, although the number of positive and negative samples is imbalanced, both types of samples are of great importance. Therefore, it is necessary to consider the classification performance of both positive and negative samples

simultaneously. From a macro perspective, Macro-F1 reflects the overall performance of different classifiers: the Macro-F1 of SVM is 0.74, which is slightly higher than that of Random Forest (0.73). From a visualization perspective, confusion matrices of the SVM and Random Forest on the predicted data are plotted, as shown in Figures 10A,B, respectively.

The depth of color in the figures clearly indicates a severe imbalance in the number of positive and negative class samples, where the number of negative class samples is much larger than that of positive class samples. The Random Forest correctly predicts only 3 positive class samples, showing a distinct negative-class bias during the prediction process. This result suggests that the Random Forest overfits the features of the majority class, leading to insufficient discrimination ability for the minority class. This phenomenon may

TABLE 5 Classification performance of different classifiers.

Performance metrics	SVM	Random forest	K-nearest neighbors
Accuracy	0.98	0.99	0.86
Macro_F1	0.74	0.73	0.46
Precision-	1.00	0.99	0.99
Recall-	0.98	1.00	0.87
F1-	0.99	0.99	0.92

FIGURE 10
Confusion Matrices of Classifiers in the Predicted Data. **(A)** SVM. **(B)** Random forest.

stem from the fact that when constructing decision trees, the Random Forest tends to select features that better distinguish the majority class while ignoring those of the minority class. In this study, the positive class corresponds to signals evaluated as poor quality, so the ability to identify the positive class is of great importance. In contrast, when constructing the decision boundary, the SVM relies on support vectors that include both majority and minority class samples, without neglecting the minority class. Thus, it performs better in positive class prediction and can identify more positive class samples. Although it misclassifies 14 negative class samples as positive ones, this result demonstrates to a certain extent that the SVM does not completely favor the majority class. Instead, it sacrifices a small degree of precision for the negative class to improve the recall rate of the positive class, reflecting its sensitivity to the minority class and adaptability to imbalanced data distribution. Based on the above analysis, the SVM achieves superior comprehensive performance compared with the Random Forest and the K-Nearest Neighbor and thus it is selected as the classifier for subsequent research.

Results of Gaussian window standard deviation screening

To select the optimal standard deviation values of the Gaussian window for PPG signals of different quality, multiple groups of gradient combinations of standard deviations were designed for

performance comparison experiments. For PPG signals with good quality, 3, 5, and 8 were selected as gradient parameters with 5 as the core value; for PPG signals with poor quality, 8, 10, and 12 were selected as gradient parameters with 10 as the core value. Experimental verification was completed based on pairwise combinations of the two parameter sets.

The comparison results of two representative parameter combinations are presented in Tables 6, 7 respectively: Table 6 focuses on the scenario where the standard deviation for good-quality signals is fixed at 5, comparing its performance with that of poor-quality signals using standard deviations of 8, 10, and 12; Table 7 focuses on the scenario where the standard deviation for poor-quality signals is fixed at 10, comparing its performance with that of good-quality signals using standard deviations of 3, 5, and 8. The experimental results show that the parameter combination with a standard deviation of 5 for good-quality signals and 10 for poor-quality signals achieves the optimal matching degree with the heart rate estimation algorithm task.

The above conclusions possess universality and representativeness: when a standard deviation of 3 or 8 is used for good-quality signals, the performance of using a standard deviation of 10 for poor-quality signals is superior to that of 8 or 12 in the vast majority of test scenarios; similarly, when a standard deviation of 8 or 12 is used for poor-quality signals, the performance of using a standard deviation of 5 for good-quality signals outperforms that of 3 or 8 in most test scenarios. Therefore, only the experimental results of representative parameter combinations are presented in this study to simplify data presentation and highlight the core conclusions.

TABLE 6 Performance comparison of poor-quality signals with different standard deviations when the standard deviation of good-quality signals is fixed at 5.

No.	Motion type	8		10		12	
		MAE/BPM	MAPE	MAE/BPM	MAPE	MAE/BPM	MAPE
1	T2	1.00	0.72%	0.96	0.70%	1.00	0.72%
2	T1	3.24	3.58%	2.56	2.83%	2.96	3.27%
3	T2	3.75	2.75%	2.76	2.05%	3.08	2.27%
4	T2	1.72	1.38%	1.63	1.31%	1.72	1.38%
5	T1	0.63	0.74%	0.63	0.74%	0.63	0.74%
Mean	-	2.07	1.83%	1.71	1.52%	1.88	1.67%

TABLE 7 Performance comparison of good-quality signals with different standard deviations when the standard deviation of poor-quality signals is fixed at 10.

No.	Motion type	3		5		8	
		MAE/BPM	MAPE	MAE/BPM	MAPE	MAE/BPM	MAPE
1	T2	4.12	2.95%	0.96	0.70%	1.68	1.22%
2	T1	4.31	4.43%	2.56	2.83%	2.67	3.00%
3	T2	2.58	1.92%	2.76	2.05%	2.98	2.17%
4	T2	4.00	3.20%	1.63	1.31%	5.57	4.39%
5	T1	0.74	0.87%	0.63	0.74%	0.63	0.74%
Mean	-	3.15	2.67%	1.71	1.52%	2.71	2.30%

Comparison of heart rate results before and after post-processing

Figure 11 shows the comparison diagram of heart rate results before and after post-processing. As can be seen from the figure, the estimated heart rate curve without post-processing exhibits obvious sawtooth fluctuations, with particularly significant errors observed at the 21st and 83rd windows. After the introduction of the post-processing strategy, the degree of fitting between the estimated heart rate and the true heart rate is significantly improved. Specifically, backtracking verification reduces abnormal jumps, while Savitzky-Golay filtering effectively suppresses high-frequency interference while retaining the main trend of the signal, thereby improving the accuracy of the estimated heart rate.

Performance comparison and analysis of heart rate estimation models

The specific results are presented in Table 8, and the corresponding MAE comparison plot is plotted as shown in Figure 12.

Overall, the MAE and MAPE values of the proposed algorithm on the test data are both lower than those of other comparative algorithms. In terms of maximum error, the values are 3.44 BPM for WFPV, 4.80 BPM for SPECMAR, 3.38 BPM for SSR_Kalman, and 2.76 BPM for the proposed algorithm. This result indicates that the proposed algorithm has higher stability than other algorithms,

further verifying that it can maintain favorable performance even on unseen data.

Grouped by exercise type, the MAE values of each algorithm are illustrated in Figure 13. In the T1 exercise type, WFPV achieves the smallest error; the error of the proposed algorithm is slightly higher than that of WFPV, while SSR_Kalman yields the largest error. In the T2 exercise type, the proposed algorithm attains the minimum error, whereas SPECMAR has the maximum error. From the perspective of stability, the proposed algorithm exhibits a small difference in error between the two exercise types, demonstrating good consistency. In contrast, all other algorithms show a large discrepancy in error across the two exercise types. Overall, the proposed algorithm outperforms the other comparative algorithms in terms of performance.

As shown in Figure 14 are the statistical results of the proposed algorithm on the test set.

Panel (A) is a linear correlation plot, where the fitting line of the true heart rate versus the predicted heart rate is expressed as $y = 1.00849x - 0.79783$, with a Pearson correlation coefficient of 0.99201. The difference between this value and the Pearson correlation coefficient of the training set is small, which indicates that the SVM has a certain generalization ability in signal quality prediction and also verifies the high accuracy of the proposed algorithm. Panel (B) is a Bland-Altman analysis plot, with the mean of errors being 0.085 BPM. As can be seen from the plot, most scatter points are concentrated around the mean of errors, among which 93.4% of the points fall within the 95% confidence interval. The distribution of scatter points shows no obvious trend, suggesting that the
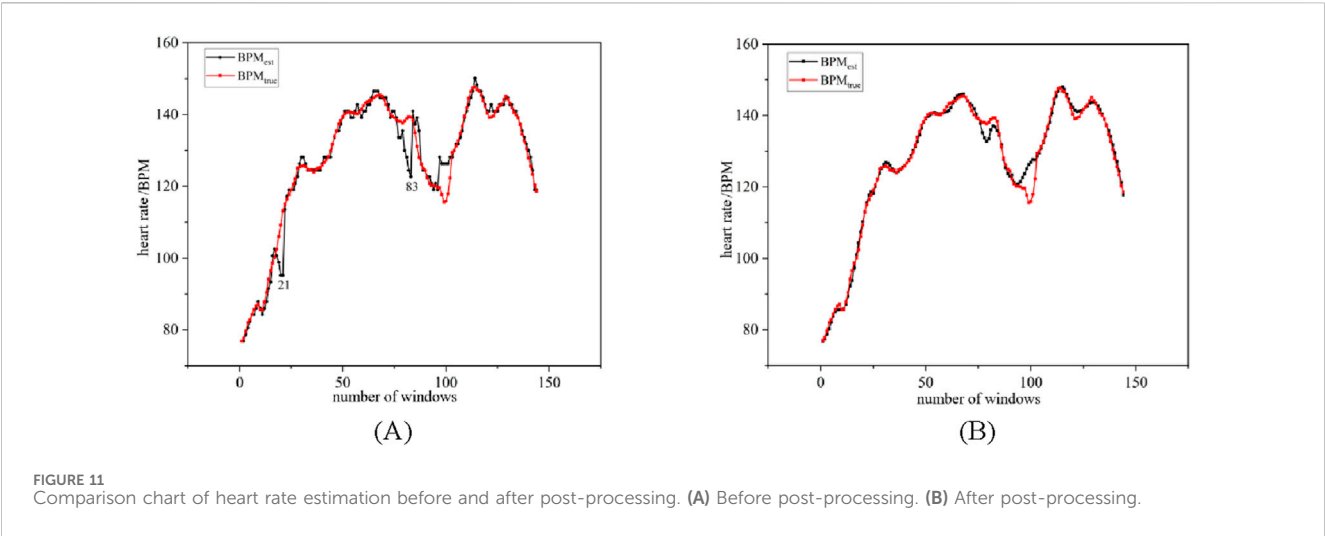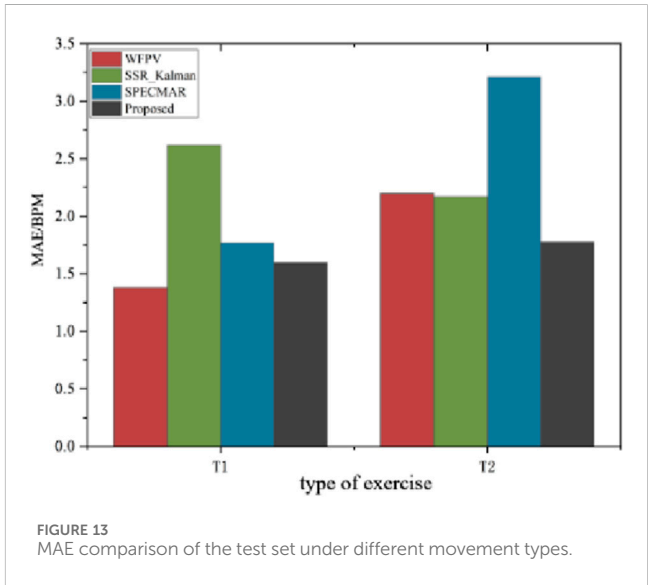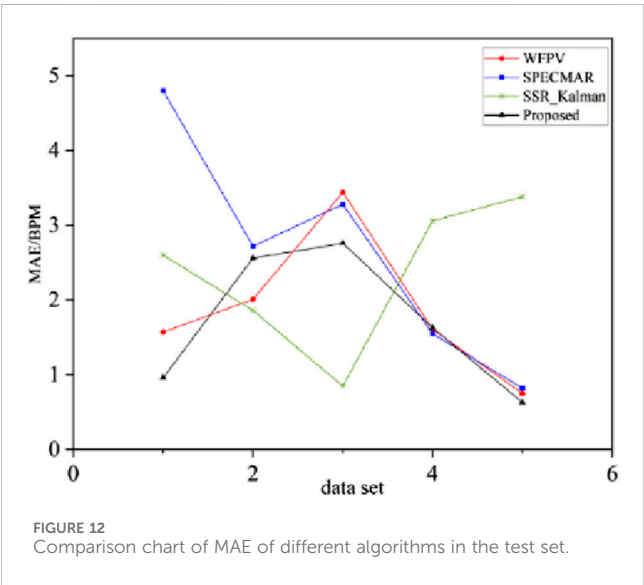
FIGURE 11
Comparison chart of heart rate estimation before and after post-processing. **(A)** Before post-processing. **(B)** After post-processing.

TABLE 8 Heart rate results of test data.

No.	Motion type	WFPV		SPECMAR		SSR_Kalman		Proposed	
		MAE/BPM	MAPE	MAE/BPM	MAPE	MAE/BPM	MAPE	MAE/BPM	MAPE
1	T2	1.57	1.15%	4.80	-	2.60	2.86%	0.96	0.70%
2	T1	2.01	2.41%	2.72	-	1.86	1.44%	2.56	2.83%
3	T2	3.44	2.45%	3.28	-	0.85	0.99%	2.76	2.05%
4	T2	1.61	1.26%	1.55	-	3.06	2.54%	1.63	1.31%
5	T1	0.75	0.88%	0.82	-	3.38	2.32%	0.63	0.74%
Mean	-	1.88	1.63%	2.63	-	2.35	2.03%	1.71	1.52%

FIGURE 12
Comparison chart of MAE of different algorithms in the test set.

FIGURE 13
MAE comparison of the test set under different movement types.

consistency between the estimated heart rate and the true heart rate is relatively stable across different heart rate ranges.

The test results are presented in Table 9, where Proposed1 refers to the algorithm proposed by Lan et al. (2024) and Proposed2 denotes the

proposed algorithm in this study. Overall, Proposed2 outperforms Proposed1 in both MAE and MAPE metrics. The proposed algorithm achieves certain improvements on partial datasets, with the improvement being relatively significant especially on the 4th and
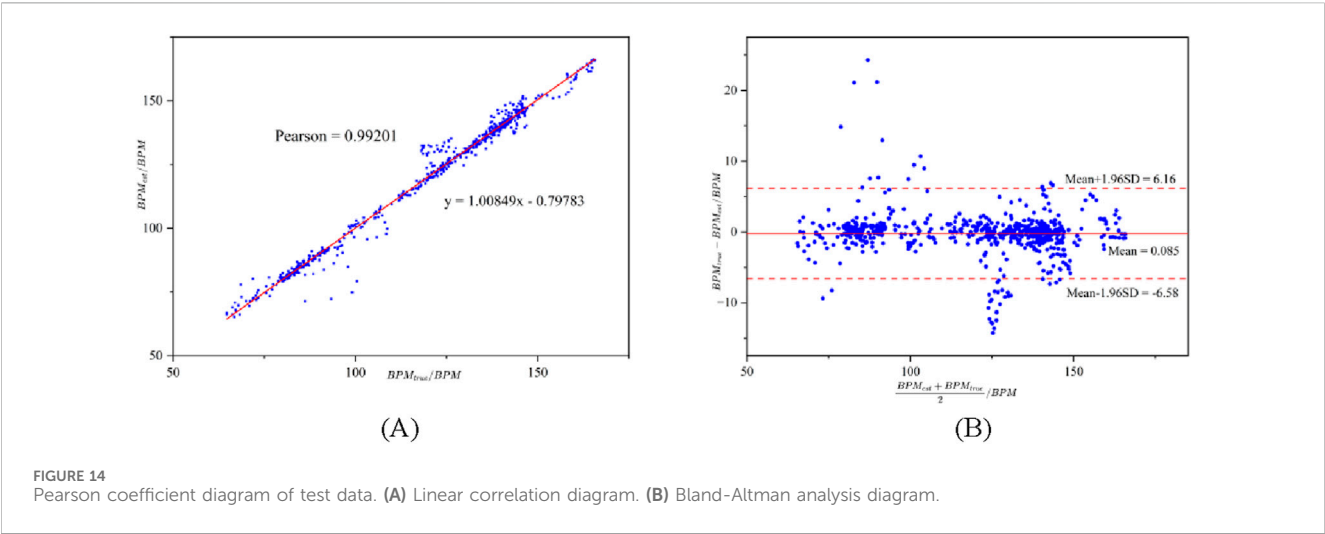
FIGURE 14
Pearson coefficient diagram of test data. **(A)** Linear correlation diagram. **(B)** Bland–Altman analysis diagram.

TABLE 9 Comparison of self-collected data results.

No.	Proposed1		Proposed2	
	MAE/BPM	MAPE	MAE/BPM	MAPE
1	1.95	2.35%	1.75	2.13%
2	1.40	1.59%	1.05	1.19%
3	2.07	1.95%	2.09	1.89%
4	1.53	1.29%	0.75	0.62%
5	1.42	1.28%	0.79	0.72%
6	1.50	1.60%	1.57	1.69%
7	1.65	2.23%	2.07	2.77%
8	2.56	3.28%	2.67	3.47%
Mean	1.76	1.95%	1.62	1.81%

5th datasets. However, both algorithms yield large errors on the 8th dataset, indicating that the anti-noise capability of the improved algorithm still needs further enhancement.

Discussion

Nowadays, the prevalence of cardiovascular diseases continues to rise, having become one of the major threats to human health. Real-time heart rate monitoring based on PPG signals is conducive to the prevention, control, and management of diseases, while also promoting the development of personalized health management and intelligent medical care. However, interference from motion artifacts and differences in physiological states among individuals (such as age, weight, skin color, etc.) increase the difficulty of signal processing and lead to a decline in the generalization ability of algorithms. Especially in dynamic scenarios, such as running and daily activities, motion artifacts can significantly affect the accuracy of heart rate extraction. Existing algorithms generally suffer from insufficient robustness and struggle to cope with complex and varied application scenarios. Therefore, this paper conducts research on heart rate extraction from PPG signals, covering traditional signal processing to machine learning, with the main work completed as follows:

To improve the stability and generalization ability of heart rate calculation, improvements were made to the signal quality assessment and spectrum analysis methods. Since the FSM (Finite State Machine) considers fewer parameters when judging signal states, has limited adaptive capacity, and is easily affected by fluctuations in signal quality, it was thus improved. Signal quality assessment was implemented using SVM, with 11 features selected from the time-frequency domain. SVM can effectively handle non-linear features, enabling the algorithm to identify the essential characteristics of signals from complex ones and integrate multi-dimensional features to enhance the algorithm's stability. In addition, to better adapt to the non-stationary and time-varying characteristics of PPG signals, STFT was combined to analyze their time-frequency properties, which can better capture such dynamic changes. During spectrum analysis, a Gaussian window was used instead of a rectangular window. Based on the amplitude distribution of the Gaussian window, higher weights can be assigned to the interval of historical heart rates, increasing the tracking of historical heart rates and reducing sensitivity to instantaneous fluctuations. Compared with the high-performance WFPV algorithm, the proposed algorithm shows smaller calculation errors, more stable performance in heart rate results under different motion states, and better generalization.

The manual feature extraction method adopted in this study has achieved ideal heart rate extraction performance on the test dataset. Its core advantage lies in the high adaptability of feature design to the physiological mechanisms and noise characteristics of PPG signals. The selected time-domain and frequency-domain features all have clear physical meanings, which avoids the problem of insufficient interpretability caused by the "black-box" nature of deep learning models. Furthermore, it does not require large-scale labeled data or high computational resources, demonstrating certain efficiency and stability in heart rate monitoring tasks with small samples and specific scenarios (e.g., static or low-intensity exercise).

Although deep learning methods (such as Long Short-Term Memory) have become a research hotspot in physiological signal analysis due to their ability to automate feature engineering—for

instance, the end-to-end model proposed by Oğuz et al. (2023) can adaptively capture complex nonlinear relationships in signals—the manual feature method in this study still possesses irreplaceable academic value and application scenarios. On the one hand, in scenarios with relatively single data distribution and clear noise types, manually designed targeted features can effectively reduce the model's overfitting risk. Additionally, with low computational complexity, it is more suitable for platforms with limited computational resources such as wearable devices. On the other hand, its strong interpretability meets the core requirement of technical reliability in the field of biomedical engineering, providing a clear decision-making basis for heart rate monitoring in clinical scenarios. This research approach does not negate the advantages of deep learning; instead, through targeted feature engineering, it offers a "lightweight and highly interpretable" complementary solution for PPG heart rate extraction, enriches technical choices in different application scenarios, and provides a potential direction for the subsequent design of hybrid models combining deep learning and manual features.

This study focuses on heart rate extraction from PPG signals. Despite achieving favorable results, it still has limitations. The data processed by SMOTE still exhibits extreme class imbalance, as quantitative balance does not equate to feature balance, and the core contradiction remains unresolved. Although this imbalance alleviates the difference in sample quantity distribution through augmentation, it fails to address key issues such as insufficient representativeness of minority class features, feature distribution overlap, and noise amplification of synthetic samples. Consequently, the decision boundary shifts toward the majority class during model training, traditional accuracy metrics are distorted due to their bias towards the majority class, and the model's generalization ability is limited. It is difficult to adapt to clinically critical scenarios corresponding to minority class samples in biomedical signal classification (e.g., low-quality PPG signals, abnormal heart rate samples), which directly affects the practical application value of the model.

The main directions for future research and optimization are as follows:

1. The use of fixed windows in the algorithm to analyze PPG signals under different states has the problem of insufficient adaptability. To improve the time-frequency resolution of signals, an adaptive window length method can be attempted to more flexibly respond to signal changes under different states.
2. The dataset selected in this paper has certain limitations, with a limited amount of data and insufficient consideration of the diversity of motion states. In the future, data sources will be further expanded, and more data collected under different states will be selected, including those of different ages, skin colors, and health conditions, to enhance the adaptability and generalization of the algorithm. Furthermore, to address the data imbalance problem, we focus on two cores: "feature balance" and "algorithm adaptation". On the one hand, improved oversampling algorithms such as ADASYN and SMOTE-ENN can be adopted to reduce pseudo-samples and noise interference, or domain adaptation and transfer learning can be used to supplement minority class features, thereby achieving true feature balance at the data level. On the other hand, weighted loss functions (e.g., Focal Loss, Weighted

Cross-Entropy) can be combined to suppress the dominant role of the majority class during training, or ensemble learning strategies such as EasyEnsemble and BalanceCascade can be employed to split majority class samples, so as to balance the training process. This enhances the model's ability to recognize minority class samples at the algorithm level and fundamentally alleviates the impact of extreme data imbalance on model performance.
3. Combining Lan et al. (2024)'s research on motion artifact removal, efforts will be made from the perspective of multi-signal fusion. On the one hand, signal enhancement can be achieved through cross-validation and information fusion of multi-channel PPG signals; on the other hand, the motion artifact information contained in acceleration signals can be deeply explored to gain a deeper understanding of the interference caused by motion to PPG signals. By integrating the synergistic effect of multiple signals, the quality of PPG signals in dynamic environments will be further improved.

Data availability statement

The datasets presented in this article are not readily available because The Self-test dataset includes preliminary experimental data related to the optimization of human physiological parameter monitoring equipment. This part of the data is still part of the team's ongoing pre-research work and has not yet completed final verification and standardization. To ensure the accuracy and integrity of subsequent research results, it is temporarily not available for external request. Requests to access the datasets should be directed to JG, guojw@mail.ustc.edu.cn.

Ethics statement

The studies involving humans were approved by the Ethics Committee of the People's Hospital of Suzhou New District (Approval No. 2024-021). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

JG: Writing – original draft, Writing – review and editing. SC: Writing – review and editing. TL: Writing – original draft. RL: Writing – review and editing. LW: Writing – review and editing. YW: Writing – review and editing. JZ: Writing – review and editing. WZ: Writing – review and editing.

Funding

Technology, Chinese Academy of Sciences. The funder is precisely the corresponding author of this paper, JZ.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Arunkumar, K. R., and Bhaskar, M. (2020). CASINOR: combination of adaptive filters using single noise reference signal for heart rate estimation from PPG signals. *Signal, Image Video Process.* 14, 1507–1515. doi:10.1007/s11760-020-01692-6

Breiman, L. (2001). Random forests. *Mach. Learning* 45, 5–32. doi:10.1023/a:1010933404324

Choe, B., Kim, H. Y., and Uhm, J. (2024). "Heart rate imputation using accelerometers for wearable devices," in 2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (IEEE), 1–5.

Chung, H., Lee, H., and Lee, J. (2018). Finite state machine framework for instantaneous heart rate validation using wearable photoplethysmography during intensive exercise. *IEEE J. Biomed. Health Inf.* 23 (4), 1595–1606. doi:10.1109/JBHI.2018.2871177

Chung, H., Lee, H., and Lee, J. (2019). State-dependent Gaussian kernel-based power spectrum modification for accurate instantaneous heart rate estimation. *PLOS ONE* 14 (4), e0215014. doi:10.1371/journal.pone.0215014

Cover, T., and Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* 13 (1), 21–27. doi:10.1109/tit.1967.1053964

Geng, Y., and Zhang, D. (2008). Review of adaptive filtering algorithms. *Inf. Electron. Eng.* (4), 315–320. doi:10.1016/j.sigpro.2021.108276

Holmes, C. C., and Adams, N. M. (2002). A probabilistic nearest neighbour method for statistical pattern recognition. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 64 (2), 295–306. doi:10.1111/1467-9868.00338

Huang, W. A., Li, F., Zhang, Y., Li, J. H., and Gao, J. F. (2023). Heart rate estimation method combining LSTM network and U-Net model. *Sci. Technol. and Eng.* 23 (5), 1875–1881. doi:10.3969/j.issn.1671-1815.2023.05.010

Islam, M. T., Ahmed, S. T., Shahnaz, C., and Fattah, S. A. (2019). SPECMAR: fast heart rate estimation from PPG signal using a modified spectral subtraction scheme with composite motion artifacts reference generation. *Med. and Biol. Eng. and Comput.* 57, 689–702. doi:10.1007/s11517-018-1909-x

Khan, E., Al Hossain, F., Uddin, S. Z., Alam, S. K., and Hasan, M. K. (2015). A robust heart rate monitoring scheme using photoplethysmographic signals corrupted by intense motion artifacts. *IEEE Trans. Biomed. Eng.* 63 (3), 550–562. doi:10.1109/TBME.2015.2466075

Lan, T., Bie, Y. A., Hai, D., and Zhong, J. . (2024). Adaptive estimation algorithm for photoplethysmographic heart rate based on finite state machine. *Appl. Sci.* 14 (24), 11631. doi:10.3390/app142411631

Maeda, Y., Sekine, M., and Tamura, T. (2011). Relationship between measurement site and motion artifacts in wearable reflected photoplethysmography. *J. Med. Syst.* 35 (5), 969–976. doi:10.1007/s10916-010-9505-0

Mazumdar, H., Khondakar, K. R., Das, S., and Kaushik, A. (2025). Soft robotics for Parkinson's disease supported by functional materials and artificial intelligence. *BME Front.* 6, 0143. doi:10.34133/bmef.0143

Meng, R., Li, Z., Yu, H., and Niu, Q. (2022). Heart rate extraction algorithm based on adaptive heart rate search model. *Sheng Wu Yi Xue Gong Cheng Xue Za Zhi* 39 (3), 516–526. doi:10.7507/1001-5515.202101091

Oğuz, F. E., Alkan, A., and Schöler, T. (2023). Emotion detection from ECG signals with different learning algorithms and automated feature engineering. *Signal, Image Video Process.* 17, 3783–3791. doi:10.1007/s11760-023-02606-y

Ray, D., Collins, T., and Ponnapalli, P. V. S. (2022). "DeepPulse: an uncertainty-aware deep neural network for heart rate estimations from wrist-worn photoplethysmography," in 2022 44th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (IEEE), 1651–1654.

Sun, L., and Jia, Y. (2020). "An improved PPG denoising methodology based on EEMD and wavelet threshold," in IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), 2020, 467–471.

Sun, J. X., Du, Y., Zeng, Y., Meng, H. R., Wang, Y., Guo, P. B., et al. (2024). Advances in the application of wearable devices in cardiovascular disease treatment. *Chin. J. Integr. Med. Cardio-/Cerebrovascular Dis.* 22 (20), 3728–3730. doi:10.12102/j.issn.1672-1349.2024.20.013

Temko, A. (2017). Accurate heart rate monitoring during physical exercises using PPG. *IEEE Trans. Biomed. Eng.* 64 (9), 2016–2024. doi:10.1109/TBME.2017.2676243

Xiao, Y., and Feng, C. (2010). A time–frequency representation method of short-time fourier transform with composite window functions. *J. Det. Control* 32 (3), 43–47.

Xiong, J., Cai, L., Wang, F., and He, X. (2017). SVM-based spectral analysis for heart rate from multi-channel WPPG sensor signals. *Sensors* 17 (3), 506. doi:10.3390/s17030506

Zhang, Z., Pi, Z., and Liu, B. (2015). TROIKA: a general framework for heart rate monitoring using wrist-type photoplethysmographic signals during intensive physical exercise. *IEEE Trans. Biomed. Eng.* 62 (2), 522–531. doi:10.1109/TBME.2014.2359372

Zhang, Y. T., Xu, J. Y., Xie, M. M., Wang, W., Ye, K. Y., Wang, J., et al. (2022). "PPG-based heart rate estimation with efficient sensor sampling and learning models," in 2022 IEEE 24th Int Conf on High Performance Computing and Communications; 8th Int Conf on Data Science Systems; 20th Int Conf on Smart City; 8th Int Conf on Dependability in Sensor, Cloud Big Data Systems Application (HPCC/DSS/SmartCity/DependSys) (IEEE), 1971–1978.